Gender, Skills and Infrastructure: Constraints to Private Sector Development in South Asia

Hannah Irmela Uckat

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Trinity Term 2020

Gender, Skills and Infrastructure: Constraints to Private Sector Development in South Asia

HANNAH IRMELA UCKAT

Hertford College
University of Oxford

Abstract This thesis studies firm development and labour markets in South Asia from three different perspectives – gender, skills and infrastructure. Chapter 1 investigates how female career advancement in the workplace impacts on bargaining in the household. It exploits the design of a promotion programme for women in Bangladesh's garment industry, a sector where women continue to be underrepresented even in low-level management positions. The results suggest that women's promotions lead to higher bargaining power at home not just for the promoted women, but also for the female subordinates that they supervise. Chapter 2 – written with Christopher Woodruff – turns to analyse the effects of the same promotion programme on factory outcomes. It reports on the importance of selecting managers based on their initial characteristics as compared to training candidates in the same skills. The evidence suggests that formal diagnostics lead factories to choose candidates with higher initial skills, which are predictive of official promotions and performance. Training results in higher promotion rates but has only marginal effects on performance. Soft skills, such as confidence and leadership, are particularly relevant for the success of the new female managers. Lastly, Chapter 3 documents heterogeneous short-term effects of infrastructure improvements on firms in India. It finds large and significant negative impacts on incumbent manufacturing firms close to improved highways, which are driven by initially less profitable, younger and smaller firms. This appears to provide a snapshot of the process of reallocation of resources towards more successful firms when areas become better connected.

Acknowledgements

I am privileged to have benefited from an extraordinary support network throughout my doctorate, for which I am deeply grateful. First and foremost, I am indebted to my supervisors, Chris Woodruff and Simon Quinn. Both have provided excellent advice, guidance and support not only in academic matters, but also when I needed to take a step back due to illness. In particular, I want to thank Chris for the many doors he opened along the way, and numerous hours of intellectually stimulating, good-humoured discussions. I cannot remember a single meeting that I did not leave as a better researcher (though I certainly did not absorb as much as I could have)! I thank Simon for always being a reference point to turn towards, and for countless acutely observed comments on presentations and drafts. His founding of the Firms and Development Group (later Oxford Firms, Labour and Organisation) created a wonderful environment in which to learn, discuss ideas and receive feedback.

Throughout this doctorate, the community at and around the Centre for the Study of African Economies became my academic as well as my social home. It is filled with genuinely kind people who are also brilliant researchers and capable administrators. Among the many whose valuable insights, professional advice and delightful company I enjoyed are Abi Adams, Chris Adam, Stefano Caria, Stefan Dercon, Lizzie Dipple, Cheryl Doss, James Fenske, Doug Gollin, Clément Imbert, Pramila Krishnan, Julien Labonne, Mahreen Mahmud, Anandi Mani, Karlijn Morsink, Rossa O'Keefe-O'Donovan, Kate Orkin, Rose Page, Natalie Quinn, Climent Quintana-Domeque, Amma Serwaah-Panin, and Gabriel Ulyssea.

Past and present DPhil students have provided friendship, thoughtful feedback as well as encouragement. They made the long journey much less lonely and much more fun. My gratitude goes to Rachel Cassidy, Claire Cullen, Lukas Hensel, Michael Kölle, Jonathan Lain, Pavel Luengas Sierra, Viviana Perego, Ashley Pople, Emma Riley, Chris Roth, Victoria Schreitter, Daniela Scur, Katherine Stapleton, Erkki Vihriälä, Verena Wiedemann, Marc Witte, and many others.

This work would not have been possible without the excellent field team in Bangladesh. I would like to thank Taslima Khanam, Leigha Miyata, Mir Rafiat Salman, Jesmin Sultana, Anaise Williams, and the entire project teams. I am grateful for the cooperation with the BetterWork programme of the International Finance Corporation and the International Labour Organization, the participating factories, and all respondents who gave their time and shared their information.

The financial support I received enabled me to undertake this DPhil and the research it entailed. I am very thankful for funding from the Private Enterprise Development in Low-income Countries research initiative, the Economic Development and Institutions research programme, the International Finance Corporation, the Urban Services Initiative, the Department of Economics, Hertford College, the Studienstiftung des deutschen Volkes, and

the Deutscher Akademischer Austauschdienst.

Hertford College and Balliol College have welcomed me warmly, enabled many exciting interdisciplinary discussions, and provided the opportunity to meet inspiring individuals from all walks of life. I will cherish the friendships I have forged here for years to come.

A special thank you goes to Julia Hill, whose mentorship greatly contributed towards my sanity during my time in Oxford. Viola Lutz, Marie-Louise Arlt, and Sara Martin lovingly ensured that I look beyond the bubble, asked the right questions, and provided a shoulder when needed.

I would not be where I am today without my parents, Manfred Uckat and Christiane Uckat-Erley, and my brother, Benedikt Uckat. I thank them for supporting me wherever they could, for reminding me of my roots, and for instilling in me the desire to make the world a better place. Together with more recent additions to the family – Kamila Uckat, Noah Uckat, Nina Stroka, Gottfried Barth, Leonie van der Valk, Peter Davies, Dylan Davies, Laura van den Berg – they have been a well of great comfort and joy.

Finally, Elwyn Davies has been an unwavering source of love and encouragement during these years – through the highs and the lows, from near and far. My gratitude goes far beyond what I can express in these lines. I could not wish for a better partner in life, and I look forward to finally not having an ocean between us!

Contents

Introduction

Firms are crucial for economic growth and development. Firms use inputs and ideas to produce goods and services that final consumers and other firms value, but they do not operate in isolation. Businesses often depend on services and infrastructure that many governments provide, and the wages and salaries that firms pay for labour are important determinants of their employees' living standards. These employees, many of them spending the majority of their days at the business premises, bring their skills, attitudes and beliefs into the workplace, which may in turn be influenced by the practices they observe in the firm.

These factors have significant influence on firm development and growth, though they are not all easily captured in a standard production function. Indeed, many firms report major constraints to their operations both from within and from outside the firm, especially in low- and middle-income countries (Bloom et al., 2014a; Gibbons and Henderson, 2013). This thesis investigates three constraints on firms within the context of South Asia. It does so in three separate chapters, whose main conclusions and contributions I outline here in broad strokes.

Chapter 1 focuses on constraints related to gender. Women continue to be underrepresented in labour markets worldwide, especially at the level of managerial positions (The World Bank, 2012). While women's equity is often considered a policy objective in its own right, recent evidence suggests that this underrepresentation may also entail substantial costs for firms (Flabbi et al., 2019). Chapter 1 investigates the interrelation between the firm and the household, by analysing the effects on intra-household bargaining if constraints to female career advancement are relaxed. Using quasi-random variation created by a promotion programme that aims to promote women to supervisors in Bangladeshi

garment factories, I measure impacts on both the households of promoted women as well as on households of the female subordinates that they supervise. The garment industry is the main source of formal wage employment for women in the country, but their opportunities continue to be limited to the non-managerial level (Menzel and Woodruff, 2019).

Different estimation approaches indicate that women's bargaining power at home increases as a result of female career advancement. For women who participated in the promotion programme, the effects are largest for the share of income that households spend on assignable goods for female household members and on remittances. These direct effects are amplified by indirect effects on their subordinates. Women who were exposed to the new female managers have more say in decision-making in the household, especially about their own mobility. I find evidence that both the direct and the indirect impacts are driven by women gaining confidence to get involved in bargaining, rather than income effects that ease the budget constraint or changes in the relative wage in the household.

This chapter adds to a vast literature on the determinants of women's bargaining power and the relationship between women's work and the household. It makes two novel contributions. The chapter provides the first evidence in the literature of the effect on intra-household bargaining if a woman gets promoted. It also contributes the first analysis of the impact of female role models in the workplace on the household. The results underline that there is a complementarity between removing women's constraints in the workplace and in the household.

Chapter 2, written with Christopher Woodruff, focuses on managerial skills and business practices for promotions. Practices of human resources management are crucial for firm performance and are especially consequential when deciding about promotions (Benson et al., 2019; Bloom et al., 2014b; Bandiera et al., 2020). Managers typically carry out more diverse and complex tasks than non-managerial workers, which often require a different skill set. In making promotion decisions, the challenge for firms is to either identify candidates that already have the requisite skills, or to try to mould new managers through instruction. Chapter 2 aims to understand the importance of manager selection and train-

ing in the context of promoting women to supervisors in Bangladesh's garment industry. It reports on two field experiments that were part of the promotion programme mentioned above. The first is an intervention which allows firms to base their selection of female candidates for promotions on formal measures of their technical and soft skills, and the second is an experiment which trained the selected female candidates in these same skills.

We present four findings. First, without formal diagnostic tests, factories' nominations for the promotion training are orthogonal to candidates' initial skills. Second, factories that are provided information about their candidates' skills select female trainees who have higher baseline skills, especially soft skills such as confidence and ambition. Third, these initial skills – and again especially soft skills – are predictive for gains in skills from the programme, official promotion rates, wellbeing of subordinates, and (to some degree) better performance. Fourth, training in the same soft and hard skills leads to improvements in some of these same outcomes, though the results are not always significant. In none of our results do we find that training in technical skills has an additional effect when compared to a training only targeting soft skills.

This chapter makes two contributions to the literature on worker selection and training, as well as female leadership. It represents the first experimental evidence on manager selection in the literature. In addition, the training intervention allows us to separately identify the impacts of soft skills and technical skills on managerial success. The results underline the importance of soft skills for female career advancement in a context where women have long been excluded from even low-level management positions. More broadly, the findings suggest the need to better understand and measure skills and characteristics which are predictive of managerial performance, in order to improve business practices for promotions.

Finally, Chapter 3 addresses infrastructure constraints. Especially in low- and middle-income countries, a lack of well-functioning systems of transport, water and communications represents a significant constraint to firm growth and productivity (see e.g. Bigsten and Söderborn, 2006; Gollin and Rogerson, 2014; The World Bank, 2004). In particular, improvements in transportation infrastructure, which are the focus of this chapter, can

reasonably be expected to decrease firms' shipping costs. However, the improved market access this entails can also lead to general equilibrium effects through entry and exit of firms, productivity improvements in existing firms, or a reallocation of resources between firms (Yang, 2018). The chapter focuses on the last two mechanisms, by analysing the short-term effects of a large-scale highway improvement programme in India – the Golden Quadrilateral project – on incumbent manufacturing firms. I estimate the impacts on firms that were incidentally treated with improved infrastructure using a difference-in-differences approach.

The results suggest very heterogeneous impacts. Overall, the incumbent firms shrink in terms of their input usage. These findings are mainly driven by firms that were initially less profitable, for whom sales and profits also decreased. In contrast, initially more profitable seem to grow their sales, profits and labour productivity, though not significantly so. The existing evidence on this infrastructure intervention has found significant entry of new firms, growth among established incumbents, and improvements in allocative efficiency. To this literature, this chapter adds a snapshot of the process of reallocation of resources towards more successful firms when areas become better connected.

Viewed together, this thesis contributes to our understanding of the constraints firms face, and it emphasises that firms do not operate in a vacuum. Firm operations and development are affected by the actions of other firms, households, and governments – and vice versa. These interrelations between firms, households, and governments are crucial to further our comprehension of private sector development. This has become especially salient at this point in time as the world seeks to address the fallout of the global COVID-19 pandemic.

The next three chapters present the essays just outlined in detail, which are written as self-contained journal-style articles. Each chapter presents its respective contributions to the literature, methodology, data description, results, and a discussion. They are followed by concluding remarks that present thoughts on further research directions.

Chapter 1

Leaning in at Home: Women's Promotions and Intra-household Bargaining in Bangladesh

Abstract This paper investigates how women's promotions in the workplace affect bargaining in the household. I exploit the design of a promotion programme for women in 27 Bangladeshi garment factories, by comparing women who were quasi-randomly selected for the programme to the shortlisted runners-up. Results using three different estimation approaches (OLS with post-double selection Lasso, regression discontinuity, and matching) show that women's bargaining power increases as a result of the promotion. The effects are largest for the share of income households spend on assignable goods for women (especially clothing and accessories) and remittances. The latter appears to mask expenditures on children, since remittances increase most for women whose children live with other relatives. I find that these direct effects of the promotions are amplified by impacts on women working as subordinates of the new female managers. Using the quasi-random assignment of sewing-line operators to production lines for identification, I observe that women exposed to a female manager have more say in decision-making in the household, especially about their own mobility. Overall, I find suggestive evidence that both the direct and the indirect impacts are driven by women gaining confidence to get involved in bargaining, rather than income effects that ease the budget constraint or changes in the relative wage in the household.

1.1 Introduction

Women's leadership in politics and business has received a lot of attention in the last two decades. To improve dismal participation rates, in 2019 more than half of all countries in the world had some type of political quota for women, and 28 countries had introduced a quota on company boards.[1] Despite these efforts, progress has been slow: in Asia, women even now represent only 19.7 % of members of parliament and 12.8 % of corporate board members.[2] Policies that encourage female leadership typically have at least two goals: to improve the participation of a previously under-represented group, and to achieve alleged instrumental benefits on outcomes, such as increased effectiveness of public spending, or higher profits. While the evidence for both goals generally points to more successful interventions in the political compared to the business sphere,[3] one aspect has largely been missing in the discussion: How does female leadership impact on bargaining in the household, both for the pioneering female leaders and the women who are exposed to them?

The literature has mostly been mute on this issue, but there is more extensive work analysing the effects of women entering employment for the first time or changes in income. In the theoretical literature, standard collective models of intra-household bargaining predict that women's potential earnings outside the marriage, her — potentially endogenously determined — labour supply and favourable changes in the relative wage will increase her bargaining power (Browning et al., 2014; Basu, 2006).[4] Empirical investigations mirror this picture. As reviewed in Heath and Jayachandran (2018) and Duflo (2012), most empirical studies have found positive impacts of women's employment on their bargaining power as well as final household outcomes, such as children's health and education. The only paper focusing on the effects of promotions on the households

[1] International Institute for Democracy and Electoral Assistance at https://www.idea.int/data-tools/data/gender-quotas/quotas and Corporate Women Directors International at https://globewomen.org/CWDINet/wp-content/uploads/2019/07/Quota-Chart-June-2019.pdf (both last accessed 11 November 2019)

[2] Inter-Parliamentary Union at http://archive.ipu.org/wmn-e/world.htm and and Corporate Women Directors International at https://globewomen.org/CWDINet/wp-content/uploads/2017/07/1.jpg (both last accessed 11 November 2019). Note that the percentage of corporate board members refers to the Asia-Pacific region.

[3] See Miller (2018) and Pande and Ford (2011) for reviews.

[4] Some theoretical extensions, however, caution that female employment has the potential to lead to conflict and violence in the household (Eswaran and Malhotra, 2011; Ramos, 2018).

of the promoted, Folke and Rickne (2020), provides a cautionary tale. As I will further discuss below, the authors find that top-level promotions in Sweden double divorce rates for promoted women.

I analyse the impact of female leadership on intra-household bargaining in the context of the garment sector in Bangladesh. The work the sector provides is similar to many labour-intensive manufacturing industries in low- or middle-income countries. The garment industry is the main source of formal wage employment for Bangladeshi women, in a country where women traditionally did not work outside the home (Asadullah and Wahhaj, 2016). Nevertheless, the positions available to women in the industry continue to be limited to the non-managerial level. Even though about 80 % of production-line workers in the sewing sections are typically female, they represent only about 5 % of line supervisors, the lowest-level management position (Macchiavello et al., 2020; Menzel and Woodruff, 2019). The production-line operators each perform one sewing operation repeatedly throughout the work day, before passing the garment on to the next machine operator in the production line. Production lines are usually managed by one or several line supervisors, who communicate with workers and higher-level managers to ensure that the line is meeting its target. This paper studies the impacts on intra-household bargaining if a woman gets promoted from production-line operator to line supervisor.

The ideal empirical scenario to analyse this research question would be by randomly assigning promotions. However, this is difficult to implement in such a highly competitive and high-pressure environment in the private sector. Profit-maximising firms are reluctant to allow their promotion offers to be determined by chance. Instead, I exploit variation created by a promotion programme for about 200 women in 27 garment factories. Chapter 2 describes the programme in detail and analyses its effect on factory outcomes. As part of the promotion programme, the participating female production-line operators trialled as line supervisors for at least two months, were trained for this role and, if successful, were officially promoted. All participating women therefore gained experience in a management position whose leadership responsibilities were very different from the work they did before, even if not all of them were officially promoted.

Drawing intuition from a simple collective household framework following Browning et al. (2014), I posit that a woman's promotion could affect the bargaining process through three potential channels. First, if a woman's promotion is accompanied by an increase in her income, this eases the household's budget constraint, ceteris paribus. I call this a "Marshallian income effect". Second, a change in the woman's income can, ceteris paribus, affect the relative wage of the individuals in the household. This "relative wage effect" could affect the woman's weight in intra-household bargaining irrespective of the effect on total household income (i.e. as a distribution factor). Lastly, the experience in a leadership position without any change in income or in the relative wage could also improve the women's bargaining position in the household as a distribution factor, which I call a "position effect". This could, for example, be the case if the new position gives the woman more confidence to get involved in bargaining in the household, or if she can translate the bargaining skills learned in the new position to the household. I will return to this below when discussing results.[5]

To study the impact on the household, I conducted two waves of extensive surveys and incentivised economic games with the women involved in the programme and with their husbands, or another adult household member if the woman was unmarried. Because the vast majorities of respondents were married couples, for simplicity I refer to them as "wife" and "husband" in the rest of the paper. The respondents were interviewed simultaneously but separately at their homes. My main outcomes capture different aspects of bargaining power and household decision-making, and consist of the household's income share spent on assignable goods for women, the income share of remittances, recent experiences of domestic violence, the women's involvement in decision-making, as well as economic games inspired by Almås et al. (2018).[6] I also conducted additional high-frequency phone surveys with the wives for statistical power.

I compare outcomes of women who were selected for the programme to the nominated and shortlisted runners-up, exploiting the selection process for the programme as a quasi-

[5]Note that, as I explain Section 1.3, supervisors only work about 20 minutes per day longer than production-line operators. Therefore, it is unlikely that a promotion affects the household via less time spent at home.

[6]For funding reasons, the home surveys took place after the conclusion of the promotion programme.

experiment. At the start of this process, we determined the number of new supervisors to be selected in each factory together with the respective management. This factory-specific cutoff was set in such a way that the factory would be able to absorb all selected operators as new supervisors within a few months. Factories then provided us with a ranked list of female production-line operators as candidates, who had to fulfill pre-set eligibility criteria. Our team further excluded individuals that failed diagnostic literacy and numeracy tests. These extensive eligibility checks meant that only highly skilled and experienced operators with a supportive family were included in the shortlist for the programme. Half of the factories were randomly assigned to a two-hour selection experiment (discussed in detail in Chapter 2), which allowed them to change their ranking of shortlisted operators once after reviewing their diagnostic scores. Out of the eligible and shortlisted operators, only those with a rank that was better than the factory-specific cutoff were selected for the promotion programme. The comparison group consists of the nominated and shortlisted women who passed all diagnostic checks but continued to work as production-line operators.

I argue that the assignment to the selected group is quasi-random because factories have limited ability to predict which women would become successful supervisors among the shortlisted group of women. Because the factories had historically nearly always promoted men, they had large numbers of skilled and experienced female line operators from which to choose new supervisors. The eligibility criteria and diagnostic tests further ensured that the considered line operators were a homogenous sample. While it might be easy for factories to identify operators who would *not* make a good supervisor in their factories, it is much harder to decide the ranking among a sample of women who had *all* been judged to be suitable after an extensive screening. This argument is similar in spirit to the literature on microenterprises, recruitment and venture capitalism, which shows that humans are often inept at predicting which firms, workers or investments will be successful among an already select sample (McKenzie and Sansone, 2017; Hoffman et al., 2018; Nanda, 2016). Since the management was interested in identifying the best supervisors for the factory, it is especially unlikely that they ranked the eligible operators based on their potential

household outcomes, which are the focus of this paper. I conduct extensive checks comparing the two groups, consisting of t-tests, tests of joint orthogonality and normalised differences following Imbens (2015), for which I use a wide range of variables from the promotion programme baseline. All of these tests show that that two groups do not differ on observable criteria and provide strong support for quasi-random assignment.

I use three estimation approaches: a machine-learning technique (the Post-Double Selection Least Absolute Shrinkage and Selection Operator, PDS Lasso) to choose predictive control variables for OLS (Belloni et al., 2013, 2014a,b, 2016), a regression discontinuity approach using the operator rankings, and a matching method. They all lead to the same conclusions. I find that the promotion programme was successful in advancing women's careers. All selected women were assigned to work as supervisors for at least two months, and therefore gained experience in a leadership position. Within the first year after the promotion programme start, selected women were 24 percentage points more likely to have been officially promoted to supervisor compared to the shortlisted comparison group, and more were still waiting for an official promotion at the time of the surveys.[7] For the promoted women, personal income increased by about a third compared to the shortlisted group.

When investigating household outcomes, I find that the women who were selected for the promotion programme enjoy significantly higher bargaining power than the comparison group. The effect is largest on the share of income that households spend on assignable goods for women, especially clothing and accessories, as well as the income share of remittances. When using PDS Lasso to analyse effects, the share of household income spent on women's goods increases to about 8 %, compared to 5.3 % in the comparison group. I cannot entirely exclude that part of this effect might be driven by the need to look more professional in the new managerial position, but argue that increasing effect sizes over time make this improbable. This is also unlikely to be the reason for the large effects I find on remittances, which increase from 6 % of household income to about 9.5 %. This large

[7]Note that ten nominees from the shortlisted comparison group joined the promotion programme as replacements for women from the selected group who dropped out of the programme. Four of these replacement nominees reported that they had been officially promoted during the household surveys.

effect appears to result primarily from transfers to family members who take care of the couple's children living outside the household. In this sample, about half of all children live with relatives outside the household. The large increase in remittances hence seems to mask expenditures on children. These results are confirmed when I implement a regression discontinuity approach relying on the ranking of the shortlisted operators, as well as a matching method that uses variable from the promotion programme baseline. In addition, the findings are robust to, amongst other checks, different specifications, different variable definitions, and different data sources (i.e. the husband's survey and high-frequency phone surveys). I also show that the effects do not primarily result from an effect of the training that was part of the promotion programme, or from a potential re-ranking of candidates as part of the selection experiment.

I then investigate which, if any, of the potential three channels identified in the collective household framework are driving the results. The sample size makes it difficult to make definitive statements, but the different pieces of evidence point to the same conclusions. I show that a Marshallian income effect is unlikely for two reasons. First, I do not find an effect on unmarried women living on their own. For this group, an income effect would be expected to operate, but a bargaining effect is excluded by definition. Second, I do not see effects on households' expenditures on assignable goods for men and boys, but indications of a positive effect on girls' assignable expenditures. This suggest that a reallocation of expenditures towards female assignable goods has taken place, rather than a Marshallian income effect. I also present evidence that a relative wage effect is improbable, by showing that the impacts are not larger for selected women who started earning more than their husbands. I find most evidence in favour of the leadership position effect driving the results. I see strong positive impacts on women's expenditures and remittances even for women who were not yet officially promoted at the time of the household surveys. These women still had experience in a leadership position during the promotion programme, but did not see their income increase. I observe some evidence that this position effect operates by improving the women's attitudes and beliefs about themselves, which could enable them to get more involved in decisions in the household. These results suggest that

the promotion programme increased the women's bargaining power in the household, if one is willing to assume that women have a higher preference than their partners for female assignable goods expenditures and for supporting family outside the household.

While this is encouraging evidence for policy approaches that aim to increase women's participation in leadership, the direct effects of promotions will only affect the small share of the female population that attains a management position. In the next step, I therefore analyse whether there are any indirect effects of female leadership on women working under a female manager. These effects would potentially be relevant for a much larger number of women. I compare outcomes of female sewing-line operators who were working on lines to which the newly selected female supervisors were assigned for the promotion programme (the "exposed" group) to outcomes of women who were working on other lines with male supervisors (the "non-exposed" group). According to industry experts, the assignment of operators to production lines is first and foremost determined by production requirements. It is especially unlikely that workers can sort to lines of new supervisors in their trial period, since factories use these periods to observe the performance of potential new supervisors with the resources that they have been given. Several tests indeed indicate that the two groups I compare are very similar. These two groups are balanced on time-invariant observable characteristics, and the observable characteristics are not jointly predictive of being assigned to a line with a female manager. I again use PDS Lasso to identify control variables for OLS in my estimation approach, but find that the control variables hardly matter.

Having interviewed about 500 exposed and non-exposed wives and their husbands about nine months after the start of the exposure to female supervisors, I find that the exposed women are considerably more involved in decision-making in the household, especially regarding their own mobility. I see no changes in expenditure outcomes, but some statistically insignificant suggestions of increased conflict in the household. I again evaluate the potential channels identified in the collective household framework, and do not find any suggestions of a Marshallian income or a relative wage effect. There is some tentative evidence that the increased involvement in decision-making is resulting from women be-

ing more willing or able to get involved in negotiations in the household. The observed increase of women's involvement in decision-making is again robust to a wide range of checks, though I find a much smaller effect when looking at the husbands' data.

Taking the results together, this paper therefore provides evidence that taking a step up on the career ladder can lead to increased bargaining power for promoted women at home, irrespective of potential income effects or relative wage effects this may bring. Moreover, I see that these direct effects of a leadership position are amplified by impacts on women who work as subordinates of the new female managers, for whom involvement in decision-making in the household increases. The evidence suggests that women's promotions increase women's confidence to get involved in intra-household bargaining, both for female managers as well as the female subordinates working under them. These results underline that there is a complementarity between women's power in the workplace and in the household.

The paper is structured as follows. Section 1.2 discusses this paper's contribution to the literature. I provide an overview of the context in Section 1.3, and explain the conceptual framework and the study design in Section 1.4. Section 1.5 presents the empirical strategy and results for the analysis of the effects on the participants of the promotion programme, which I call the "promotion analysis". Section 1.6 does the same for the analysis of role model effects, which I call the "exposure analysis". Section 1.7 concludes.

1.2 Contribution to the literature

With this analysis, I contribute to the vast literature that studies the determinants of and constraints to women's bargaining power and women's empowerment more broadly.[8] I contribute to two strands of this literature which are most closely related. First, I provide the first evidence about the effect of women's promotions on bargaining in the household. In low- and middle-income countries, the empirical literature investigating how women's work impacts on bargaining power has exclusively focused on the effects of women entering employment for the first time, income increases, or general shocks to female-specific

[8] See, for example, Doss (2013) for a review.

labour demand. Most of these studies have found that better labour market opportunities for women have a positive impact on a range of measures of women's bargaining power, such as domestic violence, expenditures, involvement in decision-making, mobility or women's time use (e.g. Atkin (2009) and Majlesi (2016) for Mexico, Dharmalingam and Morgan (1996), Rahman and Rao (2004), and Luke and Munshi (2011) for India, Getahun and Villanger (2018) for Ethiopia, and Friedemann-Sanchez (2006) for Colombia). However, there are some exceptions. For example, Jensen (2012) finds no impact on mothers' involvement in decision-making or autonomy from increased labour market opportunities for their daughters in India, but he does find that the daughters delayed marriage and childbearing, and had higher labour market participation and more schooling. Also from India, McKelway (2019) reports that employment enabled women to make more independent decisions, but did not increase bargaining power in joint decisions. Blattman and Dercon (2018) find no impact of offering factory work to women in Ethiopia on their autonomy, though this might be due to a lack of statistical power.

In this strand, most closely related to this paper is the literature that analyses the relationship between work and women's bargaining power in Bangladesh. Anderson and Eswaran (2009) and Kabeer et al. (2018) both find that work, and particularly formal work outside of the household, is positively related to women's autonomy. Regarding the Bangladeshi garment industry in particular, the Asian Center for Development (2015) reports that "93% of the [surveyed female] workers perceive that this job has given them a voice in family decisions and raised their importance as a family member". These results are supported by Heath and Mobarak (2015), who find that women exposed to the garment sector delay marriage and childbirth. The authors argue that this is due both to increased enrolment of younger girls and a higher likelihood of older girls to be employed. Nevertheless, there is some evidence which indicates that the effects of work on bargaining power may depend on the women's initial circumstances. Heath (2014) shows that women with an initial low bargaining power (proxied by low education and low age at marriage) report higher domestic violence upon entering work, but does not find such an increase for women with higher bargaining power to begin with. This paper

adds to this literature in low- and middle-income countries by providing the first evidence about the impacts of a promotion on the household. It thus goes beyond the impact of women taking up work, and instead investigates how the household responds if women climb up the first step of the career ladder. I also provide evidence on potential channels related to a promotion: a leadership position effect, a Marshallian income increase and a change in the relative wage.

In high-income countries, the literature has paid more attention to women's promotions, but the studies have mainly focused on women entering top-level corporate roles and the resulting impact on firm outcomes (See e.g. Bertrand et al. (2019) and Dale-Olsen et al. (2013) analysing Norway's quota for women on company boards, Flabbi et al. (2019) for Italy, Marinova et al. (2016) for Denmark and the Netherlands, Deszö and Ross (2012) for the USA and Sabatier (2015) for France). One exception, and closest to this paper, is Folke and Rickne (2020). The authors analyse the impact of women's and men's promotions to top political and corporate positions on divorce rates in Sweden. They find that a promotion doubles divorce rates for promoted women, but not for men. These effects are particularly pronounced for couples with traditional gender roles in the early phases of the relationship. The authors provide evidence that these higher divorce rates could result from a stressful re-negotiation of household tasks after the women's promotion or from a violation of social norms if the wife becomes the dominant earner. In comparison to Folke and Rickne (2020), this paper focuses on women's promotions to entry-level management positions in a low-income context. And while Folke and Rickne (2020) exclusively discuss divorce, i.e. household breakdown, this paper focuses on a wide range of intra-household bargaining outcomes within continuing marriages.

As a second contribution, this paper presents the first evidence of the impact of female role models in the workplace on intra-household bargaining. The literature so far has primarily focused on female role models in the political sphere, in teaching environments or in the media, and has shown that female role models can be instrumental in changing norms as well as behaviour. For example, Beaman et al. (2009, 2012) exploit the random assignment of women's quotas to Indian village councils to investigate the impact of exposure to

female political leaders on perceptions and girls' outcomes. The authors find that villages which were exposed to female politicians in two election cycles had weaker gender stereotypes and perceived women as more effective leaders (Beaman et al., 2009). In addition, the exposure erased the gap in educational attainment for girls in these villages, increased their and their parents' aspirations for them, and decreased the time girls spent on household chores (Beaman et al., 2012). The literature from teaching environments points to similar effects, with many studies finding that female teachers increase female students' performance as well as their career aspirations and choices (Carrell et al., 2010; Dee, 2005; Bettinger and Long, 2005; Asgari and Dasgupta, 2010; Antecol et al., 2014).

While none of these analyses from the political and educational environments investigate the impacts on measures of women's bargaining power, the literature on media exposure has demonstrated the potentially large effects that role models can have on these outcomes.[9] In a seminal paper, Jensen and Oster (2009) show that the staggered introduction of cable television across five Indian states had large impacts on outcomes for women living in rural areas. Within the first year of the arrival of television, the authors find a lower acceptability of domestic violence, diminished son preference, and increases in women's autonomy. These changes also translate into a reduction in fertility as well as increases in school enrolment for young children. In a similar vein, Chong and La Ferrara (2009) and La Ferrara et al. (2012) demonstrate that soap operas depicting modern women and relationships reduced fertility and increased rates of separation and divorce in Brazil.

In comparison, this paper investigates the impact of a female role model in the workplace — a female manager on the production line under whom the women work for at least ten hours a day for six days a week in the factory. This exposure to a female supervisor is therefore arguably more direct compared to female political leaders, as well as more direct and more sustained compared to television characters. It most resembles the type of exposure found in teaching environments, but affects the exposed women at a different point in their economic lives. Macchiavello et al. (2020), also working in Bangladesh's garment industry, provide suggestive evidence that having a female line supervisor has

[9] See DellaVigna and La Ferrara (2015) for an excellent review of the economic and social impacts of media, and La Ferrara (2016) for a review focusing on the potential of the media to further developmental objectives.

positive, though insignificant, effects on female operators' career aspirations. I extend this analysis and investigate whether having a female manager in the workplace changes household bargaining for exposed subordinates.

1.3 Context: Women in Bangladesh's garment industry

The garment industry in Bangladesh experienced explosive growth from the 1980s onwards. It currently represents more than 80 % of Bangladesh's exports, consists of at least 4,500 factories and employs more than 4 million workers, according to the sector's largest trade association, the Bangladesh Garment Manufacturers and Exporters Association (BGMEA).[10] Figures 1.B.1 and 1.B.2 in the appendix illustrate this impressive development.[11]

The production in garment factories is usually divided into at least three sections: cutting, sewing, and finishing.[12] After the fabric is cut into the pieces required for a garment in the cutting section (e.g. a front and back piece, two sleeves and neckbinding for a simple T-shirt), the fabric pieces are sewn together to create the garment on the sewing floors, before being passed on to the finishing section for ironing and packaging. The sewing sections employ the majority of workers. They are almost always divided into production lines, which take the cut fabric pieces as inputs and output the complete garment. These sewing lines consist of 20-80 workers, depending on the garment produced, who each perform one sewing operation (e.g. binding the neck for a T-shirt) at a workstation. Lines are typically composed of line operators of different skill levels and helpers doing auxiliary tasks. The production lines are usually managed by at least one line supervisor, several if the line is

[10] According to the BGMEA, its members account account for 100 % of woven garment exports, more than 95 % of sweater exports, and around half of light knitwear exports. (http://www.bgmea.com.bd/home/about, last accessed August 7, 2019)

[11] Note that the number of BGMEA member factories dropped by more than a quarter in 2013-2014. This coincides with the period after the collapse of the Rana Plaza factory in April 2013, during which more than a thousand workers died. In the aftermath of the collapse, the Bangladeshi government, the BGMEA and buyer organisations intensified inspections and closed down some factories, but it is unclear whether this alone explains the large drop. It is also possible that factories deregistered to avoid scrutiny or that non-operating factories were purged from lists following inspections. For more information on the response to the Rana Plaza disaster, see Boudreau (2019).

[12] Some factories have additional sections, such as those that spin thread from the raw materials, knit fabric from the thread, dye the fabric, add embroidery to the fabric, or create stonewash in jeans. See Menzel and Woodruff (2019) for a more extensive discussion of the production process, and the differences in grades among sewing operators.

longer.[13]

The garment industry is the main sector in which Bangladeshi women can find formal wage employment, in a country in which women have traditionally not worked outside the home (Asadullah and Wahhaj, 2016). While statistics from a representative sample of factories are hard to come by, it is estimated that 56-65 % of all workers in the sector are female (Labowitz and Baumann-Pauly, 2015; Asian Center for Development, 2015). However, the opportunities for women in the industry still remain mostly limited to non-managerial positions. Among the line operators and helpers in the sewing section, about 80 % are typically female, but only about 5 % of line supervisors are women (Macchiavello et al., 2020; Menzel and Woodruff, 2019).

The impact of a woman's promotion from line operator to line supervisor on the household of the female supervisor herself, and the female workers she manages, is the focus of this paper. It is therefore important to understand that the work of a line supervisor is very different to that of operators. Whereas line operators sit at their workstations and repeatedly perform their step of the sewing process, line supervisors are tasked with managing workers to ensure that their (part of the) line is running well. They are typically on their feet, and communicate with workers to motivate them to reach the target set for the line, solve problems with machines or inputs, or explain processes to operators. They also communicate any issues, such as worker absenteeism or machine problems that require technicians, to higher-level managers.

Compared to line operators, supervisors work only slightly longer hours, to set up the line at the beginning of the shift and wrap up after production has finished for the day. When comparing the hours worked reported by line supervisors and by line operators in the part of my sample that is unaffected by the promotion programme, this difference is only about 11 minutes when controlling for factory and month fixed effects, and about 19

[13]The hierarchy levels above line supervisors in the production department vary with the size of the factories. Larger factories often employ line chiefs responsible for several lines, floor managers responsible for one sewing floor, assistant production managers responsible for one unit, and production managers who report to the general manager. In smaller factories, which might have only few floors and only one unit, there are typically fewer levels of hierarchy.

minutes without these additional controls.[14] Since line operators in my sample work on average 10 hours per day, line supervisors' work hours are only about 2-3 % longer. Officially promoted line supervisors also earn more than operators. In the sample of factories included in this paper, the highest-skilled operators earn approximately 10,000 Taka per month (about 100 GBP and 125 USD at the time of the surveys).[15] Established supervisors on average earn approximately 14,000 Taka per month (about 140 GBP and 175 USD). Note, however, that the income increase for newly promoted supervisors is smaller, as I find below.

It is important to note that women in the garment industry already possess a higher level of bargaining power than one would find for the average Bangladeshi woman. Traditionally, women in Bangladesh were expected to tend to the homestead, and keep out of the public sphere (Asadullah and Wahhaj, 2016). There has been impressive progress towards gender equality, especially in education and health, in the decades coinciding with the growth of the garment industry.[16] As discussed in Section 1.2, it has widely been argued that work in the sector has enabled women to have more say in decision-making in the household and to be more visible in the public sphere. Since this paper is focused exclusively on women already working full-time outside the home in the garment industry, it is not immediately obvious whether women's promotions to managerial positions can a priori be expected to have an additional positive effect on these women's bargaining power in the household, considering the high initial bargaining power.

1.4 Framework and study design

1.4.1 Intuition from a collective household framework

I illustrate the intuition for the potential effects of a woman's promotion on intra-household bargaining using a simple collective household framework, following

[14] I focus on the part of the sample whose work hours, conditional on working, are likely to be unaffected by the promotion programme. I therefore exclude the nominees and selected women for the promotion programme, as well as the women working with the selected women as supervisors. Note that this is not a representative sample, but serves as an illustration.

[15] These numbers represent net pay, which includes basic pay, overtime pay, and all allowances, minus any deductions, e.g. for absences.

[16] See e.g. The World Bank (2008) for a review of this progress and the drivers.

Browning et al. (2014). Such a collective approach assumes that a household makes Pareto efficient decisions and that it follows a stable decision-making process. In addition, I assume that the individuals' utility functions are strictly concave. This allows me to write the optimisation programme as follows, for a household with individuals a and b that chooses consumption of private good q of price p and public good Q of price P, with income m depending on wages w,[17]

$$\begin{aligned} \max_{q_a,q_b,Q} \quad & \mu(P,p,\frac{w_a}{w_b},promotion,\boldsymbol{z})\cdot u_a(q_a,Q)+u_b(q_b,Q) \\ \text{s.t.} \quad & p\cdot(q_a+q_b)+P\cdot Q=m(w_a,w_b). \end{aligned} \tag{1.1}$$

The first line of this programme shows that the household maximises a weighted sum of the individuals' utility functions, where the individual utility functions $u_s(q_s,Q)$ for $s=\{a,b\}$ depend on the individual's private consumption of q as well as the level of the public good Q. The public good, for example, could be thought of as expenditures on children. The weight $\mu(.)$ is typically called the bargaining or Pareto weight, and represents the weight given to a's utility in the household programme. The bargaining weight is a function of the economic environment (e.g. prices) as well as potential distribution factors. Distribution factors are variables that can affect the decision process but do not affect the budget constraint or preferences, e.g. the relative wage of the individuals (once total income has been taken into account), or other factors collected in the vector $\boldsymbol{z}$. The second line of the programme represents a standard budget constraint, where I assume that expenditures on the private and public goods exhaust the household's income.

I posit that a woman's promotion could affect the bargaining process through three potential effects:

1. *Marshallian income effect*: If a woman is promoted with an increase in her wage rate w_a, this eases the household's budget constraint, $m(.)$. This would lead to an increase in the set of feasible allocations for the household.

[17] Of course, one can extend this simple framework, e.g. by allowing for altruistic preferences, introducing domestic production or labour supply decisions. Nevertheless, this simple framework is instructive to gain intuition for the potential effects of a promotion.

2. *Relative wage effect*: Ceteris paribus, a change in the woman's wage rate w_a affects the relative wage of the individuals, $\frac{w_a}{w_b}$. As a potential distribution factor, this could affect the woman's bargaining weight, $\mu(.)$.

3. *Position effect*: Irrespective of any change in total income or relative wages, a $promotion$ to a leadership position could be a distribution factor that affects the woman's Pareto weight, $\mu(.)$. This could take place, for example, if the promotion to a leadership position in the workplace affords the woman a more important role in decision-making in the household, or if she can translate bargaining skills learned in the new position to the household,[18] or if her new position gives her the confidence to get more involved in bargaining in the household.

In Section 1.5.4, I investigate which of these effects, if any, I observe in my data for the women who were part of the promotion programme.

This framework can also be applied to study the effects of women being exposed to the new female managers. Since the subordinates of the female supervisors continue to work in a line operator position, a large change in the wage rate working through a Marshallian income effect or a relative wage effect is possible but arguably less likely.[19] Instead of a position effect, I posit that the third potential effect for line operators exposed to the new female managers would instead read as:

3. *Exposure effect*: The $exposure$ to a female manager can be a distribution factor affecting the women's bargaining weight, $\mu(.)$. This could be the case, for example, if working with a female supervisor as a role model changes the exposed women's perception about a woman's permissible bargaining position in the household, or if the exposed women are able to emulate the management techniques implemented by the female supervisors at their homes.

In Section 1.6.4, I investigate which of these three effects, if any, I observe in my data for

[18] This would be in the spirit of Ashraf et al. (2020b), who find that teaching girls negotiation skills allows them to bargain for a better education.

[19] Line operators' wages typically include boni for good attendance and incentives based on production line productivity, which could both potentially be affected by the supervisor. However, these components typically constitute only a small share of line operators' income.

the women who were exposed to the new female managers.

1.4.2 Promotion programme

For the analysis, I exploit variation created by a promotion programme for women in the Bangladeshi garment industry, during which selected women worked as production line supervisors for at least two months. The impacts of the programme on factory outcomes are analysed in Chapter 2, where the programme is discussed in more detail. The way in which female production line operators were selected for the programme allows me to compare household outcomes of new female supervisors to outcomes of female operators who were short-listed for the programme but did not make the cut because of space constraints.

Only skilled and experienced operators were considered for the programme in which 27 large Bangladeshi garment factories participated. At the beginning of the programme, the factories were asked to provide a ranking of female operators that they wanted to consider for new supervisor positions. Those operators had to fulfil pre-set eligibility criteria in five areas, which are described in Panel A of Table 1.1. We conducted baseline surveys in the factories with all nominated women. We further implemented diagnostic tests with all candidates and excluded women who did not pass basic literacy and numeracy tests (see Panel B in Table 1.1). Only women who passed all screening criteria in Table 1.1 were included in the shortlist. These criteria mean that all short-listed women were skilled production line operators with significant experience in the garment industry and more than primary education. They were also interested in becoming a supervisor and had a family that supported this interest. In total, the 27 factories short-listed 243 eligible women.

In each factory, only the top-ranked candidates in the shortlist were allowed to participate in the promotion programme. We determined a factory-specific cutoff in advance in collaboration with each factory management team. The goal was to only include as many participants in the promotion programme as the factory would be able to absorb as new supervisors in a few months, such that selected operators had a good chance of officially being promoted during the programme duration. Half of the factories were randomly as-

Table 1.1: Eligibility criteria for promotion programme

Area	Criterion
Panel A: Selection criteria for factories	
Education	$\geq$ 8 years
Experience in garment industry	$\geq$ 2 years
Operator grade	Operator (grade 4) or Senior operator (grade 3)
Interest in supervisor position	Yes
Supportive family	Yes
Panel B: Exclusion criteria using diagnostic tests	
Literacy	Passed basic test
Numeracy	Passed basic test

signed to a two-hour "selection experiment", which is discussed in detail in Chapter 2. These factories were able to change their ranking of shortlisted operators once, after reviewing the diagnostic scores for all tested candidates. Nominees in the other half of the factories undertook the diagnostic tests, but the results were not revealed to the factory, and the original ranking could not be changed. For factories that were assigned to the selection experiment, the final shortlist is the last ranking after the selection experiment, after exclusion of ineligible operators and a potential re-ranking. For factories that did not participate in the selection experiment, the final shortlist coincides with the original ranking of operators, after ineligible nominees were excluded.[20]

Using these final shortlists of operators, only the women with a better rank than the factory-specific cutoff were assigned to the promotion programme. In this paper, these women are the "selected" group, consisting of 199 individuals. The women who passed all eligibility checks but whose rank was worse than the factory cutoff continued working as operators and form the comparison group. I call this group, which consists of 44 individuals, the "nominated" group. Only the selected group participated in the promotion programme, which entailed trialling as an assistant supervisor with gradually increasing responsibility for at least eight weeks. Factories were encouraged to officially promote

[20]To account for potential selection effects, I include all diagnostic scores and factory fixed effects in the vector of potential control variables, as discussed in Section 1.5.1. I also show that results hold if I limit the sample to the half of the factories that did not participate in the selection experiment in Section 1.5.3.

the selected women as soon as they felt that their quality warranted a promotion. This is similar to how factories promote new supervisors in the industry. Outside of any intervention, factories typically trial candidates for supervisor positions on the job. They first shadow existing supervisors, and then move to supervising a gradually increasing number of workers. If the factory is satisfied with their performance in managing a line or line section on their own, they receive an official promotion once a supervisor vacancy becomes available.

The selected women were assigned to different training regimes. These varied in their timing of two different modules, which were designed to train the selected women in soft and technical skills to make them better production line supervisors. The soft-skills module consisted of a four-day training in stress management, assertive communication and leadership skills. The hard-skills module provided an overview of the production process and training on how to plan, manage and solve problems on production lines. Appendix 1.A provides more details. One third of the selected women were assigned to receive both training modules immediately. Another third only received the soft skills component immediately and was trained in the technical skills about six months later. The remaining third received both training modules six months later. Remember, however, that all three groups were trialling as assistant supervisors during a two-months period. Figure 1.B.3 presents the timing in detail.

The impact of these trainings on factory outcomes is analysed in Chapter 2. For the main analysis of intra-household bargaining, I pool all women assigned to be participants of the promotion programme in the "selected" group, and compare them to the "nominated" group. In Section 1.5.4, I investigate potential training effects.

1.4.3 Survey implementation

To collect the outcome variables of interest for this paper, I implemented surveys with women and one of their household members at the respondents' homes. For funding reasons, these took place after the conclusion of the promotion programme discussed in the previous subsection. For the analysis of the effects on the promoted women, which I call

the "promotion analysis", the field team targeted all 243 women who were nominated and eligible for the promotion programme by the 27 factories.[21] This sample was interviewed twice in person at their homes. The first wave took place on average four months after the beginning of the promotion programme in the factories, and the second wave on average ten months after the beginning of the programme.[22] For power reasons, I also conducted biweekly phone surveys to collect additional measurements for a subset of outcomes of interest.

The sample for the analysis of role model effects on the women working with female supervisors, which I call the "exposure analysis", was determined during the follow-up surveys for the promotion programme in the factories. On each trial line to which a participant of the promotion programme had been assigned to trial as supervisor, female operators were sampled for the household surveys.[23] These are the so-called "exposed" operators. The trial lines had been chosen by the factories as needing a new supervisor before the start of the promotion programme. In addition, in each factory, the same number of comparison lines was sampled from all other lines not involved in the promotion programme, and from these lines, female operators were again sampled to be interviewed in the household. These latter operators are the "non-exposed" group, who are nearly exclusively supervised by men. Overall, 715 operators were sampled. The interviews for the exposure analysis took place on average nine months after the start of the promotion programme.[24]

The home surveys were implemented in two steps. First, the field team conducted short pre-surveys on the phone, using the mobile phone numbers collected during the factory visits. These pre-surveys were used to collect the respondents' location information and to

[21] In addition, we also surveyed 93 nominated and ineligible operators. Out of these, 52 nominees failed the literacy and numeracy tests, 16 were absent during the evaluation and 25 withdrew before treatment assignment. They are excluded from the analysis since they are arguably not comparable to the nominated and eligible women.

[22] For logistical reasons, the promotion programme was implemented sequentially. This ensured that class size for the training were manageable and that all trainees were taught by the same trainers. The household surveys followed this pattern. I therefore present the time line in reference to the beginning of the promotion programme in the factories. See Figure 1.B.3 for a detailed time line. The first wave of surveys took place between June and December 2017, the second wave between December 2017 and August 2018.

[23] The sampling of lines and operators was implemented using simple randomisation implemented by a computer script.

[24] These surveys were implemented between August 2017 and March 2018.

identify the second household member to be interviewed during the home survey.[25] Crucially, the pre-surveys were also used to agree on an interview time that was convenient for both respondents in the household, and to find an arrangement that would allow the respondents to be interviewed separately from each other in a private room where the interview could not be overheard.[26] This was especially important because the respondents' availability was limited due to the long working hours in the garment industry, and because a majority of respondents only had one room available (when excluding bathrooms and kitchens).

In the second step, a team of two enumerators visited the household at the agreed time and administered the surveys to the respondents in the household. For reasons of respondent and enumerator safety, a team always consisted of one female and one male enumerator. The woman known to us from the factory was always interviewed by the female enumerator, and the other interviewed household member by the male enumerator. The surveys were not conducted if privacy could not be maintained.[27]

Table 1.2 shows that the survey response rates for the sampled women were very high, above 90 % in the first wave for the promotion analysis sample and above 80 % for the exposure analysis sample. The response rate for the nominated operators dropped to 77 % in wave 2 for the promotion analysis, compared to 91 % of selected operators. In Section 1.5.3, I show that this attrition is not driving results. About 70 % of both samples were found to be married and living with their husbands in the pre-surveys, who were then

[25] If the sampled woman was married and living with their husbands in the same household, the husband was identified as the second respondent to be interviewed. If the sampled woman was living with other adults but not currently married, or married but not living with the spouse, the adult in the household that the respondent identified as the main decision-maker was identified as the second respondent. If the sampled woman lived alone or only with minor children, she alone was interviewed. When interviewing the promotion sample in the second wave, we aimed to interview the same second respondent as in the previous wave. Only if the circumstances of the sampled woman had changed, for example if she married between the surveys or the respondent in the previous wave had left the household, was another respondent interviewed. Throughout the project, the following definition of a household was used: "Someone is a member of your household if they have spent at least 6 months out of the last year living in the same house and eating from the same kitchen. If you share a house with other households or families, we are only interested in the primary household/family."

[26] For example, in a number of cases a neighbours' room was used.

[27] The enumerator team was entirely separate from the team that conducted the promotion programme and the surveys in the factory, though they were hired by the same field organisation. At no point during the survey administration was the promotion programme mentioned to the respondents. In the consent form, the surveys were explained to the respondents as aiming to understand "how families of workers in the garment industry live, and how the different members in the family interact with one another".

Table 1.2: Survey response rates

Panel A: Promotion analysis

	Selected	Nominated
Wave 1	93 %	91 %
Wave 2	91 %	77 %
Phone (ever reached)	94 %	95 %
Phone (avg. surveys)	7.28	5.98

Panel B: Exposure analysis

	Exposed	Non-exposed
Wave 1	84 %	83 %

identified as the second respondent to be surveyed. Only for approximately 10 % of respondents, other adult household members were targeted, primarily parents and siblings. For readability, I therefore use the terms "wife" and "husband" in the rest of the paper to refer to the women known to us from the factories and all second household members interviewed, respectively. Close to a fifth of respondents lived without another adult in the household, for example in factory dormitories. Table 1.C.1 in the appendix shows the breakdown in detail. Since I am interested in bargaining between household members, I exclude these individuals from one-person households from the analysis.

1.4.4 Main outcome variables of intra-household bargaining

I use the survey data to construct six main outcome variables for intra-household bargaining, which are defined in Table 1.3. The percentage of the household's income spent on assignable goods for women is the first outcome.[28] Respondents were asked to report their expenditures for women in Taka for 1) cosmetics, 2) clothing and footwear, 3) jewellery and accessories, and 4) health and medical expenditures.[29] These expenditures were winsorised, harmonised on the monthly level, summed up, divided by the household's

[28] I define assignable goods as private goods whose nature allows the researcher to deduce that they are consumed exclusively or primarily by a specific type of household member.

[29] Note that, in the survey, the question for cosmetics referred to both women and girls. This was done because it is was thought likely that cosmetics are shared between women and their under-age female household members, such that these expenditures could not be identified separately.

total monthly income, and multiplied by 100.

The second main outcome variable is the percentage of the household's income spent on remittances to family or friends living outside the household in the last calendar month, including monetary and in-kind transfers. These are defined as the transfers to the sampled woman's friends and family plus transfers to the husband's friends and family, if the husband was the second respondent.

Table 1.3: Main outcome variables

Variable	Definition and construction
Women's assignable expenditures (% in household income)	Percent of household income spent on assignable expenditures for women, consisting of expenditures on 1. Cosmetics for women and girls in the last calendar month, 2. Clothing and footwear for women in the last three months, 3. Jewellery and accessories for women in the last three months, 4. Health and medical expenditures for women in the last three months, Values are winsorised at a fraction of 0.01 in each tail. Expenditures are harmonised on the monthly level, summed up, divided by the household's total monthly income, and multiplied by 100.
Remittances (% in household income)	Percent of household income spent on transfers to family or friends outside the household, including monetary and in-kind transfers, consisting of 1. Transfers to the sampled woman's family and friends 2. Transfers to the husband's family and friends, if the husband was the second respondent. Values are winsorised at a fraction of 0.01 in each tail, summed up, divided by the household's total monthly income, and multiplied by 100.
Violence (5 months, Dummy)	Dummy variable indicating whether woman was beaten or slapped by the other respondent at least once in the past five months.

(Table 1.3 continued.)

Decision-making (Index)	Index of woman's involvement in seven decisions in the household: 1. Decision for her to take up work outside the home for income, 2. Decision for her to accept a promotion[30], 3. Decision to purchase large household appliances, 4. Decision about large house repairs, 5. Decision to purchase clothing and jewellery for herself, 6. Decision for her to take a bus to run an errand, 7. Decision for her to visit a friend in the neighbourhood. For each decision, an ordinal variable was coded such that 4=Woman decides alone without needing permission, 3=Woman decides alone, but needs permission, 2=Woman is involved in joint decision-making, 1=Others decide, but need the woman's permission, 0=Woman is not involved. I create dummy variables for each category and decision, and then compute an index following Anderson (2008) and O'Brien (1984).[31]
Hiding money (Dummy)	Dummy variable indicating whether woman chooses private option in any choice between receiving money privately or publicly in the game.
Controlling money (Taka)	Highest amount in Taka that woman is willing to give up to have money paid to herself instead of her husband (ignoring multiple switching).

The wife's recent experience of physical violence perpetrated by the husband against her is the third outcome. This is coded as an indicator variable capturing whether the woman reports that she has been slapped or beaten by the husband at least once in the last five months. This recall period was chosen to cover the time period since the start of the promotion programme. This variable was only collected from the wives, whereas all other

[30]For the selected or nominated women who had been offered a promotion in the wave 2 survey, these questions refer to the decision-making about the actual offer instead of the generic question.

[31] In this paper, an Anderson index is computed as follows. First, if necessary, all variables are recoded such that they indicate the same direction. Second, any ordinal and categorical variables will be recoded into several dummies. Ordinal variables will be recoded such that the lowest value is the excluded base group (e.g. for an ordinal variable with 4 categories from 0=strongly disagree to 4=strongly agree, the dummies will be defined as 0 and above (the excluded base group), 1 and above, 2 and above, 3 and above, 4 and above. All dummies except the base group are included in the index.). For categorical, non-ordinal variables, dummies for each category will be created and all dummies except one base group will be included in the index. Third, using the recoded variables, the index developed by Anderson (2008) and O'Brien (1984) is computed. The index is produced using a self-written Stata programme that follows Anderson's stepwise approach. As a last step, the programme standardises the index by the mean and standard deviation of the sample (separately for the promotion and exposure analysis samples, and within those for husbands and wives separately), to allow for easier comparison of effect sizes.

outcome variables were also collected from the husbands.[32]

The fourth variable captures women's involvement in seven different decisions in the household. These dimensions capture decisions about the woman's daily autonomy, i.e. visiting a friend, taking a bus, and purchasing clothing and jewellery for herself. They also cover decisions which affect the entire household, i.e. regarding large household repairs and purchases of large appliances as well as decisions about the woman's career such as taking up work and accepting a promotion. Respondents were asked to describe who makes these decisions in their household in an open-ended question, which was then coded by enumerators. As a follow-up question, respondents were asked whether the decision-makers required permission from any other household members. For each decision, the two questions were recoded into one ordinal variable where a higher number indicates a more extensive involvement in decision-making. These range from zero, indicating that the woman is not involved in the decision, to four, which indicates that the woman decides alone without needing permission. I then compute a summary index of all seven decisions following Anderson (2008) and O'Brien (1984), taking care not to ascribe cardinal meaning to ordinal differences. Note, however, that the results below are robust to different definitions, e.g. coding a binary variable indicating whether the woman is involved versus not involved in a decision, or whether she is the sole decision-maker versus not. The advantage of this ordinal approach is that it captures all potential margins of change in one summary index.[33]

[32]I followed the guidelines of the World Health Organization for domestic violence research, and under all circumstances prioritised the safety of respondents and the field team (World Health Organization, 2001). Enumerators were specifically trained for the administration of these questions about domestic violence. They were only administered once the enumerator had again confirmed the privacy of the interview, and the other household members were not aware that the survey with the sampled woman included questions about domestic violence. Only two questions about domestic violence were asked to ensure that the woman's survey was not substantially longer than the husband's. Respondents were offered the phone number of Bangladesh's National Helpline for Violence against Women and Children, irrespective of their responses. To increase truthful reporting and privacy, the respondents input their own responses for this section on the tablet, as suggested in Heath (2014), after the questions and response options were read out by the enumerator. This was possible since all respondents in the survey were literate and had experience using mobile devices.

[33]This use of an index to summarise reported household decision-making is similar to previous work in the literature (See e.g. Ashraf et al. (2010); Pitt et al. (2006), and Anderson (2008) on general advantages of using summary index tests over testing individual outcomes.). Compared to the use of the raw reported variables, an index requires the researcher to make decisions about its specific construction, which may not be innocuous. For example, in this case, as defined in Table 1.3, the coding of the responses for the decision-making questions (which is later dichotomised) implies that a wife who makes decisions alone is more involved in decision-making than a wife who decides jointly with her husband. This assessment may not be shared by all women,

The last two main outcomes are derived from two economic games that I implemented as part of the survey, which measure cooperation in the household. These were modelled after Almås et al. (2018). In the first wave of the household survey, respondents played an incentivised game that asked them to choose between receiving different amounts of money privately or publicly, i.e. with their spouse's knowledge.[34] "Hiding money" is a dummy variable indicating whether the woman ever chose the private option in a series of comparisons. In the second wave, respondents played a different game, which asked them to choose between receiving an amount of money themselves or whether their husband should instead receive a different amount of money.[35] "Controlling money" is defined as the highest amount the woman is willing to give up to control the money herself in a series of comparisons.[36] Note that, because each game was only implemented in one period, the sample size for these outcomes for the promotion analysis is half that of the other outcomes. Since the sample for the exposure analysis was interviewed only once with the survey instrument for the first wave, the outcome Controlling money is not available for these respondents.

These outcome variables correlate with demographics broadly as expected. This is demonstrated in Table 1.C.2 in the appendix, where I show how the outcome variables relate to a variety of individual and household characteristics for the non-exposed respondents of the exposure analysis. The percentage of income spent on assignable goods for women are positively, though not always significantly, related to the woman's age, her education, the woman's share in income, the number of household members, her age at marriage, and whether the woman brought assets into the marriage. Being married and having a

especially regarding decisions that affect the whole household (e.g. about large house repairs). It is therefore crucial to probe whether any effects found are an artefact of the specific index construction chosen by the researcher. For this reason, I conduct extensive robustness checks of the results with regards to different ways of constructing the index and to using the raw variables themselves below.

[34]For example, 100 Taka sent via mobile phone credit to the woman tomorrow without the husband knowing vs. 150 Taka sent via mobile phone credit to the woman tomorrow, accompanied by a message to the husband's phone informing him how much the woman receives.

[35]E.g. 100 Taka in phone credit paid to you after the survey vs. 150 Taka in phone credit paid to your husband after the survey.

[36]During the administration of the wave 1 survey, it became clear that there was a strong norm against hiding among respondents. For example, only 12 % of women in the non-exposed group of the exposure analysis ever chose the private option. This had not been apparent during the piloting. Qualitative surveys confirmed this norm against hiding, but also revealed a much weaker norm against directing money to oneself instead of the partner. This motivated the use of a different game in wave 2.

larger number of children is negatively related to women's expenditures, which is not surprising given that these variables capture competing demands on income. The correlates for the household's percentage of income spent on remittances largely look similar to those of women's expenditures, with some notable differences. A higher number of household members correlates negatively with remittances, as does an indicator for whether the woman brought assets into the household. This makes intuitive sense since remittances are likely lower in households with fewer low-income relatives living outside the household. Somewhat counterintuitively, however, the woman's income share is also negatively negatively related to the household's remittances.

The quantitatively most important correlate for domestic violence in the last five months is the woman's income share, with a higher income share related to less violence. In comparison, the woman's involvement in decision-making is highly positively correlated with the indicator whether she brought assets into the marriage and with the woman being married. Married women and women who brought assets into the marriage are less likely to hide from their spouses in the economic game.

1.4.5 Descriptive statistics

The typical woman in the sample, for both the promotion and the exposure analysis, is in her mid-twenties and lives with her husband and one child in a nuclear family. Table 1.4 shows summary statistics for the promotion sample in Panel A, using the data from the factory baseline conducted for the promotion programme. Panel B shows descriptives for the exposure sample in Panel B. Because no baseline for the exposure analysis was conducted, Panel B is derived from the household surveys of the non-exposed comparison group. See Table 1.C.3 in the appendix for variable definitions. Women in both samples are similar on a number of dimensions. They are on average about 26 years old and more than 80 % are married. Nearly two thirds of the women have migrated to Dhaka from their place of birth. On average, they have one child. Interestingly, more than half of respondents with children report that at least one child does not live in the household,

but instead lives with relatives.[37] The women are also four to five years younger than their husbands. Despite this age gap, they only have about a third of a year less education than the husbands. They contribute slightly below half of the total household income.[38] Interestingly, two thirds of all husbands also work in the garment industry.

The samples differ in other aspects, which indicate that the women who were nominated or selected for the promotion programme come from a higher socioeconomic background and play a larger role in their households than women in the exposure analysis sample. For example, 12 % of women in the promotion analysis sample consider themselves the household head, compared with only 1 % in the exposure analysis sample. In addition, the former group has on average about two years of education more than the latter group. The households in the promotion analysis sample also report slightly higher household income, of about 23,500 Taka (roughly 235 GBP and 280 USD at the time of the surveys), compared to about 21,000 Taka (roughly 210 GBP and 260 USD) in the exposure sample.

[37] Note that this is derived from Panel B, since I don't have comparable data from the promotion analysis factory baseline.

[38] Note that we only collected individual income data for the wife and husband in the promotion analysis baseline, but total household income during the household surveys. For the promotion analysis baseline, I therefore approximate total household income as the sum of income of both husband and wife. This will only differ from the total household income variable used in Panel B if household members other than the husband and wife also contribute income.

Table 1.4: Summary statistics

Panel A: Promotion analysis

	N	Mean	SD	Min	Max
Age	198	25.80	3.53	19.00	38.00
Married	198	0.84	0.36	0.00	1.00
Household members	198	3.27	1.57	1.00	10.00
Household head	198	0.12	0.33	0.00	1.00
Migrant	198	0.66	0.47	0.00	1.00
Education years	198	8.33	1.76	0.00	15.00
Nr Children (if ever married)	180	1.04	0.79	0.00	5.00
Combined income of spouses (if married)	167	23630.19	6556.26	9500.00	44760.00
Education difference of spouses (if married)	167	−0.31	3.05	−10.00	9.00
Age gap of spouses (if married)	166	−4.64	2.75	−14.00	4.00
Husband works in garment industry (if married)	167	0.66	0.47	0.00	1.00
Woman's share in spouses' income (if married)	167	0.46	0.14	0.09	1.00

Panel B: Exposure analysis

	N	Mean	SD	Min	Max
Age	238	26.20	4.16	18.00	40.00
Married	238	0.89	0.31	0.00	1.00
Household members	238	3.09	1.42	2.00	8.00
Household head	238	0.01	0.11	0.00	1.00
Migrant	238	0.63	0.48	0.00	1.00
Education years	238	6.08	2.48	0.00	10.00
Nr Children (if ever married)	224	1.16	0.84	0.00	4.00
Child outside HH (if Nr Children ≥ 1)	176	0.57	0.50	0.00	1.00
Total household income	238	21166.26	7325.44	6500.00	43000.00
Education difference of spouses (if married)	208	−0.35	3.74	−10.00	9.00
Age gap of spouses (if married)	208	−5.24	3.25	−23.00	−1.00
Husband works in garment industry (if married)	208	0.66	0.47	0.00	1.00
Woman's share in household income	238	0.48	0.18	0.00	1.00

Notes: Panel A uses data from the baseline survey of the promotion programme, conducted in the factory. Because there is no baseline for the exposure analysis, Panel B uses data from the household surveys only from the non-exposed comparison group.

1.5 The effects of a promotion

1.5.1 Empirical strategy

The selection process for the promotion programme described in Section 1.4.2 allows me to compare outcomes of selected women to nominated runners-up who were not included in the programme because the factory did not have enough vacancies available. Recall that, after the factories provided us with a ranked list of candidates fulfilling pre-set criteria, we conducted extensive screening of all nominees and excluded those who failed literacy and numeracy tests. Half of all factories were able to re-rank their candidates in the shortlist once after reviewing the diagnostic tests, whereas the shortlist for the remaining factories remained the same. Out of the operators in the final shortlists, only those with a rank that was better than the factory-specific cutoff were selected and participated in the promotion programme. The nominated women who passed all diagnostic checks but ranked below the cutoff continued to work as production-line operators and represent the comparison group.

I argue that the assignment to the selected and the nominated groups is quasi-random conditional on observables because factories have limited ability to predict which women would be successful as a supervisor among the eligible women. The factories had historically nearly always promoted men, and therefore had large numbers of competent female line operators from which to choose new supervisors. Our extensive eligibility checks meant that only skilled and experienced line operators were considered for the programme. While it might be easy for factories to exclude operators who would *not* make a good supervisor in their factories, it is much harder to decide the ranking among a sample of women who had *all* been judged to be suitable after this extensive screening. This argument is similar in spirit to the literature on microenterprises, recruitment and venture capitalism, which shows that humans are often inept at predicting which firms, workers or investments will be successful among an already select sample (McKenzie and Sansone, 2017; Hoffman et al., 2018; Nanda, 2016).

Support for the empirical strategy

The results from the selection experiment in Chapter 2 provide support for this strategy. There, we find that the nominees' original rank – before a potential re-ranking in the factories that participated in the selection experiment – is orthogonal to the results of their diagnostic tests. Only once the factories are presented with these screening results as part of the selection experiment do they re-rank the candidates on their skills, and especially their soft skills. I explain below how I take this potential selection effect in the half of the factories that were randomised into the selection experiment into account in the estimation.

I also find strong support for this empirical strategy when comparing the selected and the nominated group on observable characteristics.[39] The balance tests in Table 1.5 demonstrate that the selected and nominated groups are statistically indistinguishable on a wide range of observable characteristics. See Table 1.C.3 in the appendix for variable definitions. Only two tests out of 34 reject equality at the 10 % significance level, using the conventional t-tests in column (5). This is fewer than one would expect to reject by chance if the null hypothesis is true. The selected women have about a year *less* experience in the garment industry, but they score about seven percentage points higher on the diagnostic literacy test than the nominated group.

I also show p-values using randomisation inference in column (6) and normalised differences between the means of the nominated and selected groups following Imbens (2015) in column (7). The normalised difference is a sample-size free way to investigate balance in covariates, and an absolute value smaller than 0.30 in absolute value can be considered well balanced, according to Imbens (2015).[40] These additional tests strengthen the conclusion that the nominated and selected groups are very similar. I reject only one hypothesis when using the randomisation inference p-values, and all normalised differences are weakly smaller than 0.30. Lastly, I conduct a test of joint orthogonality by investigating

[39] Note that, for all checks in this subsection, I only include women whose data is also used in the analysis, i.e. who were interviewed in at least one household survey wave and had another adult decision-maker in the household. This is to ensure that attrition is not driving an imbalance.

[40] The normalised difference is calculated as the difference in means between the two groups, divided by the square root of the average of the sample variances of the two groups.

Table 1.5: Balance of nominated and selected groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Nominated		**Selected**		p-value	Norm-
	N	(Mean)	N	(Mean)	p-value	(RI)	diff
Panel A: Variables from baseline survey							
Age	38	26.18	160	25.71	0.45	0.47	−0.13
Married	38	0.89	160	0.83	0.34	0.40	−0.18
Household members	38	3.32	160	3.26	0.85	0.91	−0.03
Household head	38	0.05	160	0.14	0.15	0.30	0.29
Migrant	38	0.68	160	0.66	0.74	0.81	−0.06
Education years	38	8.00	160	8.41	0.19	0.16	0.22
Nr Children	35	1.06	145	1.03	0.88	0.88	−0.03
Combined income of spouses	34	23408.41	133	23686.89	0.83	0.80	0.04
Spousal education difference	34	0.26	133	−0.46	0.22	0.29	−0.22
Age gap of spouses	34	−4.97	132	−4.55	0.43	0.42	0.15
Wife's income share	34	0.45	133	0.47	0.41	0.36	0.18
Decision-making	38	−0.13	160	−0.18	0.79	0.77	−0.05
Supportive family	38	0.09	160	0.06	0.89	0.90	−0.02
Experience in garment sector	38	6.97	160	5.91	0.08*	0.12	−0.30
Exposure to female SV	38	0.63	160	0.54	0.33	0.31	−0.18
Internal locus of control	38	0.20	160	0.08	0.50	0.44	−0.12
Grit	38	−0.05	160	0.13	0.28	0.19	0.20
Self-efficacy	38	−0.06	160	0.06	0.51	0.52	0.11
Emotional competence	38	0.14	160	0.11	0.86	0.86	−0.03
Multi-factor Leadership	38	0.08	160	0.08	0.97	0.96	−0.01
Life satisfaction	38	7.55	160	7.60	0.89	0.92	0.03
Self-assessment	38	0.03	160	0.14	0.76	0.71	0.06
Ambition	38	2.16	160	2.20	0.84	0.86	0.04
Panel B: Variables from diagnostic tests before selection							
Literacy score	38	0.51	160	0.58	0.09*	0.04**	0.29
Numeracy score	38	0.45	160	0.47	0.62	0.61	0.08
Processing speed score	38	0.34	160	0.33	0.75	0.72	−0.06
Garment knowledge score	38	0.56	160	0.54	0.28	0.46	−0.21
Family support score	38	0.69	160	0.72	0.39	0.31	0.16
Interest score	38	0.72	160	0.74	0.62	0.59	0.09
Confidence score	38	0.73	160	0.73	0.96	1.00	−0.01
Panel C: Time-invariant variables from household surveys							
Muslim	38	0.97	160	0.97	0.87	1.00	−0.03
Socioeconomic background	38	−0.10	160	−0.01	0.65	0.73	0.09
Marriage duration	36	8.72	149	8.13	0.51	0.64	−0.13
Brought assets in marriage	36	0.31	149	0.34	0.73	0.68	0.06

Notes: Column (5) shows the p-value of a conventional t-test of equality of the group means in columns (2) and (4). Column (6) shows the p-value of the t-test implemented using randomisation inference, with 5000 replications of treatment assignment that maintain the share of selected women within factory strata. Column (7) shows the normalised difference of the group means in columns (2) and (4), calculated as the difference in means between the two groups divided by the square root of the average of the sample variances of the two groups, following Imbens (2015). * p<0.1, ** p<0.05, *** p<0.01

whether these 34 observable characteristics jointly predict selection in a linear regression.[41] The p-value of the F-test is 0.87, which shows that selection is not determined by a wide range of baseline observable characteristics (see Table 1.C.4 in the appendix).

Main specification

I employ three different estimation approaches. I use OLS with control variables chosen by PDS Lasso in my main analysis, and discuss the regression discontinuity design and the matching method in Section 1.5.3. The following is the main specification for outcome Y of each nominee and selected operator i in factory f, while pooling both household survey waves $w = \{1, 2\}$:

$$Y_{ifw} = \alpha + \beta \cdot selected_{if} + g(\boldsymbol{X}_{if}) + \epsilon_{ifw} \tag{1.2}$$

$selected$ is a dummy variable indicating whether the individual is an operator selected for the promotion programme, which is zero for the nominated group. $\boldsymbol{X}$ is a vector of potential control variables. The vector consists of all 34 variables for which I tested balance on in Table 1.5 and their squared terms, in addition to factory, enumerator, and month dummy variables. By including all diagnostic scores as well as factory fixed effects, I account for a potential re-ranking on these scores by the factories participating in the selection experiment. The set of control variables to be included are selected using the PDS Lasso algorithm, using the data-driven penalty loadings for clustered standard errors suggested in Belloni et al. (2013, 2014a,b, 2016). Standard errors are clustered at the individual level.

The PDS Lasso algorithm achieves a parsimonious selection of control variables in two steps. It first implements the Lasso algorithm twice to select two sets of covariates from the $\boldsymbol{X}$ vector: one set which is predictive for the outcome variable Y and one set which is predictive for $selected$, the covariate of interest. In the second step, β is estimated using

[41] In this paper, to include the variables only defined for married respondents, I follow two steps. First, I set these variables to zero if they are missing and, second, I include dummy variables that indicate whether the variables are missing in the regression. As Table 1.C.4 in the appendix shows, I also do not find that covariates predict selection if I include all covariates but limit the sample to married respondents, or if I only include covariates which are defined for all respondents in the regression. The p-values for the F-statistic of the regression in these cases are 0.96 and 0.92, respectively.

an ordinary least squares regression of Y on $selected$ and the union of the two sets of covariates selected in the first step. As Belloni et al. (2013) show, this method performs well across a variety of methods, enhances efficiency, and leads to valid inference even with mistakes in variable selection.

Since I am interested in intra-household decision-making, I limit the sample to those women who report having another adult decision-maker in the household. As shown in Section 1.4, this is the husband in the vast majority of cases. I compute sharpened q-values to correct for multiple hypotheses testing for all main outcome variables defined in Table 1.3 following Benjamini and Hochberg (1995) and Benjamini et al. (2006).

1.5.2 Results

Promotions and income

The promotion programme was successful in advancing women's careers. As column (1) of Table 1.6 shows, the women selected for the programme were 24 percentage points more likely than the nominated group to have been officially promoted to supervisor after the conclusion of the promotion programme. They also saw an income increase of about 900 Taka, equivalent to about 9 % of the income in the comparison group, as per column (2). However, an official promotion increased the woman's personal income much more. When I interact being selected for the programme with the indicator for being promoted to supervisor in column (3), I find that the personal income of officially promoted women increased by about 32 %, compared to around 9900 Taka (99 GBP, 124 USD) in the nominated group. Recall, however, that all selected women were assigned to trial as supervisors for at least two months, and therefore had experience in a leadership position. A number of them were also still waiting to be officially promoted at the time of the household surveys. Of course, the official promotion is likely an endogenous outcome. I therefore focus on the intent-to-treat effects of being selected for the promotion programme for the main results, but in Section 1.5.4 unpack the results into the different potential effects identified in Section 1.4.1.

Table 1.6: Effects on promotions and income

	(1) Supervisor (Dummy)	(2) Personal income (Taka)	(3) Personal income (Taka)
selected	0.24***	887.93 *	−270.69
	(0.06)	(509.53)	(544.67)
selected*supervisor			3183.95 ***
			(711.07)
N	363	363	363
Nominated mean	0.09	9883.31	9883.31
Nominated SD	0.29	3203.56	3203.56

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Main outcomes

Turning to my main outcomes, Table 1.7 shows the headline results. I find large, positive and significant effects on the share of income that households spend on expenditures on women's assignable expenditures, as well as on remittances. The households of the selected women spend 2.71 percentage points more on women's goods than the households of the nominated women, as shown in column (1). This effect is large at about 51 % of the mean in the nominated group of 5.32 %. Column (2) also shows a large effect on remittances, of 3.48 percentage points. This is 58 % of the comparison group mean. The sharpened q-values in squared brackets support that these results remain significant even after accounting for multiple hypotheses testing.

I do not see significant effects on my other main outcomes, as shown in columns (3) to (6) of Table 1.7. The effects on domestic violence in the past five months and the woman's decision-making power in the household are small and insignificant. However, their direction supports an increase in bargaining power, with a negative point estimate on domestic violence and a positive estimate on women's involvement in decision-making.[42]

Similarly, the outcomes of the economic games do not appear to be affected. Respondents

[42]Note that I also do not find an impact on whether the women has ever experienced domestic violence, with a point estimate of 0.03 from a mean in the nominated group of 0.30. The mean of 0.30 is in line with the incidence of domestic violence that Heath (2014) finds in Bangladesh. She reports that 37.4 % of married women who work for pay have ever been beaten by their husbands.

in the selected group are a little more likely to choose a private payout in the hiding game, as seen in column (5) of Table 1.7. Their willingness to pay for controlling money themselves is virtually unchanged, however, as shown in column (6). Compared to the mean and standard deviation of the nominated group, both coefficients in columns (5) and (6) are small and insignificant.[43]

To understand these results better, I look at the components of women's assignable expenditure separately in Table 1.8. The effect is positive for all components of women's assignable expenditures—cosmetics, clothing and footwear, accessories and jewellery, and health expenditures. However, the effects are strongest on expenditures on women's clothing and footwear as well as accessories and jewellery.[44] It is possible that some of the effects on clothing and accessories expenditures are driven by a need for the selected women to look more professional in their new supervisor position. I argue that the size of the effects over time makes this less likely. It is arguably more likely that the women selected for the promotion programme need to invest in a stock of new clothing when starting their new position, which would suggest a larger effect early on. In Table 1.C.5 in the appendix, I look into the effects on women's assignable expenditures for the two waves of the household surveys separately.[45] Even though I lose power because of the smaller sample size when splitting the survey waves, the coefficient in the second wave (3.22 percentage points) is more than a third larger than the effect in the first wave (1.99 percentage points). This shows that the households of the selected women continued to spend a large income share on women's assignable expenditures even eight months after the beginning of the promotion programme, which is less likely to be driven by job requirements.

The increase in remittances is also unlikely to be driven by job requirements. In Table 1.9, I find that the effect on total remittances appears to mainly be driven by transfers to the

[43]These results from the economic games are consistent with a narrative in which the selected women's subjective marriage quality did not decline or could have improved, since they are not more likely to hide from their spouses and also do not have a higher willingness to pay to control money compared to the nominated group.

[44]When comparing the distributions of these outcomes in the selected and nominated group in Figures 1.B.4 and 1.B.5 in the appendix, I find that the overall effect is concentrated in the upper half of the distribution, though the effect on clothing expenditures is clearly affecting the entire distribution.

[45]I only include respondents interviewed in both waves to ensure that sample composition does not drive potential differences in effect size across waves.

Table 1.7: Effects on main outcomes

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)	(6) Controlling money (Taka)
selected	2.71**	3.48**	−0.03	0.09	0.03	−5.32
	(1.07)	(1.40)	(0.04)	(0.15)	(0.08)	(17.90)
	[0.04]	[0.04]	[1.00]	[1.00]	[1.00]	[1.00]
N	363	362	363	363	181	182
Nominated mean	5.32	6.00	0.09	−0.16	0.28	146.77
Nominated SD	4.48	8.63	0.29	1.13	0.45	147.72

Notes: Standard errors clustered at the individual level in parentheses. Sharpened q-values in squared brackets. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.8: Effects on the components of women's assignable goods expenditures

	(1) Cosmetics (% in HH income)	(2) Clothing (% in HH income)	(3) Jewellery (% in HH income)	(4) Health (% in HH income)
selected	0.02	1.06***	0.54**	1.10
	(0.18)	(0.23)	(0.22)	(0.96)
N	363	363	363	363
Nominated mean	0.96	1.60	0.11	2.65
Nominated SD	1.24	1.37	0.67	4.06

Notes: Standard errors clustered at the individual level in parentheses. Sharpened q-values in squared brackets. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.9: Effects on remittances (% in HH income)

	All	Married couples with children					
	(1) Total	(2) Total	(3) Wife's family	(4) Husband's family	(5) Total	(6) Wife's family	(7) Husband's family
selected	3.48**	3.53**	0.87	3.03**	1.29	0.11	1.68
	(1.40)	(1.69)	(0.96)	(1.46)	(1.70)	(0.92)	(1.28)
selected*Child outside HH					3.60	1.11	2.00
					(3.20)	(1.87)	(3.08)
Child outside HH					4.07	1.74	2.42
					(2.78)	(1.66)	(2.89)
N	362	237	238	237	237	238	237
Nominated mean	6.00	6.46	2.73	3.73	6.46	2.73	3.73
Nominated SD	8.63	8.31	4.52	6.87	8.31	4.52	6.87

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

husband's family, if I restrict the sample to married couples who have children. Columns (3) and (4) in Table 1.9 show that the effect on remittances to the husband's family is more than three times as large as the effect on remittances to the wife's family. However, it appears that these remittances are mostly benefitting children who live outside of the household. As mentioned in Section 1.4.5, in this context it is common for children to not live with their parents and to be cared for by other family members. Of course, the decision whether children live with their parents or not is endogenous, and the sample size becomes challenging when investigating interaction terms. As indicative evidence, however, columns (5) to (7) of Table 1.9 show that households with selected women where at least one child lives outside the household report higher remittances. Though the interaction terms are insignificant, they are large at 3.6 percentage points in column (5), for example. This at least suggests that the higher remittances to the family are masking higher expenditures on children. However, I do not find that the promotion programme led the selected group to send their children to live with relatives, when I investigate the dates that children started living outside the household collected during the surveys.[46]

1.5.3 Robustness checks

In this section, I show that the large positive effects on expenditures on women's assignable goods and remittances are robust to different estimation approaches, different variable definitions, different specifications, different data sources, placebo tests and checks for attrition, though I lose significance in a few instances. I also demonstrate that the effects are not mainly driven by the training that was delivered as part of the promotion programme.

[46]I do not find an effect on the way women and husbands spend their time. Across the two household survey waves, nearly all women and husbands report working, 92 % and 94 %, respectively. This means that, at the point of the surveys, some women have left their work in the garment factories since the start of the promotion programme. However, I do not find an impact of being selected for the programme on the extensive margin of working for either women or husbands. Table 1.C.6 in the appendix shows the effects on women's time use in hours in Panel A and, using the husband's survey data, on his time use in Panel B. The coefficients are small, for both the last work day in columns (1) to (4) and the last non-work day in columns (5) to (7). There is some indication that the selected women work around a quarter of an hour longer and sleep shorter by the same amount on the last work day, but these effects are small in economic terms.

Regression discontinuity design

The positive and significant impacts on women's assignable goods expenditures and remittances are robust to implementing a regression discontinuity design (RDD), despite the low statistical power of this approach.[47] For the analysis, I use the ranked shortlist of operators provided by each factory after removal of any ineligible candidates, as explained in Section 1.4.2. Recall that only those operators with a rank better than the factory-specific cutoff were assigned to the programme. Since factories nominated different numbers of operators, I scale the rank of operators by the total number of operators nominated in each factory to make the ranks comparable across factories. I use the distance of this scaled rank variable from the factory-specific cutoff as the running variable. I determine the optimal bandwidth following Imbens and Kalyanaraman (2012)[48] for each outcome variable. Using the observations within this bandwidth, I then allow for a linear function of the running variable separately on either side of the cutoff when estimating the effect of being selected for the promotion programme. Standard errors are clustered at the individual level.

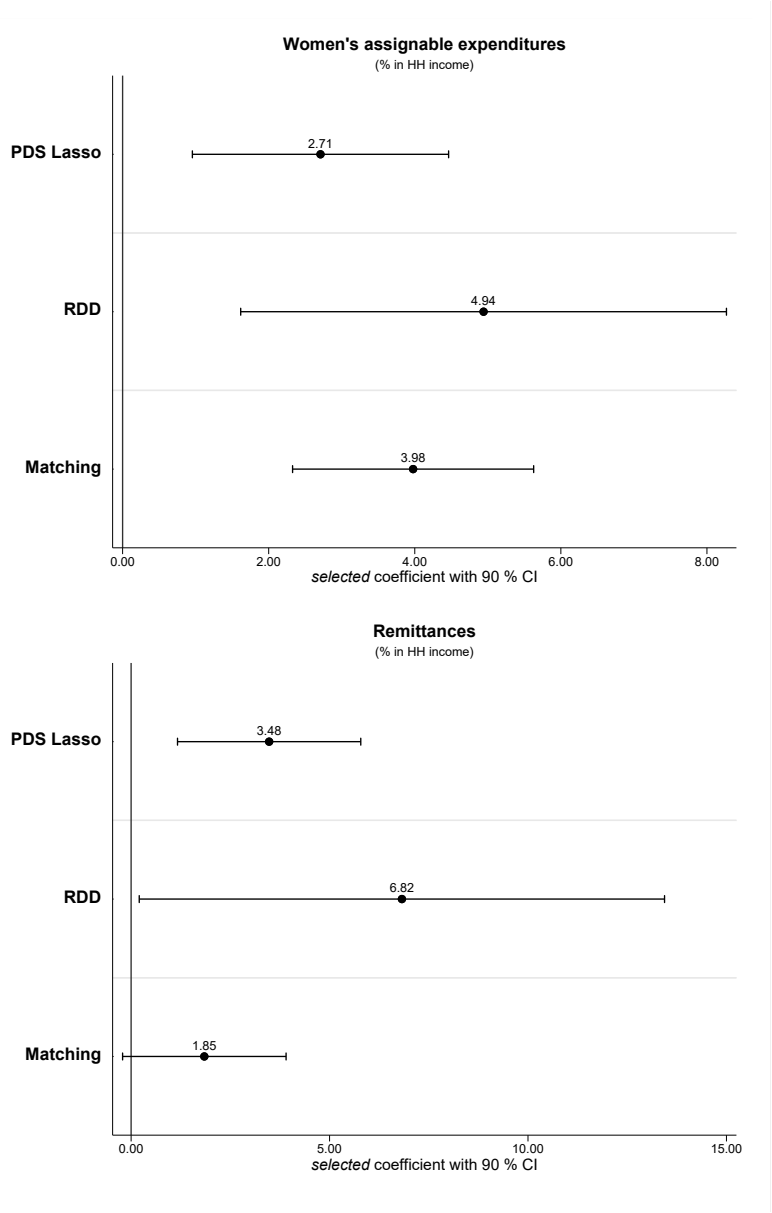
Figure 1.1 compares the coefficients from the regression discontinuity approach with the estimates from PDS Lasso and the matching approach discussed below. Table 1.C.7 in the appendix presents results for all main outcome variables. As the figure shows, the RDD point estimates for expenditures on women's assignable goods and remittances remittances are larger than those from the PDS Lasso estimation, at 4.94 percentage points for women's expenditures and 6.82 percentage points for remittances. Though they are less precisely estimated, I find significant effects on both these outcomes.[49]

[47] See, for example, Deke and Dragoset (2012) and Schochet (2008) for a discussion of the lower power of regression discontinuity designs.

[48] I use the Stata command rdbwselect to do so.

[49] If I allow for a quadratic function of the running variable on both sides of the cutoff separately, I find coefficients of 9.74 percentage points for women's expenditures (p-value of 0.02), and 3.13 percentage points for remittances (p-value of 0.67).

Figure 1.1: Comparison of estimates using three different estimation approaches



Matching

Estimating the results with a matching method qualitatively also confirms the previous results. I implement the nearest-neighbour matching estimator suggested by Abadie et al. (2004), with one-to-one matching with replacement and using the matrix of the inverse sample standard errors of the matching variables as weights in the matching. I use all balance variables in Table 1.5, as well as all factory, enumerator and month dummies as

matching variables.[50] As Figure 1.1 and Table 1.C.8 in the appendix show, the matching estimate for the effect on women's expenditures (at 3.98 percentage points) is larger than the estimates obtained using PDS Lasso and significant. The coefficient for remittances (at 1.85 percentage points), however, is smaller and loses significance with a p-value of 0.14.

Different variable definitions

I next check robustness to different variable definitions for women's assignable expenditures and remittances in Panels A and B of Table 1.C.9, respectively. I again estimate Equation 1.2 with PDS Lasso. I repeat the main results from Section 1.5.2 as a comparison in the first columns, and vary the definition, first, to take the inverse hyperbolic sine transformation of the percent of total household income and, second, to the monetary value in units of Taka (winsorised, harmonised on the monthly level, and then summed up). For both outcomes, the checks confirm large positive effects, though these are not always significant.

The inverse hyperbolic sine transformation for the income share of women's expenditures suggests an effect of about 20 % and of about 30 % for remittances in the second column of both panels, respectively. However, neither coefficient is significant, with a p-value of 0.15 for women's expenditures and 0.22 for remittances.[51] Since the inverse hyperbolic sine is a transformation which compresses the distribution similar to a natural logarithm, this again suggests that the effect is concentrated in part of the distribution, which is the upper half of the distribution here. Note, however, that the impact on the inverse hyperbolic sine of the income share of women's clothing expenditures remains highly positive and significant, with an effect of about 41 % and a p-value of 0.00.

When I vary the definition to look at the sum of women's expenditures and remittances in monetary terms in the third column in both panels, the coefficients remain highly positive and significant. The point estimate suggests that households of selected women spend about 396 Taka (about 5 USD or 4 GBP) more per month on women's assignable goods,

[50] To include balance variables which are missing for some observations, I set these to zero if they are missing and also include indicators if they are missing in the matching.

[51] See Bellemare and Wichman (2020) for a discussion on the interpretation of regression coefficients with inverse hyperbolic sine transformations.

about a third of the comparison mean. Similarly, the estimate for spending on remittances in monetary terms suggests higher expenditures of about 954 Taka (about 12 USD and 9.5 GBP), about 73 % of the comparison mean.

Different specifications

I now return to the preferred definitions of my outcome variables—the percent of household income spent on women's assignable goods or remittances. I investigate robustness of my main results to different OLS specifications and a larger set of potential controls for the PDS Lasso algorithm. This is shown in Table 1.C.10 in the appendix, where the first column repeats the results from my preferred specification in Section 1.5.2. I then sequentially introduce different sets of control variables, starting from OLS with no controls in column (2). Standard errors are always clustered at the individual level. Column (3) adds factory, enumerator and month fixed effects, and column (4) instead adds linear and squared controls for the imbalanced baseline control variables of experience in the garment industry and the literacy score. Column (5) instead adds linear controls for all baseline balance variables from Table 1.5, and column (6) adds all changes in columns (2) to (5) together. Column (7) implements OLS with controls chosen based on economic theory, where I include covariates for being married, total household income, total household income squared, number of girls, boys, women and men in the household, age of the respondent, as well as factory, enumerator and month fixed effects. Finally, column (8) implements PDS Lasso but allows for a larger set of potential control variables, including all balance variables from Table 1.5, factory, enumerator and month dummy variables, all possible interaction terms between these variables, in addition to all squared balance variables.

As seen in Panel A of Table 1.C.10, the point estimates for women's assignable goods expenditures fluctuate around the PDS Lasso estimate in column (1). They remain significant for all specifications, with only one exception. When I include all linear balance variables and all factory, enumerator and month dummy variables as controls in column (6), the coefficient reduces to 1.58 percentage points, compared to 2.71 in column (1). This could

possibly due to overcontrolling. When I allow for the larger set of potential controls in PDS Lasso in column (8), with a total of 4,486 potential controls, the coefficient decreases slightly to 2.04 percentage points compared to column (1), but it remains significant at the 5 % level.

The effect size for remittances also remains relatively stable across the specifications, and the estimates are always significant. The effect size with all controls in column (6) of Panel B in Table 1.C.10 is about 19 % larger than the PDS result in column (1). Once I allow for the larger set of potential control variables in column (8), the effect size decreases by only around 5 % compared to column (1), to 3.32 percentage points, and it remains significant.

Husband's survey

The effects on women's assignable expenditures and remittances are confirmed when using the data collected during the surveys with the husbands. Table 1.C.11 in the appendix shows the results for the main outcomes, estimating equation 1.2 with the husband's data.[52] The point estimate of the effect on women's expenditures is 2.21 percentage points and significant using the husband's survey, compared to a slightly larger 2.71 percentage points from the wives' data in Table 1.7. For the budget share of remittances, the estimate from the husband's data is 5.13 percentage points and significant, compared to 3.48 percentage points from the wives' data.

Note that husbands and wives also report similar *levels* of budget shares. The wives of the comparison group of non-selected nominees report that the household spends 5.32 % of household income on women's assignable expenditures and 6 % on remittances on average. This compares to a mean of 5.71 % for women's expenditures and 6.31 % for remittances from the husbands' data. It is a common pattern in this data that both respondents in the household report similar expenditures and income. For example, for total household income before winsorisation, I find a correlation coefficient of 0.85 between the reports of both respondents across both waves. This suggests that information about income and

[52]Note that I use the same set of potential control variables for the balance variables of the PDS Lasso algorithm as in Section 1.5.2, i.e. the variables from the wife's data. Also note that some husbands were unexpectedly unavailable at the time of the survey or refused to be surveyed, such that the number of observations is slightly smaller for husbands than for the wives.

expenditures is widely shared within the household, possibly because the regular income streams in the garment industry are both difficult to hide and make budgeting easier.[53]

High-frequency phone surveys

My main results are confirmed with the data from the wife's high-frequency surveys. Table 1.C.12 in the appendix shows the results. I estimate Equation 1.2 with the same set of potential control variables, but use survey round fixed effects instead of month fixed effects. I also keep the same estimation sample as in Section 1.5.2 for comparison. For time reasons, the high-frequency surveys asked about assignable goods expenditures for women and girls combined, and only included questions about remittances to the wife's family (not the husband's). It also did not include questions about income. My outcome measures are therefore, first, the sum of household expenditures on women's and girls' assignable goods in Taka per month (again consisting of cosmetics, clothing, accessories and health), and remittances to the wife's family in Taka per month.[54] I find large and significant effects on women's and girls' assignable expenditures, with a point estimate of 658 Taka (about 8 USD and 6.60 GBP), about 37 % of the mean in the comparison group. My results also show significant and large increases of remittances to the wife's family, with a coefficient of 1009 Taka (about 13 USD and 10 GBP) that represents about 43 % of the comparison group mean.

Placebo test: Comparing top and bottom half of selected group

I implement a placebo test by comparing outcomes of the women in the top half of the selected group to outcomes of women in the bottom half of the selected group. This means that I am now only comparing women who were participants of the promotion programme. Finding no effects for this comparison would support the claims that (1) the ranking of eligible women in the shortlist is orthogonal to my outcomes of interest, and that (2) the cutoff which separates the selected from the non-selected is the relevant mar-

[53]This could potentially also explain why I see little effects on the outcomes from the economic games I implemented, which measure respondents willingness to pay for hiding or controlling money.

[54]The monetary values are again winsorised, harmonised on the monthly level and, if applicable, summed up.

gin of interest. I create the placebo treatment as follows. I only keep the women who were selected for the promotion programme. For each factory ranking, I then assign the selected women who are in the top half of the ranking to the placebo treatment, and the bottom half of the selected women to the placebo control. For example, imagine a factory had 15 women in the shortlist after excluding ineligible women, and ten of these women were selected for the promotion training, while the other five are the non-selected nominees. I drop the non-selected nominees for this placebo test and assign the five selected women with the top rank in the shortlist to the placebo treatment and the remaining five selected women in the shortlist to the placebo control.

I estimate Equation 1.2 as described in Section 1.5.1, but now the $placebo$ treatment as created above is my covariate of interest. Table 1.C.13 shows the estimates for the placebo treatment effect for all my main outcomes. Importantly, I do not find effects of the placebo treatment on women's expenditures and remittances. The coefficients for both women's expenditures and remittances are small and insignificant, at 0.31 and 0.60 of a percentage point, respectively. These compare to estimates of 2.71 and 3.48 percentage points for my actual treatment in Section 1.5.2. The effects of the placebo treatment on the other main outcomes are also small and insignificant, with the exception of the willingness to pay for controlling money in the economic game, for which I find a negative and significant effect.

Attrition

I now show that survey attrition is not driving the results for women's expenditures and remittances, by implementing two checks. First, I restrict the estimation of Equation 1.2 to a balanced panel in columns (1) and (2) of Table 1.C.14 in the appendix, which means that I only keep respondents who are part of the sample in both waves. Both point estimates, for women's expenditures (at 2.49 percentage points) and remittances (at 2.93 percentage points), are similar to my main results in Section 1.5.2 and significant despite the smaller sample size.

As a second check, I only include the data from Wave 1 of the household survey when estimating Equation 1.2, for which the response rates for the selected and the non-selected

nominated group were very high and very similar (See Table 1.2). I show the results in columns (3) and (4) of Table 1.C.14 in the appendix. The effect on women's expenditures in Wave 1 remains significant at 1.92 percentage points, though it is about 30 % smaller than in Table 1.7. The effect on remittances in Wave 1, at 3.08 percentage points, is only about 12 % smaller than in Table 1.7, and it remains significant. These two checks suggest that a change in sample composition due to attrition is not explaining the results.

Training effects

I also provide evidence that the results for women's expenditures and remittances shown in Section 1.5.2 are not primarily driven by the effects of the training provided as part of the promotion programme. To check for a potential effect of training, I investigate the effects for the three different groups discussed in Section 1.4.2.[55] Remember that all selected women were assigned to work as assistant supervisor for at least two months. In Table 1.C.15 in the appendix, I interact the $selected$ variable with variables indicating these training groups. As the first row of the table shows, the effects appear to be primarily driven by being selected for the promotion programme, not the training. The effects remain significant for the group that did not immediately receive training, at 3.32 percentage points for women's expenditures and 3.29 percentage points for remittances. None of the interaction terms are individually significant. The interaction term which is largest in magnitude is actually negative for the soft only training group, at -2.60 percentage points. Nevertheless, I cannot reject that the interaction terms are jointly equal to zero for both women's expenditures (p-value of 0.13) and remittances (p-value of 0.87). Overall, these results suggest that training is not a large driver behind the observed results.

Re-ranking effects

Lastly, I show that the results are not driven by the re-rankings of shortlisted candidates in the factories that participated in the selection experiment. Table 1.C.16 in the appendix

[55]Recall from Section 1.4.2 that those operators selected for the promotion programme were randomly assigned to one of three groups. One group received training in soft and technical skills immediately, one group only received soft skills training immediately and one group did not receive any training until later. After about six months, the soft skills only group and the group that had initially not been trained also received the missing training.

Table 1.10: Effects on other assignable goods expenditures (% in HH income)

	(1) Men (% of HH income)	(2) Girls (% of HH income)	(3) Boys (% of HH income)
selected	0.32	1.22**	−0.37
	(0.57)	(0.62)	(0.72)
N	363	146	161
Nominated mean	3.92	1.65	2.69
Nominated SD	3.99	2.41	3.65

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies. Columns (2) only includes observations where the respondent reported having at least a daughter, or a girl living in the household. Column (3) only includes observations where the respondent reported having at least a son, or a boy living in the household.

shows the results for the main outcomes only for the half of the factories that did not participate in the selection experiment. Since this cuts the sample into half, I am mainly interested in the magnitude of coefficients rather than their significance. As columns (1) and (2) demonstrate, the effects on women's assignable expenditures and remittances in this subsample are even larger than the results for the entire sample shown above – at 3.2 percentage points for women's expenditures (compared to 2.71 percentage points in Table 1.7) and 4.31 percentage points for remittances (compared to 3.48 percentage points in Table 1.7).

1.5.4 Unpacking the results

I now return to the simple collective household framework described in Section 1.4.1, and evaluate which, if any, of the three potential effects could be driving the large and positive increases in household budget shares spent on women's assignable goods and remittances. The small sample size makes it somewhat difficult to exactly pin down mechanisms, but the different pieces of evidence point into the same direction.

Marshallian income effect

A Marshallian income effect resulting from the woman's wage increase would ease the household's budget constraint, and increase the set of feasible consumption allocations. This could affect the demand for all goods consumed in the household, though the income

elasticity could of course differ between goods.[56] However, one would by definition expect a positive income elasticity for normal goods.

I present two pieces of evidence which suggest that the observed effects are not primarily due to Marshallian income effects. First, I investigate the effects on the percent of household income spent on assignable goods for men, girls and boys in Table 1.10.[57] Note that I only include the subsample of respondents with a daughter or a female child in the household when estimating the effects on girls' expenditures (and equivalently for boys' expenditures), which reduces the sample size. As can be seen in column (2), I find a significant positive effect on girls' expenditures of 1.22 percentage points, an increase of 74 % of the comparison mean.[58] This compares to a negative point estimate on boys' expenditures, which is insignificant at -0.37 percentage points. The effect on men's expenditures is positive and insignificant at only 0.32 percentage point, which is much smaller than the effect of 2.71 percentage points for women's expenditures in Table 1.7. These much larger effects on expenditures on female assignable goods, for both women and girls, compared to male assignable goods, for men and boys, suggest that a re-allocation of the household's budget towards female expenditures is taking place, rather than a Marshallian income effect.[59]

As a second piece of evidence, I investigate the effects on unmarried women who live in one-person households, for example in factory dormitories. For this subsample, intra-household bargaining is excluded by definition, but any possible Marshallian income effects should still be at play. Since the number of observations is small at only 49, I am mainly interested in the direction and size of effects. Table 1.11 shows a very small positive

[56] The functional form of the Engel curves will depend on the underlying demand system. See, for example, Banks et al. (1997)'s estimation of quadratic Engel curves based on a Quadratic Almost Ideal Demand System.

[57] Men's assignable goods expenditures consist of tobacco and alcohol, cosmetics, clothing and footwear, jewellery and accessories, and health expenditures. Boys' and girls' assignable expenditures consist of clothing and footwear and medical expenditures each. See Table 1.C.3 for definitions.

[58] Additional analysis shows that the effect is largest on the budget share of education expenditures for female household members.

[59] I also find that the effect sizes on the monetary value of assignable expenditures for women and girls are much larger than the effect sizes on the monetary value of assignable expenditures for men and boys. The effect on women's expenditures is 395.59 Taka compared to a mean in the nominated group of 1201.99 Taka (See Table 1.C.9), and the effect on girls' expenditures is 202.22 Taka from a mean of 391.78 Taka in the nominated group. This compares to an effect on men's expenditures of 76.12 Taka from a mean of 861.67 Taka in the nominated group, and a negative effect on boys' expenditures of -45.49 Taka from a mean of 566.15 Taka in the comparison group. This again indicates that the increased spending is concentrated on women's and girls' assignable expenditures.

Table 1.11: Effects on women in one-person households

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)
selected	0.35	−3.74
	(2.09)	(6.86)
N	49	49
Nominated mean	11.58	19.32
Nominated SD	14.70	20.45

Notes: Table only includes unmarried respondents living in one-person households. Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.12: Relative wage effect

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)
selected	3.61**	3.33**
	(1.41)	(1.57)
selected*		
Starts earning more	−4.02 *	−3.11
	(2.16)	(3.88)
Starts earning more	0.98	4.06
	(1.56)	(3.16)
N	301	301
Nominated mean	4.99	6.21
Nominated SD	4.36	8.21

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

effect on women's expenditures of 0.35 percentage points (3 % of the comparison mean), and a negative and effect on remittances of -3.74 percentage points (about 19 % of the comparison mean).[60] Taken together with the evidence from Table 1.10, this suggests that the large effects on women's assignable goods expenditures and remittances are unlikely to be driven primarily by a Marshallian income increase.

Relative wage effect

I next investigate whether a relative wage effect could explain the results, that is, whether the increase in the woman's wage rate compared to the husband's affects the bargaining weight as a distribution factor. The evidence suggests that this is unlikely. I look into this by analysing whether the effects differ for women who started earning more than their husbands between the factory baseline and the household survey.

We know from Table 1.5 that the women earn about 46 % of the joint spousal income at factory baseline. In addition, about a third already earn more than their husbands at this point. I then compare this baseline information with the information about the woman's income share from the household surveys, and define a dummy variable "Starts earning more" as equal to one if the women did not earn more than her husband at baseline, but did earn more than the husband at the point of the respective household survey.[61] Across both waves, about 14 % of the women started earning more than their husbands between baseline and household survey.

I interact the new variable, $Starts\,earning\,more$, with being selected for the promotion programme in Table 1.12 and also include the non-interacted $Starts\,earning\,more$ in the regression. The results do not suggest that a relative wage increase in the women's favour is driving the results. In fact, the interaction term is negative and sizeable for both women's expenditures and remittances, but only marginally significant for the first. Further investigation reveals that the negative interaction term for the budget share of women's expenditures is mainly driven by a decrease in health expenditures, whereas the magnitudes of the interaction terms for the other components of women's expenditures are small.

This evidence suggests that the relative wage increase in the women's favour is not driving the large and positive impacts on women's expenditures and remittances I found before.

[60] Note that 47 % of the selected women living in one-person households have been officially promoted with an income increase at the time of the household surveys, so that this lack of effect can not be attributed to a lack of income increase.

[61] Note that the income information at baseline was collected from the wife. To compute the income ratio at the point of the household survey, I use the reports from the wife about her own income and from the husband about his own income. Since some husbands refused to participate or were unavailable, this reduces the sample size.

Table 1.13: Effects on women not yet officially promoted

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)
selected	3.59**	3.12*
	(1.42)	(1.69)
N	256	255
Nominated mean	5.49	6.29
Nominated SD	4.65	8.90

Notes: Table only includes subsample of women who were not yet officially promoted at the time of the household surveys. Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

These findings are in line with the results in Bertrand et al. (2015), which found in the United States that a woman who earns more than her husband does not appear to have more favourable intra-household outcomes. Instead, the authors find that these couples are less satisfied with their marriage, more likely to divorce and more likely to have a higher gender gap in non-market work.

Position effect

The previous two subsections suggest that the large effects on women's expenditures and remittances are not driven by a Marshallian income effect or a relative wage effect. I therefore investigate whether the leadership position could affect the woman's bargaining weight irrespective of income changes. To do so, I focus on the subsample of women who had not yet been officially promoted at the time of the household surveys, but who participated in the promotion programme and trialled in the supervisor position for two months. This means that this subsample has not experienced any wage increase. While an official promotion is an endogenous outcome, it is nevertheless instructive to investigate whether the effects observed in Section 1.5.2 hold for this subgroup.

Table 1.13 presents the results. The large, positive and significant effects on women's expenditures and remittances are also observed for this subsample of women who were not yet officially promoted. The point estimates, at 3.59 percentage points for women's expenditures and 3.12 for remittances, remain significant and are close to the coefficients found

Table 1.14: Effects on attitudes and beliefs

	(1) Overall self-beliefs (Index)	(2) Confidence (Index)	(3) Negotiating (Index)	(4) Gender attitudes (Index)	(5) Aspirations (Dummy)
selected	0.22*	0.15	0.13	0.07	0.07
	(0.13)	(0.13)	(0.11)	(0.09)	(0.07)
N	363	363	363	363	363
Nominated mean	−0.17	−0.12	−0.10	0.04	0.49
Nominated SD	1.14	1.11	0.95	0.74	0.50

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

for the whole sample in Section 1.5.2. This evidence suggests that even women who participated in the promotion programme but were not (yet) officially promoted saw their bargaining weight increase as a result of (at least temporarily) attaining a leadership position in the factory.

I find some evidence that the leadership effect operates by improving the women's attitudes and beliefs about themselves. Table 1.14 shows the results for several different measures of attitudes and beliefs (all defined in Table 1.C.3), as well as an overall index in column (1). I find positive point estimates on the women's confidence, her involvement in negotiation, gender attitudes and her aspirations, though none of these are individually significant. If I include all outcomes in columns (2) to (5) into one summary index, the effect on this index is positive and significant at the 10 % significance level. While this is not conclusive evidence, these results are consistent with the leadership position improving the selected women's self-beliefs and attitudes.

1.5.5 Summary

The results in this section have shown that quasi-random participation in the promotion programme leads to a higher share of household income spent on female assignable goods (both for women and girls) and on remittances to the family. The latter result appears to mask expenditures on children, since transfers are highest if children live outside the household. Thinking within a collective household framework, I provided suggestive evidence that the results are mainly driven by the leadership position improving the women's

beliefs and attitudes about themselves, rather than Marshallian income effects or an effect working through a change in the relative wage within the household. Overall, this indicates that the promotion programme increased the women's bargaining power in the household through the leadership position, if one is willing to assume that women have a higher preference than their partners for female assignable goods expenditures and for supporting family outside the household.

1.6 The effects of exposure to a female manager

The previous section provided evidence that the promotion programme increased bargaining power in the household for the participating women. This is encouraging evidence for the many policy approaches that aim to increase women's participation in leadership positions. However, since a large share of the workforce is made up of non-managerial positions (in the garment industry and otherwise), promotions to a management position will usually only affect a small share of the female population. In this section, I therefore analyse whether there are amplifying effects of female leadership on the non-managerial female workforce.

1.6.1 Empirical strategy

I analyse the potential effects of exposure to the new female supervisors discussed in the previous section on their female subordinates. I compare outcomes of production line operators who have worked with the new female supervisors from the promotion programme to outcomes of production line operators on other, male-supervised lines. I argue that the assignment of workers to production lines is largely determined by production requirements and can be considered quasi-random. As discussed in Section 1.3, production lines typically consist of 20-80 operators, depending on the type of garment produced. Each operator is responsible for one step in the production process at a work station (e.g. attaching the sleeve to a T-shirt), which differ in the skills required. According to discussions with an industrial engineer with substantial experience in several garment factories, the assignment of line operators to sewing lines is done by the management of the sewing

department, typically line chiefs or production managers. In this process, the most important consideration is to match operators to the work stations that they are skilled at (e.g. overlock machines are only operated by workers skilled in using this machine).

The management of the production floor typically prefers to keep the composition of well-functioning lines unchanged. However, there are some situations when it can be necessary to move operators for production reasons, for example to temporarily cover absences of operators on other lines, or when the style produced changes and with it the production requirements. Operators might also be moved to lines with a lower target if they struggle to meet targets or increase alteration rates. Line operators and supervisors are sometimes able to express their preferences for the assignment of operators to lines, but — barring any serious allegations of misconduct — the production requirements always take precedence in dealing with any potential requests. According to an industry expert, it would be especially unlikely to observe sorting of line operators towards lines of new supervisors in their trial period. During this trial period, supervisors are expected to demonstrate good performance with the team they have been given, and any requests by themselves or operators to move to their lines would be unlikely to be granted.

Balance tests

I provide three pieces of evidence that support this strategy.[62] First, I show that the exposed and non-exposed operators are balanced on a number of observable characteristics in Table 1.15. Since there is no baseline data for the random operators, these characteristics are from the household surveys. Out of the 15 mean-equality tests in Table 1.15, one test — for assets brought into the marriage — rejects at the 10 % significance level using the conventional p-values in column (5), and one additional test — for age — rejects when I use the randomisation inference p-values in column (6). This is in line with what one would expect to happen by chance. However, the normalised differences in column (7) are small

[62] As in Section 1.5.1, to take attrition into account, I only include women in these checks whose data is also used in the analysis, i.e. who were interviewed in the household survey and had another adult decision-maker in the household. In the ideal scenario, I would be able to use data of line operator movements to exclude sorting towards lines exposed to new female supervisors based on characteristics that determine the operators' household outcomes of interest. Unfortunately, the factories in this sample do not keep a reliable account of operator movements.

Table 1.15: Exposure analysis: Balance of non-exposed and exposed groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Non-exposed		**Exposed**		p-value	Norm-
	N	(Mean)	N	(Mean)	p-value	(RI)	diff
Age	238	26.20	250	26.91	0.10	0.08*	0.15
Married	238	0.89	250	0.92	0.21	0.19	0.11
Household members	238	3.09	250	3.21	0.39	0.38	0.08
Migrant	238	0.63	250	0.58	0.22	0.31	−0.11
Education years	238	6.08	250	5.94	0.56	0.55	−0.05
Nr Children	224	1.16	239	1.24	0.25	0.22	0.11
Household income	238	21166.26	250	20964.74	0.76	0.74	−0.03
Age gap of spouses	208	−5.24	219	−5.22	0.96	0.96	0.00
Muslim	238	1.00	250	0.98	0.20	0.35	−0.12
Socioeconomic background	238	0.08	250	−0.06	0.13	0.13	−0.14
Marriage duration	224	9.50	239	10.02	0.29	0.25	0.10
Brought assets in marriage	224	0.23	239	0.17	0.08*	0.08*	−0.16
Household head	238	0.01	250	0.03	0.23	0.32	0.11
Spousal education difference	208	−0.35	219	−0.29	0.86	0.85	0.02
Wife's income share	238	0.48	250	0.48	0.94	0.94	0.01

Notes: Column (5) shows the p-value of a conventional t-test of equality of the group means in columns (2) and (4). Column (6) shows the p-value of the t-test implemented using randomisation inference, with 5000 replications of treatment assignment that maintain the share of exposed women within factory strata. Column (7) shows the normalised difference of the group means in columns (2) and (4), calculated as the difference in means between the two groups divided by the square root of the average of the sample variances of the two groups, following Imbens (2015). * p<0.1, ** p<0.05, *** p<0.01

to moderate. None of them are greater than 0.16 in absolute value, much smaller than the cutoff used by Imbens (2015).

Second, I test whether all 15 variables jointly predict exposure to a female supervisor by estimating a linear regression of $exposed$ on all 15 variables.[63] The p-value of the overall F test of the regression is 0.41, as seen in Table 1.C.17 in the appendix. I hence conclude that the 15 variables are jointly not predictive for exposure. Third, I test whether the lines of the operators who were exposed to the female supervisors are different from the non-exposed lines on observable characteristics measured before the start of the promotion programme.[64] I use daily line-level production data we collected and harmonised from the factories for four important production outcomes: efficiency, alteration rates, absen-

[63] As in Section 1.5.1, to include the variables only defined for married women, I follow two steps. First, I set these variables to zero if they are missing and, second, I include dummy variables that indicate whether the variables are missing in the regression. As Table 1.C.17 in the appendix shows, I also do not find that covariates predict selection if I include all covariates but limit the sample to married respondents, or if I only include covariates which are defined for all respondents in the regression. The p-values for the F-statistic of the regression in these cases are 0.29 and 0.32, respectively.

[64] Again, I only use lines of operators who are actually included in the analysis, to account for attrition.

Table 1.16: Balance of exposed and non-exposed line characteristics

	(1) Efficiency	(2) Alteration rate, %	(3) Absenteeism, %	(4) Product complexity
exposed	−1.02	0.02	−0.34 **	−0.14
	(0.87)	(0.51)	(0.14)	(0.72)
N	18025	24960	21247	22148
Not exposed mean	48.16	8.52	4.23	17.28

Notes: All regressions include factory and day fixed effects, and are limited to the time before project start in each factory. Standard errors are clustered at the line level.

teeism and product complexity.[65] As Table 1.16 shows, before the start of the promotion programme, the lines exposed to the new female supervisors had very similar production statistics as the non-exposed lines. These groups did not differ in their efficiency, their complexity of garments produced and their alteration rates. The exposed lines have only slightly lower absenteeism rates, about 0.34 percentage points lower than the non-exposed mean of 4.23 %.[66]

Main specification

The main specification for the exposure analysis is similar to Equation 1.2 in Section 1.5.1. For outcome Y of each exposed and non-exposed subordinate operator s in factory f,

$$Y_{sf} = \alpha + \beta \cdot exposed_{sf} + g(\boldsymbol{X}_{sf}) + \epsilon_{sf}. \tag{1.4}$$

$exposed$ is a dummy variable indicating whether the individual was working on a line to which a selected participant of the promotion programme was assigned, where the individual's line code was collected at the survey follow-up for the promotion programme.

[65] Product complexity is measured by the standard minute value (SMV), and is defined as the time that the industrial engineers of each factory estimate a production line with qualified operators working at a standard pace would require to produce one piece of the garment. The efficiency of the lines measures how the daily output of each line in garment units, adjusted for the standard minute value of the garment, compares to the total output possible that day with the workers present on the line and the hours worked. As is standard in the industry, it is calculated for each line and day as

$$efficiency = \frac{(Pieces\,produced \cdot SMV)}{Workers\,on\,the\,line \cdot Hours\,operated \cdot 60}. \tag{1.3}$$

The alteration rate is the percentage of all garments that needs to be altered out of all produced units, and the absenteeism rate is the percentage of workers absent amongst all workers (present and absent) on the line.

[66] Note that these comparisons are within-day-within-factory comparisons. I include day and factory indicators in the regression to account for seasonality of production and different ways of measurement across factories, respectively. Standard errors are clustered at the line level because of high auto-correlation.

Table 1.17: Exposure analysis: Main outcomes

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)
exposed	−0.89	−1.39	0.02	0.36***	0.07
	(0.74)	(1.10)	(0.02)	(0.09)	(0.04)
	[0.22]	[0.22]	[0.22]	[0.00]	[0.22]
N	488	483	482	488	488
Non-exposed mean	6.48	8.49	0.02	−0.26	0.12
Non-exposed SD	6.63	11.81	0.13	1.08	0.33

Notes: Standard errors clustered at the production line level in parentheses. Sharpened q-values in squared brackets. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory, enumerator and month dummies.

$exposed$ is equal to zero if the individual was working on a line unrelated to the promotion programme.[67] Note that this comparison is likely to be a conservative estimate, since the female supervisor can be expected to have been visible to operators working on non-exposed lines, too. $\boldsymbol{X}$ is a vector of potential control variables, which consists of all 15 variables for which I tested balance on in Table 1.5 and their squared terms, in addition to factory, enumerator and month dummy variables.[68] The set of control variables to be included are selected using the PDS Lasso algorithm as explained in Section 1.5.1. Standard errors will be clustered at the production-line level, following Abadie et al. (2017). I again limit the sample to those women who report having another adult decision-maker in the household.

1.6.2 Results

Table 1.17 presents the results for the main outcomes when estimating Equation 1.4. I find that the women working with the new female supervisors report having more say in decision-making processes in the household, especially about their autonomy. Column (4) in Table 1.17 shows that the effect on the decision-making index is large at 0.36 of the overall standard deviation, and highly significant. It also is robust to multiple hypotheses

[67] This means that no selected participant for the promotion programme was assigned to that line and no selected participant was found to be working on that line at the survey follow-up for the promotion programme.

[68] Again, to include variables which are only defined for married respondents, I set these to zero if they are missing and also include variables indicating whether these variables are missing in the vector of potential control variables.

Table 1.18: Exposure analysis: Effects on components of decision-making

	(1) Visit a friend (Index)	(2) Take a bus (Index)	(3) Purchase clothing (Index)	(4) House repairs (Index)	(5) Purchase appliances (Index)	(6) Accept promotion (Index)	(7) Take up work (Index)
exposed	0.13*	0.25**	0.22**	0.17**	0.12	0.11	0.12
	(0.08)	(0.10)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
N	488	488	488	488	488	488	488
Non-exposed mean	−0.14	−0.27	−0.17	−0.26	−0.22	−0.12	−0.22
Non-exposed SD	0.85	1.00	1.13	0.87	0.85	1.26	1.18

Notes: Standard errors clustered at the production line level in parentheses. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory, enumerator and month dummies.

testing, with a sharpened q-value of 0.00.

I do not find significant effects on the other main outcomes, though there are some suggestions of more conflict in the household. The point estimates for the budget shares spent on women's assignable expenditures in column (1) and remittances in column (2) are negative, but not significant. Their magnitudes in comparison with the non-exposed mean are also smaller than the positive effects I found for the women who participated in the promotion programme. Recent domestic violence increases insignificantly by 2 percentage points in column (3), and exposed women are also a little more likely to choose a private option in the economic game, as shown in column (5).[69] These indications of conflict could potentially be a response to the wife's higher involvement in decision-making, but I lack the data to confirm this.

When investigating which of the the seven different decisions are affected, I find that all point estimates are positive, as Table 1.18 demonstrates.[70] However, I find the largest effects of more than 0.2 SD on the decisions for the woman to take a bus to run an errand, and on purchasing clothing and jewellery for the woman (columns (2) and (3)). The effect on the decisions to visit a friend and house repairs are also significantly affected. It hence appears that the women exposed to new female supervisors primarily gain more

[69]Note that the incidence of the women ever having been slapped or beaten by their husband is 0.30 in the non-exposed sample. I do not find an effect of exposure on this outcome.

[70]Note that the outcome variables for the separate decisions again are indices, since I dichotomise the four different categories of involvement explained in Table 1.3 into dummy variables, and then create a standardised Anderson index.

Table 1.19: Exposure analysis: Effects on different definitions of the decision-making index

	(1) Ordinal (Index)	(2) Involved (Index)	(3) Sole (Index)	(4) Joint (Index)
exposed	0.36***	0.27***	0.10**	−0.07
	(0.09)	(0.10)	(0.05)	(0.06)
N	488	488	488	488
Non-exposed mean	−0.26	−0.17	−0.20	0.18
Non-exposed SD	1.08	1.29	0.67	0.83

Notes: Standard errors clustered at the production line level in parentheses. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory, enumerator and month dummies.

autonomy in their mobility and spending decisions.[71]

When analysing how the decision-making process changes, I find that the exposed women are both more likely to be involved in the decision-making process, and more likely to be the sole decision-maker. The decision-making index presented in Table 1.17 is constructed using dichotomised ordinal variables and hence captures all potential margins of change, as discussed and defined in Section 1.4. In Table 1.19, I present results for three other ways of constructing the index, with column (1) repeating the main result for comparison. In column (2), the index is created using one binary variable for each of the seven decisions, where the binary variable is 1 if the woman is involved in the decision, and 0 if not.[72] In column (3), I also create the index using binary variables for each decision, but this time the variable indicates whether the woman makes the decision alone without needing permission, and is 0 otherwise.[73] Lastly, column (4) creates the index using binary variables indicating a joint decision involving the woman and other household members, which are

[71]There are also some indications that the exposed women are able to translate some of this decision-maker power into less time spent on domestic work, though this effect is economically small. As column (5) of Panel A in Table 1.C.18 illustrates, the exposed women spend about ten minutes less on domestic work on the last day off work. This is mirrored by the results from the husband's survey, which shows that the husbands' time spent on chores increases by about 12 minutes (see column (5) of Panel B in the same table). Note, however, that these offsetting effects are relative to very different means for wives and husbands. Whereas the wives in the non-exposed group spend on average more than 4.5 hours on chores on the non-working day, the husbands of the non-exposed wives only spend about 2 hours on these tasks. Economically speaking, the magnitude of the effects I find is therefore small. The patterns of time use on the last work day are less clear, as columns (1) to (4) in Table 1.C.18 show. The exposed wives appear to sleep marginally less than the non-exposed wives, which seems to be driven by more time spent on domestic work and leisure. The husbands instead report less leisure time, mainly resulting from more time spent on work.

[72]This means that the categorical variable is at least 1=Others decide, but need the woman's permission, or higher.

[73]This means that the categorical variable is 4=Woman decides alone without needing permission.

0 otherwise.[74]

As column (2) in Table 1.19 demonstrates, there appears to be a large effect on women being involved in decision-making, of 0.27 of a standard deviation. In addition, there is also a smaller (at 0.1 SD) but significant effect on women being able to make decisions alone, as seen in column (3). The effect on joint decision-making in column (4) is small and insignificant, though the point estimate is negative. These patterns are confirmed when I look at the seven decisions separately in Table 1.C.19 in the appendix. The likelihood of joint decision-making decreases for the decisions about visiting a friend and purchasing clothing for the women, whereas the exposed women are both more involved and more likely to be the sole decision-maker than the non-exposed group for the decisions of taking a bus and purchasing clothing.

Table 1.C.19 also underlines that the women working in the Bangladeshi garment industry are already highly involved in decision-making in their households. In the non-exposed group, for all seven questions about decision-making, at least 85 % of women report being involved in decision-making. Nevertheless, their autonomy is low: only 10 % of non-exposed women report that they on their own can make decisions about them visiting a friend. 8 % of non-exposed women can decide alone whether they want to take a bus to run an errand, and 17 % decide on their own about buying clothes and jewellery for themselves.

The results in this subsection therefore suggest that the exposure to female supervisors allowed women to gain more say in intra-household decision-making. However, they are not able to translate this decision-making power into changes in the household's spending. There are some non-significant indications of an increase in conflict in the household, which could possibly result from this increased involvement.

1.6.3 Robustness

In this section, I show that the large and positive effect on women's involvement in decision-making is robust to different variable definitions, a matching approach, different

[74]This means that the categorical variable is 2=Woman is involved in joint decision-making.

OLS specifications, a larger set of potential controls for the PDS Lasso algorithm, and to restricting the estimation to a subsample to take into account possible operator sorting. The results from the husband's data do not contradict a positive effect on the women's involvement in decision-making, but the point estimate is smaller.

Different variable definitions

I already demonstrated the robustness of the impact on women's involvement in decision-making to different variable definitions in the previous section in Table 1.19. As I showed then, the results also hold if I analyse the impact on women being involved (vs. not involved) in decision-making, or on being the sole decision-maker. There appears to be a slight negative effect on joint decision-making, which is not significant, however.

Matching

I implement the nearest-neighbour matching estimator suggested by Abadie et al. (2004) for the exposed and non-exposed comparison. I follow the strategy described in Section 1.5.3, and match on all variables for which I test balance in Table 1.15, in addition to factory, enumerator and month dummy variables. The results in Table 1.C.20 confirm the positive and significant effect on women's involvement in decision-making, with an estimate of 0.24 SD. This is a third smaller than the estimate I obtain using PDS Lasso, of 0.36 SD. In addition, I also find additional evidence that there is some increase in conflict when I use the matching approach, mainly due to more efficient estimates. The effects on both recent domestic violence (0.04 percentage points) and hiding money (0.06 percentage points) in the economic game are significant at the 5 % level.

Different specifications

I next show that the positive and significant effect on women's involvement in decision-making is robust to different OLS specifications and a larger set of potential control variables for the PDS Lasso algorithm. The columns of Table 1.C.21 are set out as described in Section 1.5.3. The coefficient remains highly significant and large across all columns of Table 1.C.21. I obtain the smallest estimate of 0.21 SD with OLS and all linear control

variables in column (5), but even this remains significant at the 5 % significance level.

Husband's survey

Using data from the husband's survey does not contradict the previous results, though I do not find such a large impact on women's involvement in decision-making. Table 1.C.22 shows the effects on the main outcomes using the husband's data. The effect on women's decision-making is again positive, but at 0.05 SD the effect size is much smaller than the estimate from the wife's survey, and insignificant. When looking into the seven decisions separately, I again see the largest effects on the decisions about the women's mobility, but none of the estimates are significant. Similar to their wives, the husbands in the exposed group are significantly more likely to choose the private option in the economic game (column (4) of Table 1.C.22), which could be considered another piece of evidence that conflict in the household increased.

Operator sorting

To investigate whether the results could be driven by operators sorting to the lines of the new female supervisors, I estimate the effects on the main outcomes for the subsample of operators who joined the lines exposed to the new female supervisors before the start of the promotion programme, comparing them to the previous sample of non-exposed operators. I use the data from the factory follow-up for Chapter 2 to determine this condition. I show the results in Table 1.C.23 in the appendix. Column (4) shows that the result is not due to operator sorting. The coefficient of 0.38 SD for this subsample is very similar to the result for the whole sample of exposed operators of 0.36 SD, and highly significant.

1.6.4 Unpacking the effects

Marshallian income effect

I again return to the framework introduced in Section 1.4.1 and first ask whether the large and positive effect on decision-making for the women working under the new female supervisors could be due to a Marshallian income effect. I show the effect on the exposed

Table 1.20: Exposure analysis: Effect on income

	(1) Personal income (Taka)
exposed	226.16 *
	(118.21)
N	488
Non-exposed mean	9387.21
Non-exposed SD	2863.31

Notes: Standard errors clustered at the production line level in parentheses. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.21: Exposure analysis: Effects on other assignable goods expenditures

	(1) Men (% in HH income)	(2) Girls (% in HH income)	(3) Boys (% in HH income)
exposed	−1.06 **	−0.21	−0.06
	(0.44)	(0.37)	(0.37)
N	488	255	262
Non-exposed mean	4.69	2.41	2.35
Non-exposed SD	5.25	2.50	2.41

Notes: Standard errors clustered at the production line level in parentheses. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory, enumerator and month dummies. Column (2) only includes observations where the respondent reported having at least a daughter, or a girl living in the household. Column (3) only includes observations where the respondent reported having at least a son, or a boy living in the household.

women's personal income in Table 1.20, estimated using Equation 1.4. The effect on the women's personal income is positive, but small and marginally significant at 2.40 % of the income in the non-exposed group. This small effect is not surprising since the exposed women continued working as line operators. At the time of the household surveys, none of the exposed or non-exposed women were working in a supervisor position.

Despite this small increase in the women's income, the main results in Section 1.6.2 showed that there was no effect on the budget shares of women's assignable goods expenditures or remittances, with point estimates actually being negative. I compare these results with the effects on the budget shares spent on other assignable goods expenditures in Table 1.21. I find no effects on assignable expenditures for girls and boys, and again find insignificant

Table 1.22: Exposure analysis: Relative wage effect

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)
exposed	0.59	−1.13	0.01	0.34***	0.10*
	(0.97)	(1.43)	(0.02)	(0.08)	(0.05)
exposed*earns more	−3.25 **	−0.13	0.05	−0.03	−0.04
	(1.54)	(2.09)	(0.03)	(0.17)	(0.06)
earns more	3.48***	0.43	−0.03 *	−0.13	−0.03
	(1.35)	(1.76)	(0.02)	(0.19)	(0.05)
N	449	446	445	449	449
Non-exposed mean	6.55	8.51	0.02	−0.24	0.11
Non-exposed SD	6.67	12.04	0.13	1.08	0.31

Notes: Standard errors clustered at the production line level in parentheses. Sharpened q-values in squared brackets. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory, enumerator and month dummies.

negative point estimates. The point estimate for assignable expenditures on men is negative and significant at 23 % of the comparison mean. The effect, however, is of similar size as for women's expenditures in Table 1.17, with -1.06 percentage points for men's goods compared to -0.89 percentage points for women.

Overall, these results suggest that a Marshallian income effect is unlikely to be at play, though the significant negative effect on men's assignable expenditures could be interpreted as a weak indication that expenditures are allocated away from men's goods.

Relative wage effect

I next analyse whether the effects differ depending on the share of income earned by the woman. I define an indicator "earns more" as equal to one if the woman earns more than 50 % of the combined income of the two respondents.[75] This is the case for 41 % of women in the sample. Table 1.22 shows the results when I interact this indicator with being exposed to a female manager and also include $earnsmore$ on its own in the regression. All of the interaction terms point to more negative outcomes for the exposed women who earn more, though the interaction term is only significant for the outcome of women's expenditures in column (1). For the decision-making index, the interaction terms is negative and

[75] I do not have baseline data for the sample in this analysis, so I do not know whether the respondents were already earning more than their husbands at the start of the programme.

Table 1.23: Exposure analysis: Effects on attitudes and belifs

	(1) Overall self-beliefs (Index)	(2) Confidence (Index)	(3) Negotiating (Index)	(4) Gender attitudes (Index)	(5) Aspirations (Dummy)
exposed	−0.03	0.00	0.17	−0.05	0.03
	(0.13)	(0.11)	(0.11)	(0.09)	(0.05)
N	488	488	488	488	488
Non-exposed mean	0.04	0.09	0.02	0.10	0.50
Non-exposed SD	1.23	0.84	1.04	0.93	0.50

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

insignificant. Taken together, this suggests that the woman earning more than the husband does not drive the large positive effect on the decision-making index.

Exposure effect

The previous two subsections suggested that Marshallian income and relative wage effects are likely not explaining the large positive effect on the exposed women's involvement in decision-making. I therefore investigate whether exposure to new female supervisors could have affected the attitudes or beliefs of the women working under these new female managers. Table 1.23 presents the results, and Table 1.C.3 in the appendix defines all variables.

The effects on the overall index and all individual components are small, except for the women's negotiation behaviour. For this variable, I find a positive effect of 0.17 SD, with a marginally insignificant p-value of 0.11. This outcome is an index derived from vignettes that present the women with a specific situation where a female character meets resistance from her husband against her plans (e.g. going to evening school to earn a school-leaving certificate). The respondent is then asked to describe how the character should respond to the husband to achieve her goals for each vignette, which is coded according to its sophistication (see Table 1.C.3) and summarised in an index according to Footnote 31. I interpret this outcome as capturing the women's willingness and ability to put herself forward in intra-household negotiations. Though this evidence can only be considered tentative, the large effect on this outcome suggests that the role model of a female supervisor in the workplace might have given the exposed women the ability or the confidence to get more

involved in decision-making in the household.

1.7 Conclusion

This paper analysed the impacts of women's career advancement on intra-household bargaining in the context of the Bangladeshi garment industry. I investigated both the direct effects on the promoted women themselves as well as potential indirect effects on the women working under the new female supervisors. I find that the women who participated in the promotion programme gained significant bargaining power in the household, as compared to the shortlisted runners-up. The effects are strongest for spending on women's assignable goods expenditures, and remittances to family members who take care of the couple's children. When analysing the effects of working under a female manager, as compared to working on comparison lines with male supervisors, results indicate that exposed women are more involved in decision-making in the household. Overall, I find suggestive evidence that the findings for both the direct and indirect effects are driven by women gaining the confidence to get involved in bargaining in the household, rather than Marshallian income effects or changes in the relative wage in the household.

Viewed together, these results demonstrate that there are potential complementarities between women's power in the workplace and women's power in the household. They suggest that policies to promote female career advancement have the potential to address inequalities in the household at the same time. Importantly, I find that the direct effects on the promoted women are amplified by the impacts on women working as subordinates of the female managers. These indirect effects are potentially relevant for a much larger number of women than the direct effects of promotions, which will usually only affect a small share of the female working population.

Acknowledgements I thank Christopher Woodruff and Simon Quinn for invaluable advice and support. The paper has benefited greatly from discussions with and comments by Abi Adams, David Atkin, Abigail Barr, Laura Boudreau, Emily Breza, Leonardo Bursztyn, Elwyn Davies, Cheryl Doss, James Fenske, Douglas Gollin, Rachel Heath, Matthew Klein, Paula López-Peña,

David McKenzie, Amanda Pallais, Natalie Quinn, Natalia Rigol, Heather Schofield, Tavneet Suri, Kate Vyborny, and from audiences at NEUDC 2019, Advances for Field Experiments 2019, Empirical Management Conference 2019, OxDev 2019, CSAE Workshop 2019, EDI Workshop 2017, and OxFLO 2017-18. The project would not have been possible without the cooperation of participating factories and workers, and the excellent Innovations for Poverty Action (IPA) field team. I gratefully acknowledge funding for the surveys for this project by the Private Enterprise Development in Low-Income Countries research initiative (PEDL); and the Economic Development and Institutions research programme (EDI), funded by UK aid from the UK government. The promotion programme was funded by the International Finance Corporation and the Urban Services Initiative, and run in co-operation with BetterWork Bangladesh, a programme of the International Labour Organization. This study, together with Chapter 2, was registered with the AER RCT Registry as AEARCTR-0001843, and received ethical approval from the University of Oxford and Innovations for Poverty Action.

Appendices

1.A Description of the training modules

The training consisted of one soft-skills component of four days, and one hard-skills component of five days. The soft skills training aimed to equip female operators with the non-cognitive skills to be successful as line supervisors. It was developed in cooperation with a Bangladeshi psychologist and covered the following topics:

- Stress management, self-awareness and setting boundaries
- Assertive communications and confidence
- Leadership and worker management skills

The hard-skills component aimed to give female operators the technical skills required for a line supervisor. It was developed in cooperation with Bangladeshi industrial engineers experienced in the garment industry and covered the following topics:

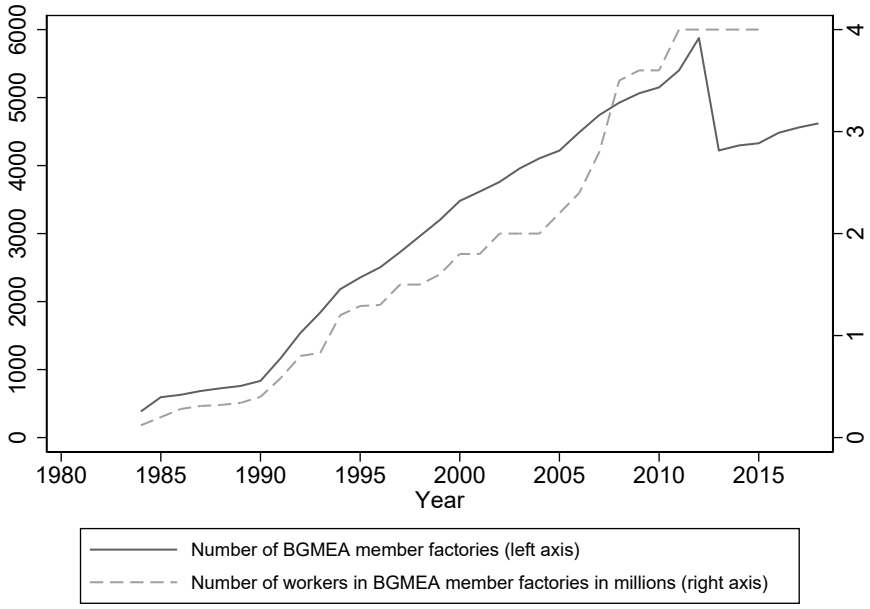
- Overview of the entire production process (from first contact with buyer to shipment)
- Planning and managing production lines (tracking production, efficiency calculations, machine layout, machine types and attachments, quality control)
- Solving problems on the production line (bottlenecks, line balancing, quality problems)

The soft-only and soft & hard-skills group both received the soft skills training immediately, whereas only the soft & hard-skills group received the hard-skills training immediately. The control group received training after the experimental period, about six months later.

The training was implemented part-time. Trainees participated in training at a training centre for two days a week, and worked in the factory the other four days of the working week. On the days in the factory, trainees from all three groups worked as Assistant Line Supervisors with gradually increasing responsibilities during the eight-week trial period. This set-up mirrors the traditional way in which workers are promoted to supervisors in the industry

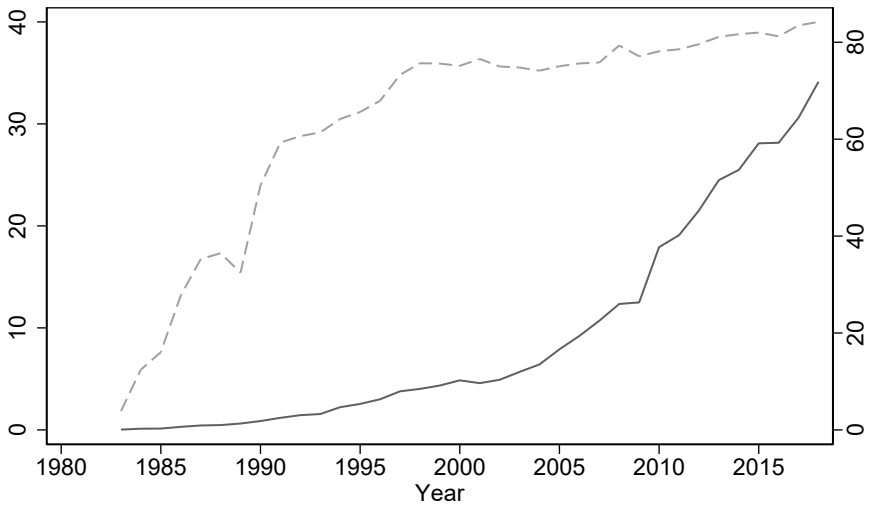
1.B Additional figures

Figure 1.B.1: Number of factories and employment in the Bangladeshi ready-made garment sector



Data source: BGMEA Trade Statistics 2017 and 2019

Figure 1.B.2: Exports of the Bangladeshi ready-made garment (RMG) sector



Data source: BGMEA Trade Statistics 2019

Figure 1.B.3: Timeline

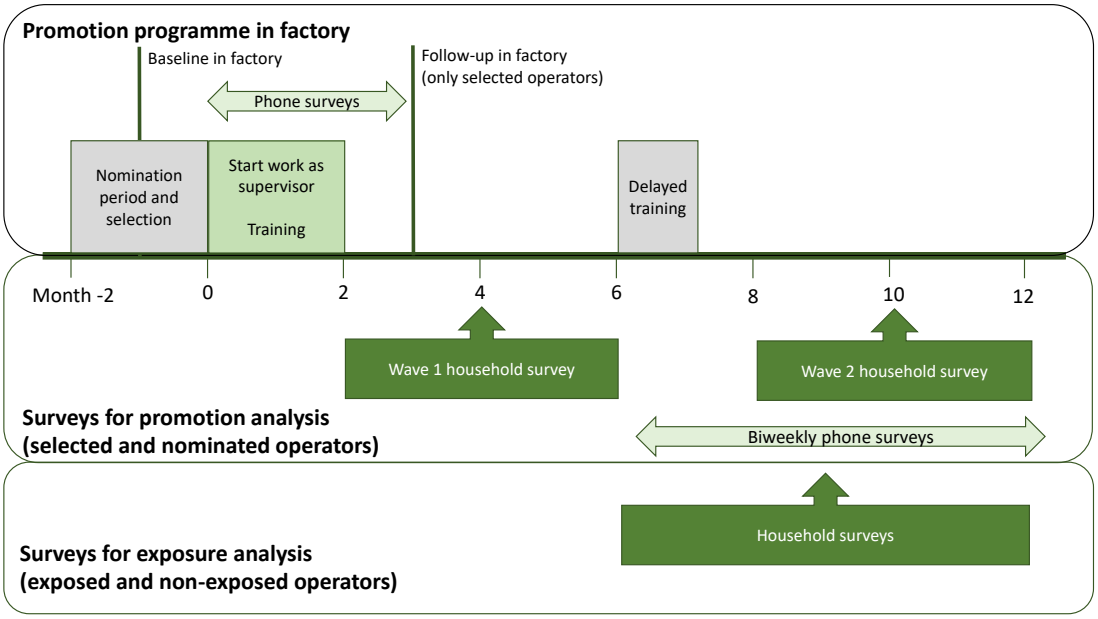
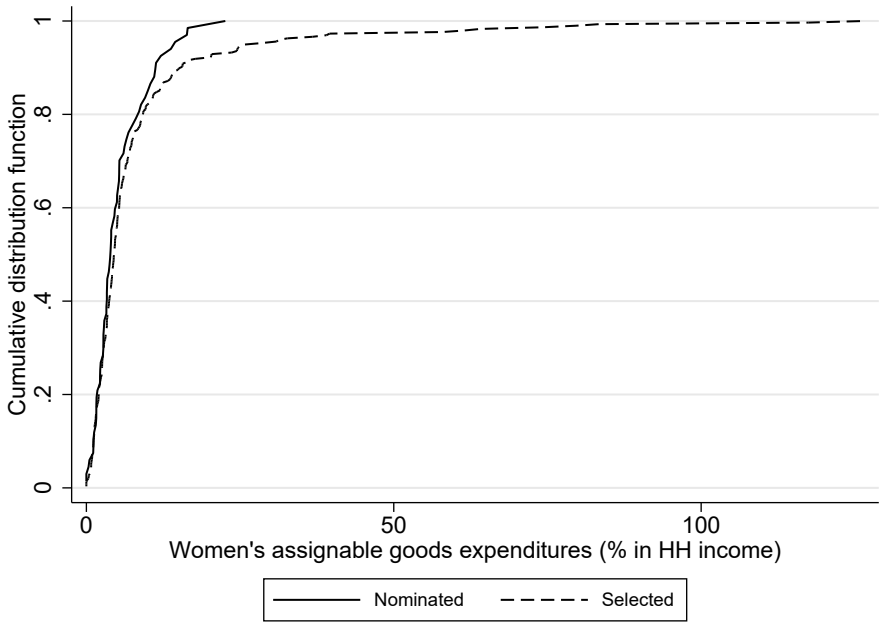
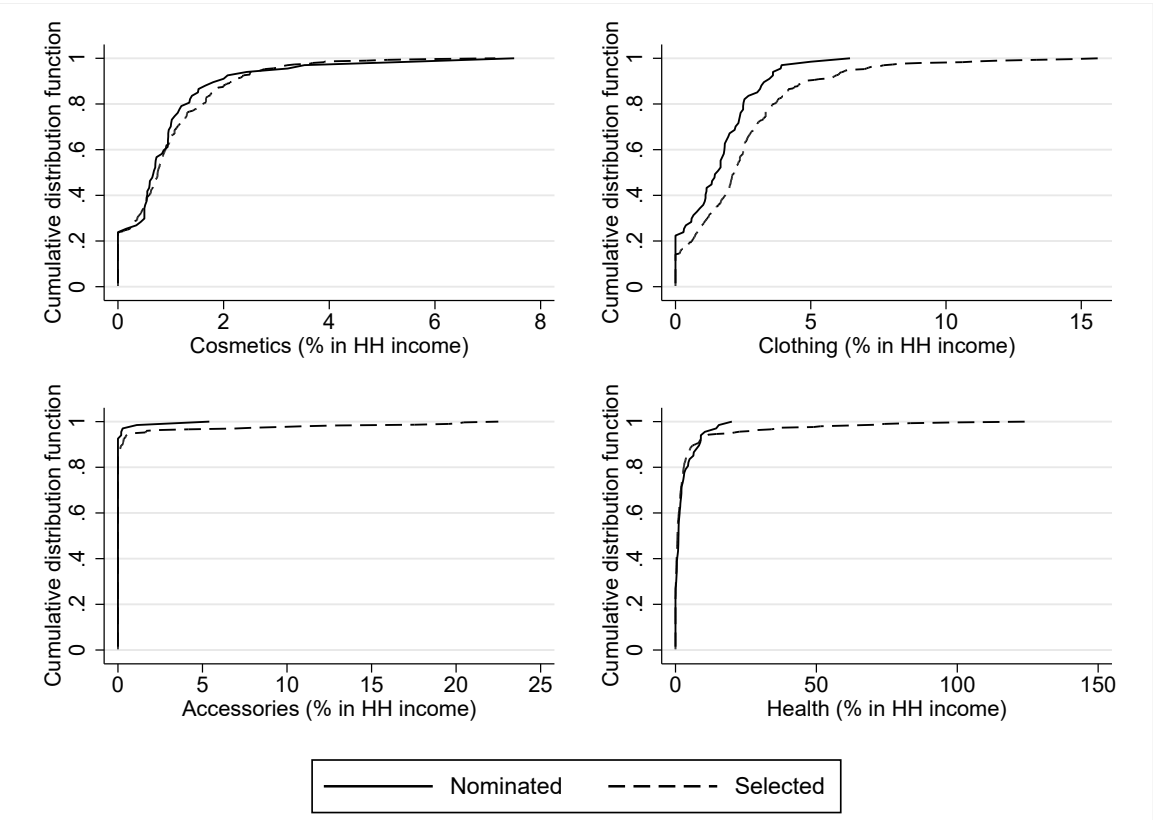


Figure 1.B.4: Cumulative distribution functions of women's assignable goods expenditures for nominated and selected women



Note: Both waves of household surveys are pooled for this graph.

Figure 1.B.5: Cumulative distribution functions of different women's assignable goods expenditures for nominated and selected women



Note: Both waves of household surveys are pooled for this graph.

1.C Additional tables

Table 1.C.1: Second respondents identified in pre-survey

Panel A: Promotion analysis (Wave 1)

	N	Percent
Husband	160	69.26
Father	11	4.76
Brother	4	1.73
Mother	4	1.73
Sister	5	2.16
Brother-in-law	1	0.43
No other adult	46	19.91
Total	231	100.00

Panel B: Exposure analysis

	N	Percent
Husband	453	72.36
Father	22	3.51
Brother	6	0.96
Mother	19	3.04
Sister	9	1.44
Other from your family	1	0.16
Father-in-law	1	0.16
Mother-in-law	1	0.16
Sister-in-law	2	0.32
No other adult	112	17.89
Total	626	100.00

Table 1.C.2: Correlations between main outcome variables and demographics

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)
Age	0.07	0.04	0.00	0.03	0.00
	(0.14)	(0.27)	(0.00)	(0.02)	(0.01)
Education years	0.18	0.43	−0.01	0.02	0.00
	(0.20)	(0.40)	(0.01)	(0.04)	(0.01)
Married	−5.32	−2.34	0.00	0.69	−0.24 *
	(3.43)	(2.95)	(0.03)	(0.58)	(0.12)
Household income	0.00*	0.00	0.00	0.00*	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Women's share in HH income	6.44**	−3.79	−0.04	−0.16	−0.01
	(2.76)	(4.92)	(0.05)	(0.32)	(0.15)
Nr Children	−1.15	1.73	−0.01	0.00	−0.01
	(0.92)	(1.29)	(0.01)	(0.10)	(0.04)
Household members	1.43*	−2.75 ***	0.00	−0.05	−0.02
	(0.78)	(0.45)	(0.01)	(0.07)	(0.02)
Age at marriage	0.23	0.19	0.00	−0.04	−0.01
	(0.18)	(0.33)	(0.00)	(0.03)	(0.01)
Brought assets in marriage	1.29	−0.68	0.02	0.38***	−0.03
	(1.41)	(1.93)	(0.03)	(0.11)	(0.05)
Age gap of spouses	−0.05	0.25	0.00	0.01	0.01
	(0.15)	(0.22)	(0.00)	(0.02)	(0.01)
Education gap of spouses	−0.09	0.30	0.00	0.00	0.00
	(0.15)	(0.28)	(0.00)	(0.02)	(0.01)
Constant	0.52	11.19	0.10	−0.59	0.53*
	(4.84)	(9.66)	(0.08)	(0.87)	(0.29)
N	238	238	237	238	238

Notes: Standard errors clustered at the production line level in parentheses. The table only includes observations in the non-exposed group for the exposure analysis. To include variables that are only defined for married respondents, these are set to zero for non-married respondents, and I include dummy variables that indicate whether they are missing as additional controls.

Table 1.C.3: Variable definitions

Variable	Definition
Panel A: Demographic and control variables from surveys	
Age	Age in completed years
Married	Dummy variable =1 if married
Household members	Number of household members
Household head	Dummy variable =1 if respondent is the household head
Migrant	Dummy variable =1 if respondent is not born in Dhaka
Education years	Highest educational attainment in years
Nr Children	Number of children (only available if ever married)
Child outside HH	Dummy variable =1 if any children live outside the household (only available if Nr Children ≥ 1)
Combined income of spouses	Sum of husband and wife's income from economic activities last month (if married). Only available for promotion sample.
Household income	Total household income from economic activities last month
Education difference of spouses	Wife's - husband's education years (if married)
Age gap	Wife's - husband's age (if married)
Woman's share in spouses' income	Ratio of wife's income from economic activity and combined income of spouses last month (if married). Only available for promotion sample.
Woman's share in household income	Ratio of wife's income from economic activity and household income last month (if married)
Decision-making	See Table 1.3.
Supportive family	Anderson index of the following 3 questions measuring support for female relatives working in garment industry, computed following Footnote 31: 1. Whether the female relatives were ever asked to quit their jobs (=0 if Yes) 2. How supportive the family is towards female relatives working in the garment industry (categorical variable) 3. How supportive the family would be if a female relative received an offer for a promotion to supervisor (categorical variable)
Experience in garment sector	Number of completed years in the garment industry
Exposure	Dummy variable =1 if respondent has worked under a female supervisor or manager
Internal locus of control	Anderson index of 8 locus of control questions indicating whether nominee chooses the internal option, based on Rotter (1966)
Grit	Anderson index of 6 grit items, higher score means higher grittiness, based on Duckworth and Quinn (2009)

(Table 1.C.3 continued.)

Self-efficacy	Anderson index of 10 self-efficacy items, higher score means higher self-efficacy, based on Schwarzer and Jerusalem (1995)
Emotional competence	Anderson index of 16 emotional competence items, higher score means higher competence, based on Mikolajczak et al. (2014)
Leadership	Anderson index of 3 multi-factor leadership items, higher score means higher leadership, based on Avolio and Bass (2014)
Life satisfaction	Rung chosen on Cantril ladder from 0-10, higher rung means higher satisfaction, based on Cantril (1965)
Self-assessment	Assessment of own performance as supervisor on a scale of 1-10 minus assessment of typical supervisor in factory
Ambition	Number of positions nominee would accept a promotion to (supervisor, line chief, assistant production manager)
Muslim	Dummy variable =1 if religion is Islam
Socioeconomic background	Anderson index of variables proxying nominee's socioeconomic background during youth: 1. Father's years of education 2. Mother's years of education 3. Type of house respondent lived in at 14 years of age 4. How often respondent went to bed hungry at 14 years of age
Marriage duration	Completed years of marriage (if married)
Assets into marriage	Dummy variable =1 if respondent brought assets into marriage (if married)
Age at marriage	Age - Marriage duration

Panel B: Variables from diagnostic tests

Numeracy score	Nominee's score on 10 numeracy questions, score out of 1 which represents a perfect score
Literacy score	Nominee's score on 20 literacy questions, score out of 1 which represents a perfect score
Processing speed score	Nominee's score on 193 processing speed questions, based on the Wechsler Adult Intelligence Scale, score out of 1 which represents a perfect score
Garment knowledge score	Nominee's score on 81 garment knowledge questions, score out of 1 which represents a perfect score
Family support score	Nominee's score on 8 family support questions, score out of 1 which represents a perfect score • 5 statements about family support, asks the respondent to respond on a four-point scale from agree to disagree • 3 questions about the level of support given to other women in the family who work in garment factories

(Table 1.C.3 continued.)

Interest score	Nominee's score on 6 interest questions, score out of 1 which represents a perfect score • 2 questions about whether respondent would want to be promoted to supervisor or line chief • 4 questions that indirectly ask whether they are interested in the supervisor position, four-point scale from agree to disagree
Confidence score	Nominee's score on 4 confidence questions, score out of 1 which represents a perfect score • Question how respondent would rate their performance compared to a typical supervisor on a 5-point scale, if they were promoted to supervisor today • 3 forced choice questions between one statement that says "I am confident" using various words, and a dummy statement about the factory

Panel C: Production line statistics

Product complexity	Standard minute value (SMV) of garment. Time that the industrial engineers of each factory estimate a production line with qualified operators working at a standard pace would require to produce one piece of the garment
Efficiency	Measures how the daily output of each line in garment units, adjusted for the standard minute value of the garment, compares to the total output possible that day with the workers present on the line and the hours worked. Calculated as $$efficiency = \frac{(Pieces\,produced \cdot SMV)}{Workers\,on\,the\,line \cdot Hours\,operated \cdot 60} \cdot 100. \quad (1.5)$$
Alteration rate	Percentage of all garments that needs to be altered out of all produced units
Absenteeism rate	Percentage of workers absent amongst all workers (present and absent) on the line

(Table 1.C.3 continued.)

Panel D: Additional outcome variables

Men's assignable expenditures (% in household income)	Percent of household income spent on assignable expenditures for men, consisting of expenditures on 1. Tobacco and alcohol in the last 7 days, 2. Cosmetics for men and boys in the last calendar month, 3. Clothing and footwear for men in the last three months, 4. Jewellery and accessories for men in the last three months, 5. Health and medical expenditures for men in the last three months, Values are winsorised at a fraction of 0.01 in each tail. Expenditures are harmonised on the monthly level, summed up, divided by the household's total monthly income, and multiplied by 100.
Girls'/boys' assignable expenditures (% in household income)	Percent of household income spent on assignable expenditures for girls/boys, consisting of expenditures on 1. Clothing and footwear for girls/boys in the last three months, 2. Educational expenditures for female/male household members or female/male children living outside the household in the last three months, Values are winsorised at a fraction of 0.01 in each tail. Expenditures are harmonised on the monthly level, summed up, divided by the household's total monthly income, and multiplied by 100.
Domestic work	Time spent on domestic work on last working/non-working day.
Leisure	Time spent on leisure on last working/non-working day.
Sleep	Time spent on sleep on last working/non-working day.
Work	Time spent on work on last working day.
Confidence	Anderson index of women's confidence • 10 self-efficacy questions • 1 question whether feeling about self changed in last 5 months, recoded such that 1=Yes, for the better, 0=No, it is the same/Yes, for the worse. • 1 forced choice question about confidence at work, recoded such that =1 if respondent chooses "I hold my head high because I have done really well at work." over "With a bit of luck, things could have turned out better for me at work."

(Table 1.C.3 continued.)

Negotiation	Anderson index of questions measuring negotiation behaviour • 2 questions with vignettes where women were asked in an open-ended way to describe what the woman in the vignette should have done to convince her husband. These are coded as follows: 1=She should have behaved as she did, 2=She should have apologised to the husband, 3=She should have asked nicely again, in a soft tone, 4=She should have explained to him why she wants to make this decision, 5=She should have explained to him why she wants to make this decision, and acknowledged his concerns, 6=She should have explained to him why she wants to make this decision and acknowledged his concerns. She should then try to find a solution that allows for a compromise. • 1 forced choice question, recoded such that =1 if respondent chose "If I really want something, I can easily convince my family." over "At times, I find it hard to explain what I want."
Gender attitudes	Anderson index of questions measuring attitudes about gender. 3 vignettes describing situations in which women exercised mobility and independence in purchases. Respondents were asked to state their personal opinion. Questions will be recoded such that 1=Socially very inappropriate, 2=Socially somewhat inappropriate, 3=Socially somewhat appropriate, 4=Socially very appropriate.
Aspirations	Dummy variable =1 if respondent agrees more with "There is still much more I want to achieve in my work career." over "I am looking forward to enjoying a quieter life back home as soon as I have enough money."

Table 1.C.4: Joint orthogonality of observable characteristics for nominated and selected groups

	(1)	(2)	(3)
Age	0.00	0.01	0.00
	(0.01)	(0.02)	(0.01)
Married	−0.58		−0.06
	(0.57)		(0.10)
Household members	−0.02	−0.01	−0.01
	(0.02)	(0.03)	(0.02)
Household head	0.09	0.15	0.12
	(0.12)	(0.15)	(0.11)
Migrant	0.01	0.01	0.01
	(0.07)	(0.08)	(0.07)
Education years	0.01	0.02	0.01
	(0.02)	(0.02)	(0.02)
Decision-making	−0.02	−0.02	−0.02
	(0.04)	(0.04)	(0.04)
Supportive family	−0.01	−0.02	−0.01
	(0.04)	(0.04)	(0.03)
Experience in garment sector	−0.02	−0.01	−0.02
	(0.01)	(0.01)	(0.01)
Exposure to female SV	−0.05	−0.04	−0.05
	(0.07)	(0.07)	(0.06)
Internal locus of control	−0.08 **	−0.06	−0.06
	(0.04)	(0.04)	(0.04)
Grit	0.05	0.07	0.04
	(0.04)	(0.04)	(0.04)
Self-efficacy	0.05	0.04	0.04
	(0.04)	(0.04)	(0.03)
Emotional competence	−0.01	−0.02	−0.01
	(0.04)	(0.04)	(0.04)
Multi-factor Leadership	−0.02	−0.02	−0.02
	(0.04)	(0.05)	(0.04)
Life satisfaction	0.01	0.01	0.01
	(0.02)	(0.02)	(0.02)
Self-assessment	0.00	0.00	−0.01
	(0.02)	(0.02)	(0.02)
Ambition	0.00	−0.02	−0.01
	(0.04)	(0.04)	(0.04)
Literacy score	0.31	0.17	0.28
	(0.19)	(0.22)	(0.18)
Numeracy score	−0.04	0.03	−0.02
	(0.21)	(0.25)	(0.20)
Processing speed score	−0.49	−0.22	−0.65
	(0.48)	(0.54)	(0.45)
Garment knowledge score	−0.43	−0.44	−0.40
	(0.35)	(0.40)	(0.34)
Family support score	0.14	0.21	0.15
	(0.24)	(0.28)	(0.22)

(Continued on next page.)

(Table 1.C.4 continued from previous page.)

Interest score	−0.03	−0.01	0.03
	(0.19)	(0.22)	(0.18)
Confidence score	0.01	−0.09	0.02
	(0.18)	(0.22)	(0.18)
Muslim	0.07	0.05	0.03
	(0.19)	(0.20)	(0.18)
Socioeconomic background	0.01	0.01	0.00
	(0.03)	(0.03)	(0.03)
Nr Children	0.03	0.04	
	(0.06)	(0.06)	
Combined income of spouses	0.00*	0.00	
	(0.00)	(0.00)	
Education difference	−0.02 *	−0.02 *	
	(0.01)	(0.01)	
Age gap of spouses	0.01	0.01	
	(0.01)	(0.01)	
Income ratio of spouses	0.77**	0.65	
	(0.36)	(0.39)	
Marriage duration	0.00	−0.01	
	(0.01)	(0.01)	
Brought assets in marriage	0.04	0.08	
	(0.07)	(0.08)	
Constant	0.82	0.12	0.98**
	(0.70)	(0.66)	(0.46)
N	198	166	198
p-value of F-statistic	0.87	0.96	0.92
Missing indicators	Yes	No	No
Only married	No	Yes	No

Notes: Standard errors in parentheses. In column (1), to include the variables only defined for married women and include currently unmarried respondents, I follow two steps. First, I set these variables to zero if they are missing and, second, I include dummy variables in the regression that indicate whether the variables are missing. Column (2) does not make these adjustments and hence only includes married respondents. Column (3) includes only covariates defined for all respondents. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 1.C.5: Effects on women's assignable expenditures in Wave 1 and Wave 2 separately

	(1) Wave 1	(2) Wave 2
selected	1.99*	3.22
	(1.16)	(1.98)
N	165	165
Nominated mean	5.21	5.68
Nominated SD	4.95	4.44

Notes: Table only includes respondents interviewed in both waves. Column (1) only includes observations from Wave 1, column (2) from Wave 2. Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.C.6: Effects on time use

Panel A: Women's time use (Hours)

	Last work day				Last non-work day		
	(1) Domestic work	(2) Leisure	(3) Sleep	(4) Work	(5) Domestic work	(6) Leisure	(7) Sleep
selected	0.09	−0.09	−0.20 *	0.19	−0.10	−0.01	−0.07
	(0.13)	(0.16)	(0.11)	(0.27)	(0.13)	(0.23)	(0.18)
N	336	336	336	336	363	363	363
Nominated mean	2.86	1.83	7.02	10.26	4.56	4.00	8.96
Nominated SD	1.02	1.26	1.03	1.86	1.48	2.29	1.22

Panel B: Husband's time use (Hours)

	Last work day				Last non-work day		
	(1) Domestic work	(2) Leisure	(3) Sleep	(4) Work	(5) Domestic work	(6) Leisure	(7) Sleep
selected	0.03	−0.12	−0.13	0.05	0.11	0.05	0.01
	(0.11)	(0.15)	(0.13)	(0.25)	(0.18)	(0.19)	(0.17)
N	330	330	330	330	350	350	350
Nominated mean	1.06	2.92	7.34	10.60	2.33	4.89	9.05
Nominated SD	0.82	1.71	0.84	1.47	1.25	2.47	1.18

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.C.7: Regression discontinuity design—Effects on main outcomes

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)	(6) Controlling money (Taka)
selected	4.94**	6.82*	0.01	−0.16	0.14	−35.46
	(2.00)	(3.98)	(0.08)	(0.44)	(0.24)	(82.44)
N	145	162	180	184	86	89
Nominated mean	4.37	5.45	0.09	−0.14	0.32	146.15
Nominated SD	3.49	7.59	0.29	1.16	0.48	147.60
IK bandwidth	0.24	0.26	0.31	0.33	0.28	0.32

Notes: Standard errors clustered at the individual level in parentheses. Regression discontinuity estimation allows for a linear function of the running variable on either side of the cutoff. Bandwidth selected following Imbens and Kalyanaraman (2012).

Table 1.C.8: Matching—Effects on main outcomes

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)	(6) Controlling money (Taka)
selected	3.98***	1.85	−0.02	0.32	0.07	9.89
	(1.00)	(1.25)	(0.04)	(0.20)	(0.14)	(30.12)
N	363	362	363	363	181	182
Nominated mean	5.32	6.00	0.09	−0.16	0.28	146.77
Nominated SD	4.48	8.63	0.29	1.13	0.45	147.72

Notes: Heteroskedasticity-robust standard errors in parentheses. One-to-one nearest-neighbour matching with replacement implemented on all balance variables in Table 1.5 as well as factory, enumerator and month dummies. The weighting matrix for matching variables is the diagonal matrix of the inverse sample standard errors.

Table 1.C.9: Different variable definitions—Robustness for women's expenditures and remittances

Panel A: Women's assignable expenditures

	(1)	(2)	(3)
	% in HH income	IHS of % in HH income	Taka
selected	2.71**	0.19	395.59 **
	(1.07)	(0.13)	(186.12)
N	363	363	363
Nominated mean	5.32	2.07	1201.99
Nominated SD	4.48	0.83	1050.47

Panel B: Total remittances

	(1)	(2)	(3)
	% in HH income	IHS of % in HH income	Taka
selected	3.48**	0.30	954.15 ***
	(1.40)	(0.24)	(324.62)
N	362	362	362
Nominated mean	6.00	1.43	1305.97
Nominated SD	8.63	1.60	1986.61

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies. IHS is the inverse hyperbolic sine transformation.

Table 1.C.10: Different specifications—Robustness for women's expenditures and remittances

Panel A: Women's assignable goods expenditures (% in HH income)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
selected	2.71**	3.01***	2.63*	2.68**	2.67*	1.58	2.48**	2.01**
	(1.07)	(1.10)	(1.51)	(1.06)	(1.40)	(1.73)	(1.18)	(1.00)
N	363	363	363	363	363	363	354	363
Nominee mean	5.32	5.32	5.32	5.32	5.32	5.32	5.34	5.32
Nominee SD	4.48	4.48	4.48	4.48	4.48	4.48	4.51	4.48
Method	PDS	OLS	OLS	OLS	OLS	OLS	OLS	PDS
Controls	PDS	No	No	No	Linear	Linear	Theory	PDS with interactions
Factory F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Enumerator F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Month F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Imbalanced controls	PDS	No	No	Yes	No	Yes	No	No

Panel B: Remittances (% in HH income)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
selected	3.48**	3.47**	3.32**	3.13**	3.35**	4.13**	2.63*	3.48**
	(1.40)	(1.37)	(1.54)	(1.34)	(1.41)	(1.63)	(1.55)	(1.42)
N	362	362	362	362	362	362	353	362
Nominee mean	6.00	6.00	6.00	6.00	6.00	6.00	5.97	6.00
Nominee SD	8.63	8.63	8.63	8.63	8.63	8.63	8.70	8.63
Method	PDS	OLS	OLS	OLS	OLS	OLS	OLS	PDS
Controls	PDS	No	No	No	Linear	Linear	Theory	PDS with interactions
Factory F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Enumerator F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Month F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Imbalanced controls	PDS	No	No	Yes	No	Yes	No	No

Notes: Standard errors clustered at the individual level in parentheses.

Table 1.C.11: Husband's survey—Robustness for main outcomes

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Wife's decision-making (Index)	(4) Husband's Hiding money (Dummy)	(5) Husband's controlling money (Taka)
selected	2.21*	5.13***	−0.05	−0.03	−26.07
	(1.14)	(1.50)	(0.10)	(0.07)	(18.74)
N	350	349	349	173	175
Nominated mean	5.71	6.31	0.09	0.19	134.48
Nominated SD	4.52	9.39	0.65	0.40	131.68

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from the wife's demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.C.12: Phone surveys—Effects on women's and girls' assignable goods expenditures

	(1) Women's and girls' expenditures (Taka)	(2) Remittances to wife's family (Taka)
selected	657.55 **	1008.50 **
	(290.91)	(479.94)
N	1395	1390
Nominated mean	1781.96	2361.05
Nominated SD	2562.26	4787.86

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and survey round dummies.

Table 1.C.13: Placebo test—Comparing top and bottom half of selected group

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)	(6) Controlling money (Taka)
placebo	0.31	0.60	0.01	0.04	−0.11	−46.06 **
	(1.93)	(1.62)	(0.03)	(0.11)	(0.08)	(19.23)
N	296	295	296	296	145	151
Placebo control mean	8.13	9.25	0.06	−0.14	0.37	120.42
Placebo control SD	14.83	12.91	0.23	1.02	0.49	154.36

Notes: Table only includes women selected for the promotion programme. Standard errors clustered at the individual level in parentheses. Sharpened q-values in squared brackets. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.C.14: Attrition—Robustness for women's assignable expenditures and remittances

	Balanced panel		Wave 1 only	
	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Women's expenditures (% in HH income)	(4) Remittances (% in HH income)
selected	2.49**	2.93*	1.92*	3.08*
	(1.15)	(1.50)	(1.02)	(1.68)
N	330	329	181	181
Nominated mean	5.45	6.00	5.19	5.08
Nominated SD	4.67	8.89	4.62	8.22

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.C.15: Training effects—Robustness for women's assignable expenditures and remittances

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)
selected	3.32**	3.29*
	(1.43)	(1.71)
selected*softonly	−2.60	0.80
	(1.77)	(1.92)
selected*hardsoft	0.78	−0.22
	(2.68)	(1.96)
N	363	362
Nominated mean	5.32	6.00
Nominated SD	4.48	8.63
p(selected*softonly= selected*hardsoft=0)	0.13	0.87

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.C.16: Selection effects – Robustness of main outcomes for factories not in the selection experiment

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)	(6) Controlling money (Taka)
selected	3.20	4.31*	0.02	−0.23	0.03	−79.86 ***
	(2.18)	(2.31)	(0.05)	(0.26)	(0.17)	(21.42)
N	160	159	160	160	78	82
Nominated mean	6.98	5.23	0.05	0.22	0.30	240.91
Nominated SD	5.85	10.20	0.22	1.14	0.48	102.02

Notes: Standard errors clustered at the individual level in parentheses. Controls chosen using pdslasso from demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies. Table only includes factories that did not participate in the selection experiment.

Table 1.C.17: Exposure analysis: Joint orthogonality of observable characteristics for exposed and non-exposed groups

	(1)	(2)	(3)
Age	0.01	0.02	0.00
	(0.01)	(0.01)	(0.01)
Married	0.40**		0.15*
	(0.17)		(0.09)
Household members	0.01	0.00	0.01
	(0.02)	(0.02)	(0.02)
Migrant	−0.06	−0.05	−0.06
	(0.05)	(0.05)	(0.05)
Education years	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)
HH income	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Muslim	−0.34	−0.53 **	−0.31
	(0.23)	(0.25)	(0.23)
Socioeconomic background	−0.02	−0.03	−0.03
	(0.02)	(0.03)	(0.02)
Household head	0.13	0.07	0.21
	(0.17)	(0.30)	(0.17)
Wife's share in HH income	−0.04	−0.17	−0.03
	(0.14)	(0.16)	(0.14)
Nr Children	0.01	−0.01	
	(0.04)	(0.05)	
Age gap of spouses	0.00	0.00	
	(0.01)	(0.01)	
Marriage duration	0.00	−0.01	
	(0.01)	(0.01)	
Brought assets in marriage	−0.12 **	−0.11 *	
	(0.06)	(0.06)	
Spousal education difference	0.00	0.00	
	(0.01)	(0.01)	
Constant	0.35	0.76**	0.55*
	(0.36)	(0.38)	(0.31)
N	488	427	488
p-value of F-statistic	0.41	0.29	0.33
Missing indicators	Yes	No	No
Only married	No	Yes	No

Notes: Standard errors in parentheses. In column (1), to include the variables only defined for married women and include currently unmarried respondents, I follow two steps. First, I set these variables to zero if they are missing and, second, I include dummy variables in the regression that indicate whether the variables are missing. Column (2) does not make these adjustments and hence only includes married respondents. Column (3) includes only covariates defined for all respondents. * p<0.1, ** p<0.05, *** p<0.01

Table 1.C.18: Exposure analysis: Effects on time use

Panel A: Women's time use (Hours)

	Last work day				Last non-work day		
	(1) Domestic work	(2) Leisure	(3) Sleep	(4) Work	(5) Domestic work	(6) Leisure	(7) Sleep
exposed	0.08	0.03	−0.16 *	−0.05	−0.16 *	0.12	0.09
	(0.08)	(0.08)	(0.10)	(0.18)	(0.10)	(0.14)	(0.10)
N	465	465	465	465	488	488	488
Non-exposed mean	3.17	1.57	6.85	9.80	4.58	3.27	8.22
Non-exposed SD	0.97	0.98	1.08	1.20	1.06	1.80	1.31

Panel B: Husband's time use (Hours)

	Last work day				Last non-work day		
	(1) Domestic work	(2) Leisure	(3) Sleep	(4) Work	(5) Domestic work	(6) Leisure	(7) Sleep
exposed	0.05	−0.19 ***	0.03	0.16	0.20***	−0.11	−0.20 *
	(0.08)	(0.07)	(0.08)	(0.18)	(0.08)	(0.08)	(0.10)
N	414	414	414	414	449	449	449
Non-exposed mean	0.98	2.59	7.31	9.93	2.04	4.19	8.86
Non-exposed SD	0.64	1.32	0.75	1.37	0.94	1.99	0.93

Notes: Standard errors clustered at the production line level in parentheses. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory and enumerator dummies.

Table 1.C.19: Exposure analysis: Effects on components of decision-making with different definitions

Panel A: Involved in decision (Dummy)

	(1) Visit a friend	(2) Take a bus	(3) Purchase clothing	(4) House repairs	(5) Purchase appliances	(6) Accept promotion	(7) Take up work
exposed	0.00	0.07**	0.02**	0.07**	0.01	0.02**	0.01*
	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)
N	488	488	488	488	488	488	488
Non-exposed mean	0.99	0.89	0.98	0.86	0.95	0.98	0.99
Non-exposed SD	0.11	0.32	0.13	0.35	0.21	0.14	0.11

Panel B: Sole decision (Dummy)

	(1) Visit a friend	(2) Take a bus	(3) Purchase clothing	(4) House repairs	(5) Purchase appliances	(6) Accept promotion	(7) Take up work
exposed	0.05	0.07**	0.07*	0.01	0.02	0.00	−0.02
	(0.03)	(0.03)	(0.04)	(0.01)	(0.01)	(0.02)	(0.02)
N	488	488	488	488	488	488	488
Non-exposed mean	0.10	0.08	0.17	0.01	0.01	0.24	0.06
Non-exposed SD	0.30	0.27	0.37	0.11	0.09	0.43	0.24

Panel C: Joint decision (Dummy)

	(1) Visit a friend	(2) Take a bus	(3) Purchase clothing	(4) House repairs	(5) Purchase appliances	(6) Accept promotion	(7) Take up work
exposed	−0.12 ***	−0.05	−0.08 **	0.05	0.00	0.01	0.00
	(0.03)	(0.04)	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)
N	488	488	488	488	488	488	488
Non-exposed mean	0.54	0.75	0.77	0.84	0.95	0.66	0.91
Non-exposed SD	0.50	0.44	0.42	0.36	0.23	0.48	0.29

Notes: Standard errors clustered at the production line level in parentheses. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.C.20: Exposure analysis: Matching—Main outcomes

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)
exposed	−0.36	0.12	0.04**	0.24***	0.06**
	(0.63)	(1.30)	(0.02)	(0.07)	(0.03)
N	488	483	482	488	488
Non-exposed mean	6.48	8.49	0.02	−0.26	0.12
Non-exposed SD	6.63	11.81	0.13	1.08	0.33

Notes: Heteroskedasticity-robust standard errors in parentheses. One-to-one nearest-neighbour matching with replacement implemented on all balance variables in Table 1.15 as well as factory, enumerator and month dummies. The weighting matrix for matching variables is the diagonal matrix of the inverse sample standard errors.

Table 1.C.21: Exposure analysis: Different specifications—Robustness for women's involvement in decision-making

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
exposed	0.36***	0.23**	0.36***	0.26***	0.21**	0.32***	0.32***	0.30***
	(0.09)	(0.09)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)
N	488	488	488	488	488	488	488	488
Non-exposed mean	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26	−0.26
Non-exposed SD	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08
Method	PDS	OLS	OLS	OLS	OLS	OLS	OLS	PDS
Controls	PDS	No	No	No	Linear	Linear	Theory	PDS with interactions
Factory F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Enumerator F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Month F.E.	PDS	No	Yes	No	No	Yes	Yes	PDS
Imbalanced controls	PDS	No	No	Yes	No	Yes	No	PDS

Notes: Standard errors clustered at the production line level in parentheses.

Table 1.C.22: Exposure analysis: Husband's survey—Effects on main outcomes

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Women's decision-making (Index)	(4) Husband's Hiding money (Dummy)
exposed	0.41	−0.78	0.05	0.08*
	(0.89)	(1.28)	(0.10)	(0.04)
N	449	449	449	449
Non-exposed mean	6.55	9.03	0.07	0.12
Non-exposed SD	6.51	12.80	1.18	0.33

Notes: Standard errors clustered at the production line level in parentheses. Controls chosen using pdslasso from the wife's demographic controls and baseline bargaining measures, their squares, and factory, enumerator and month dummies.

Table 1.C.23: Exposure analysis: Operator sorting—Effect on main outcomes for subsample on production line before programme

	(1) Women's expenditures (% in HH income)	(2) Remittances (% in HH income)	(3) Violence (5 months, Dummy)	(4) Decision-making (Index)	(5) Hiding money (Dummy)
exposed	−1.00	−0.25	0.02	0.38***	0.02
	(0.70)	(1.34)	(0.02)	(0.09)	(0.04)
N	354	352	351	354	354
Non-exposed mean	6.48	8.49	0.02	−0.26	0.12
Non-exposed SD	6.63	11.81	0.13	1.08	0.33

Notes: Table excludes the exposed women who joined the production line after the start of the promotion programme. Standard errors clustered at the production line level in parentheses. Controls chosen using pdslasso from demographic controls and bargaining measures, their squares, and factory, enumerator and month dummies.

Chapter 2

Learning How to Choose or Learning How to Lead? Experiments on Selecting and Training Female Managers in Bangladesh's Garment Industry

with Christopher Woodruff, University of Oxford

Abstract We report the results of field experiments designed to understand the importance of the selection of and training for new female supervisors in Bangladesh's garment factories. Participating factories have little prior experience with promoting women. We show that formal diagnostic tests lead factories to select candidates that are more likely to be promoted and who, according to their subordinates, perform better as supervisors. Diagnostics measuring attitudes and soft skills are particularly relevant for factories and predictive of later outcomes. Supervisory training for the selected candidates leads to higher rates of promotions, but has only marginal effects on performance. In none of our results do we find that training in technical skills has an additional effect when compared to a training that focuses on attitudes and soft skills. These results indicate the importance of hard measures of soft skills and attitudes in the process of selecting female supervisors, and suggest that training in non-cognitive skills could be a promising avenue to increase the participation of women in managerial positions.

2.1 Introduction

Lazear et al. (2015) find that replacing a supervisor in the 10^{th} percentile with one in the 90^{th} percentile increases output of a work team by more than 10 %. However, identifying the best candidates for supervisory positions is challenging. Benson et al. (2019), using detailed internal records from a large US firm, show that the best workers do not necessarily make the most effective supervisors. For higher-level managers, making promotion decisions is essentially a prediction problem. The difficulty is to identify, measure, and base promotions upon characteristics that are correlated with more successful performance after promotion. Evidence from similar settings suggests that this type of prediction exercise is hard (McKenzie and Sansone, 2017; Fafchamps and Woodruff, 2017). Firms may focus on a suboptimal set of characteristics when selecting employees or leaders, especially when they have limited prior experience in making these decisions (Ashraf et al., 2020a; Bandiera et al., 2020).

The challenge of selecting the best supervisors arises at least in part because, compared with workers, managers carry out a much more diverse and complex set of tasks. This increase in complexity and variety of tasks implies that new managers need to learn new skills, either on-the-job or through training. However, moulding managers through education is fraught with difficulties. Training programs for managers and entrepreneurs across a wide range of contexts have produced only limited improvement in firm outcomes (McKenzie and Woodruff, 2014).

We conduct experiments to understand the importance of manager selection and training in a particularly challenging context: promoting women to supervisors in Bangladesh's garment industry. Garment factories in Bangladesh are increasingly interested in promoting women. The factories face increasing competition for male employees in domestic labour markets and pressure from international brands to provide opportunities for women. More than three quarters of workers in the sewing sections of factories are women, but female supervisors remain very rare; less than 10 % of supervisors are female (Menzel and Woodruff, 2019).

Previous research highlights several factors which suggest that factories have limited information both on how to select and on how to train the best female candidates for supervisory positions. Candidates for promotions in the factories are typically recommended by existing supervisors and higher-level managers. Beaman et al. (2018) report experimental evidence that men are more likely to refer other men for jobs, even when they are aware of better-qualified women. Moreover, the characteristics of the best female candidates may differ from those of the most qualified males, which impedes learning. Managers may overlook "obvious" characteristics because they lack experience in choosing female candidates. In addition, the feedback to decision-makers whether the decisions were correct is slow. In this context, all new supervisors typically spend many weeks in a trial period, and then often underperform initially before growing into the job. Outside of this on-the-job-training during a trial period, formal training for new supervisors is rare. Macchiavello et al. (2020) show that formal training can increase the share of female supervisors in Bangladeshi garment factories, who perform as well as new male supervisors after an adjustment period.

With this in mind, we conduct experiments in two acts with 27 large garment factories in Bangladesh. We aim to investigate how predictive initial skill diagnostics are for the success of new female supervisors, and whether a supervisory training can produce successful supervisors. The factories were asked to nominate women for a training program designed to prepare them for supervisory positions. At the beginning of the project, one-third of the participating factories had no female supervisors, and there was no factory where more than 20 % of supervisors were female. We asked managers to rank the candidates in order of their expected performance as supervisors.

In the first intervention, we implemented a selection experiment with higher-level management staff in a randomly selected half of the factories. The participants of the selection experiment received a two-hour session introducing them to diagnostic tests as a means of selecting candidates for promotions. At the end of the training, managers were shown scores from seven diagnostic tests conducted with the women they had nominated for the supervisory training program. Four of the diagnostic tests related to cognitive ability or

technical skills: tests of literacy, numeracy, fluid intelligence and garments knowledge. We aggregate these scores into an index of initial aptitude. The other three scores were measures of non-cognitive skills and attitudes that previous research showed were important to the success of female supervisors: confidence in the ability to work as a supervisor, interest in the position, and support from the family for being a supervisor.[1] We aggregate these into an initial attitude index. The attitude index therefore combines aspects of non-cognitive skills which are often termed soft skills in economics (Heckman and Kautz, 2012), as well as attitudes of the individual and the family that quantitative and qualitative research has shown to be predictive of women's promotion rates in this context (Macchiavello et al., 2020). After reviewing the diagnostic scores for all of the factory's nominees, the managers that participated in the selection experiment were invited to re-rank the candidates if they desired to do so. The factories that did not participate in the selection experiment were not able to see their nominees' diagnostic scores nor could they change their ranking from the one they had initially provided.

In the second intervention, we provided supervisory training to the selected trainees in all factories. We randomised those selected for training into three groups. The first received a four-day training program focusing on attitudes and soft skills – stress management, assertive communication and leadership – and a five-day training programme in aptitude and hard skills required of a successful supervisor – line balancing, quality control, etc. – immediately. The second group received the attitude training immediately and the aptitude training about six months later. The third group received both training modules six months later. All of the nominees were assigned to trial immediately as assistant supervisors for at least two months, which is the status quo method of training candidates for supervisory positions.[2]

[1]Data from the supervisor training programme in Macchiavello et al. (2020) show that baseline interest in promotion to supervisor is associated with completion of the training program (rather than dropping out), and continuing to work as a supervisor at the time of the final survey. The claim on family support is based on qualitative research with factory management. None of the factories reported using similar standardised tests in their promotion decisions prior to the project. Qualitative research suggests that, in the absence of the project, the identification of potential new managers is an informal process. For example, it appears to be common that established line supervisors suggest workers from their lines for supervisor vacancies, often those who have shown leadership ability in the group or who have approached the supervisors to express their interest in a promotion.

[2]Of course, the trainees either in the status quo or in our programme may drop off of the supervisory ladder

Using the variation created by these experiments, we document four results with a combination of survey and administrative data. First, factories do not take into account the initial attitude and aptitude measures of candidates when they nominate women for the promotion training. That is, the initial ranking of nominees before the diagnostic tests are taken is orthogonal to the initial attitude and aptitude indices we measure. Second, the factories that are provided information about their candidates' skills and can change their ranking as part of the selection experiment react to this information. They select candidates who have higher baseline skills, especially on the attitude dimension. We posit that the attitude diagnostics have a large effect on managers' ranking because they are characteristics that are likely to be less relevant for male supervisory candidates. Similar to Hanna et al. (2014), the attitude diagnostics lead the managers to notice characteristics that they otherwise are not conditioned to notice. This also ties in with evidence on self-selection of workers into jobs. For example, Ashraf et al. (2020a) find that emphasising career prospects in advertisements for health worker positions in Zambia leads to a more talented but less pro-social applicant pool, and better-performing hires.

Third, we find that the initial skills – and especially attitudes and soft skills – matter for the outcomes of the promotion programme. They are predictive of a gain in skills, the promotion rates to official line supervisor, higher wellbeing of their subordinates and – to some degree – higher performance evaluations as judged by subordinates. We do not find that they matter for production outcomes such as efficiency and alteration rates. Fourth, attitude and aptitude training for the participants of the supervisory training leads to some improvements in skills, promotions, and performance, though the results are not always significant. In none of our results do we find that training in technical skills has an additional effect when compared to a training that only focuses on attitudes and soft skills.

The evidence we present suggests that both the selection process and the supervisory training matters. We find that initial skills predict the new female supervisors' success, and observe some evidence that training can make the supervisors more successful. For the new female low-level managers that are the focus of this paper, attitudes and soft skills appear

either because they themselves decide they do not want to be a supervisor, or because managers decided they are not qualified for the position.

to be particularly important. The results of the selection experiment suggest that factories had some idea that attitudes and soft skills would matter for their trainees' success, but the little experience in choosing female supervisors likely meant that they lacked the ability to convincingly measure these skills. In the case of female career advancement, this paper therefore not only provides evidence for the importance of hard measures of soft skills and attitudes, but also indicates that training in non-cognitive skills is a promising avenue to increase the participation of women in managerial positions.

In addition to the potential to increase the pool of management talent in factories, increasing the rate of promotion of women may also have transformative effects on women's lives. Female labour force participation generates changes in development dynamics in lower-income countries, increasing female empowerment and educational attainment of children (Heath and Jayachandran, 2018; Duflo, 2012; Qian, 2008; Heath and Mobarak, 2015). The garment sector provided an entry into wage work for women in Bangladesh. Even in 2017, garment factories were the workplace for 40 % of female wage workers with less than tertiary education. As in other countries, employment in the sector is responsible for higher educational attainment and later age of marriage for women (Heath and Mobarak, 2015). Chapter 1 examines the impact of a promotion on the women's position in the household, both for the women selected for training and for those working under their direction, and finds significant increases in the women's bargaining power.

The paper proceeds as follows. We discuss the contribution to the literature in Section 2.2, and explain the two interventions in Section 2.3. We discuss our empirical strategy and present results in Section 2.4. Section 2.5 discusses the results and concludes.

2.2 Contribution to the literature

This paper is contributing to the literature that describes challenges in worker and manager selection, and investigates whether and how managerial success can be predicted. For example, Benson et al. (2019) for US firms find that prioritising current job performance over observable characteristics that predict managerial success in promotion decisions entails high costs – which, however, could be justified by the benefits of promotion-based

incentives for workers. In a similar vein, Hoffman et al. (2018) and Autor and Scarborough (2008), again in the US, find that using skills testing in recruitment decisions increases the quality of new hires compared to relying on human judgement.

In low- and middle-income countries, however, the literature on managerial selection has largely focused on managers of micro and small enterprises, which represent the vast majority of firms in these economies.[3] Fafchamps and Woodruff (2017) in Ghana, McKenzie and Sansone (2017) for Nigeria, and Hussam et al. (2020) in India find that baseline survey data of entrepreneurs or firms can predict firm growth. Even though these key characteristics outperform the predictions of expert judges in business competitions in both Fafchamps and Woodruff (2017) and McKenzie and Sansone (2017), Hussam et al. (2020) observe that truthfully reported peer predictions can outperform baseline characteristics. As Quinn and Woodruff (2019) point out, hard skills of entrepreneurs are typically found to be more important than soft skills in this strand of the literature. Overall, however, the variation in future outcomes explained by the characteristics the authors look at is small.

Compared to the literature on managers of micro or small enterprises, our extensive diagnostic tests allow us to focus on predicting success for a very different set of managers. Line supervisors are the lowest level of managers in highly hierarchical, large garment factories. Compared to the owner-managers of micro and small firms discussed in most of the literature, who are responsible for leading and running every aspect of their enterprise, line supervisors have very different responsibilities. As will be further discussed below, line supervisors are assigned to a production line or part of a production line and are tasked with ensuring that their line is running well and meeting its production target. This typically involves motivating the subordinate line operators, solving simple problems on the line, and communicating more complex problems to higher-level managers.

Compared to the literature on manager selection in high-income countries, we present the first experimental evidence on the selection of managers, by allowing a random subsample of factories to reconsider the selection of trainees for a promotion programme after present-

[3] See Quinn and Woodruff (2019) for a recent overview of experimental work on entrepreneurship in developing countries.

ing the results of the diagnostic tests to them. This allows us to test whether factories are well-informed about candidates for management positions in advance, and whether they respond to the baseline data presented to them. If we do see a response, we can analyse whether factories respond to the characteristics that we identify as predictive.

We also contribute to a large literature on managerial training in low- and middle-income countries, which has found little evidence that standard business trainings for micro and small firms have large impacts on business performance or growth. This literature is reviewed in detail in Quinn and Woodruff (2019) and McKenzie and Woodruff (2014). While there is a lot of variation in the content and the length of the trainings investigated in the literature, they typically aim to improve entrepreneurs' business practices, mainly by teaching hard skills such as accounting, financial planning, or marketing. These practices have been shown to be correlated with business performance for both small and large firms.[4] However, the typical business training does not appear to lead to large improvements in business practices over the longer term, which likely explains their limited effects on business performance or growth (Quinn and Woodruff, 2019).

Two newer approaches have shown more promise and are now being researched more widely. First, consulting services tailored to each firm's circumstances have shown success for small, medium and large firms. Bloom et al. (2013) offered intensive consulting services to large Indian textile plants. They found improvements in management practices and productivity in the first year, and an increase in new plant openings after three years, compared to a control group. Similarly, Bruhn et al. (2018) find that providing consulting to micro and small firms in Mexico leads to large increases in employment in the long term and some positive effects on firm profitability in the short term.

Second, trainings focusing on soft rather than hard skills have shown potential. For example, Campos et al. (2017) and Glaub et al. (2014) investigate so-called personal initiative trainings in Togo and Uganda, which aim to develop a proactive mindset by teaching "self-starting behavio[u]r, innovation, identifying and exploiting new opportunities, goal-setting, planning and feed-back cycles, and overcoming obstacles" (Campos et al., 2017).

[4]See, for example, McKenzie and Woodruff (2017) and Bloom et al. (2014b).

Compared to a standard business training, these papers show that personal initiative trainings have larger, positive effects on profits, input use, product diversification and access to finance.

By assigning women to different training groups, the supervisory training we implemented allows us to investigate the impacts of training in attitudes and soft skills only and the additional effects of training in technical skills. The training contents focused on the specific skills needed in a supervisor position in the garment industry (e.g. communication skills, managing bottlenecks on a production line), as compared to the standard business skills (e.g. accounting, financial planning) contained in the trainings described above. It is a shortened and improved version of the training in Macchiavello et al. (2020). However, Macchiavello et al. (2020) compare between female and male candidates for promotion and conduct attitude and aptitude training for all selected candidates. This paper focuses on female candidates and randomly assigns the training group to distinguish between the effects of building attitude versus aptitude.

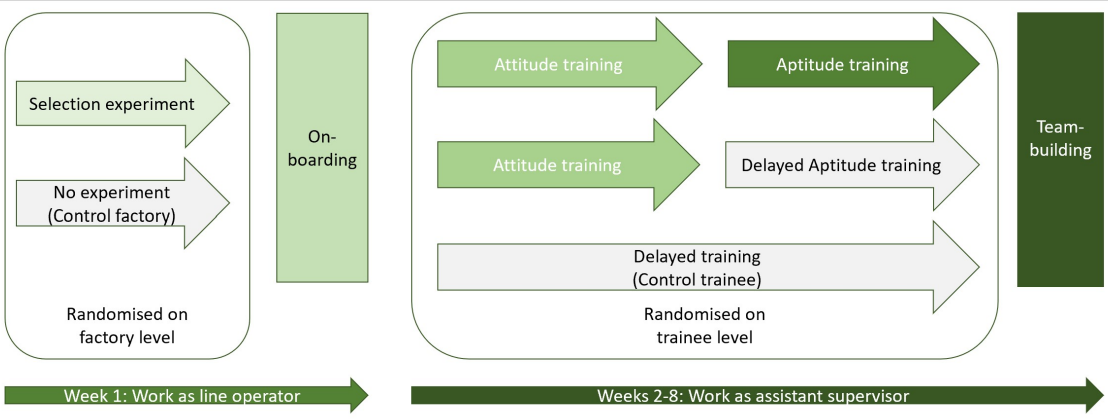
The information on the effects of attitude and aptitude training on official promotions, skills, performance and production outcomes enable us to conduct a horse-race type analysis between baseline skills and a training in the same skills for the female candidates. We focus on female candidates for line supervisor positions since women continue to be underrepresented in managerial roles in the Bangladeshi garment industry (Menzel and Woodruff, 2019). Women only represent about 5 % of all supervisors in the industry (Macchiavello et al., 2020). The evidence in this paper therefore helps to illuminate how women's participation in management positions in this labour-intensive manufacturing industry could be improved – by better selection of suitable female candidates according to skills that predict managerial success, or by training women in these skills.[5]

[5]With this, we also contribute to the literature on female leadership, which is discussed in detail in Chapter 1.

2.3 Study design

We implemented two distinct interventions sequentially. In the first, we conducted extensive diagnostic tests with all candidates for a promotion programme and implemented a selection experiment. During the first experiment, a random subsample of factories was able to reconsider their candidate selection based on the results of the diagnostic tests. In the second intervention, we trained female operators to become line supervisors. Figure 2.1 provides an overview of the design.

Figure 2.1: Overview of study design



At the beginning of the project, we determined the number of operators to train in each factory in conversations with the management of the 30 factories that initially agreed to participate. We advised them to select this number based on the number of openings for supervisory positions they expected to have in the following few months.[6] We then asked managers to nominate 1.67 times the number of workers they intended to train, and to rank all nominees based on their expected performance as supervisors. Thus, a factory that intended to train six women nominated ten for the programme, and provided us with a ranked list of the ten candidates. A smaller rank indicated a higher expected performance as supervisor.

Approximately two weeks after receiving this information from the factory, we conducted

[6]The variance in this expected demand led to variance in the number of trainees across factories. For example, some factories anticipated opening new production lines, and hence anticipated the need for a larger number of new supervisors.

baseline surveys and diagnostics with all nominees. The survey elicited information on education, labour history, and other demographic data from each of the nominees. We also asked questions about relationships in their household, including measures of their participation in household decision-making.

For each nominee in all factories, we conducted diagnostic tests with seven components. The diagnostic, described in more detail in Appendix 2.A, included tests of literacy, numeracy, fluid intelligence (measured by processing speed tests from the Wechsler Adult Intelligence Test), and technical knowledge related to garments production. In the analysis, we show results for the individual diagnostics and for a standardised index following Anderson (2008) that combines these four diagnostics into a single index measuring aptitude. We also asked questions which we group into three scores reflecting non-cognitive skills and attitudes: a measure of how much the nominee's family would support her working as a supervisor, a measure of the nominee's own interest in the position, and a measure of the nominee's confidence. In the analysis, we will group these three diagnostics into a single attitude index following Anderson (2008).[7]

Our choice of diagnostic tests was informed by previous work in Macchiavello et al. (2020) and by qualitative research with factory management. Taking first the three measures of attitude, we included the measures of confidence and interest because of earlier evidence indicating that female supervisor trainees had lower confidence levels than male trainees, and that the training closed this confidence gap (Macchiavello et al., 2020). Discussions with factory managers and focus groups research also revealed the importance of family support. Regarding the aptitude measures, previous work reports data from baseline surveys of workers at all level of the factory showing a wide-spread belief that men are better at "Understanding machines: which machines are appropriate for which tasks, and knowing when machines are not functioning properly" (Macchiavello et al., 2020). The garments

[7]The indices are computed as described in Footnote 31. Following Anderson (2008), the variables entering the index are first standardised and the index is then created as a weighted sum of the standardised variables, where the weights are the inverse of the covariance matrix of the standardised variables. This approach entails that variables which have comparatively low correlation with the other variables in the index have a relatively higher weight. Note that the indices were not made available to managers, only the seven diagnostic scores as a percentage out of 100 %. Therefore, the index construction could not have affected the ranking. We use the indices as summary measures of attitude and aptitude for the main analysis, but also provide results with the seven diagnostic scores in the appendix.

knowledge diagnostic measures exactly this. Literacy was included in the diagnostics because it is required for successful participation in the training. Discussions with factory managers indicated that numeracy and fluid reasoning are widely believed to be important for supervisor performance. Note that we do not claim that these diagnostic measures are an exhaustive list of either aptitude or attitude. The measures almost certainly could be improved.

2.3.1 Selection experiment

Our first intervention allowed factory managers to use these quantitative diagnostic tools to select the best candidates for the promotion training. We refer to this as the "selection experiment". Factories were randomised into two equally sized groups.[8] The first received a two-hour introduction to the diagnostic tests, and the second served as a control group without any further intervention. The intervention in the treated factories provided data from previous projects that supported the case for promoting female supervisors. It also introduced the managers to the seven different diagnostic scores and allowed them to take part of the test themselves. This session was aimed at mid- and high-level managers – those involved in nominating the female operators and in making decisions about whether they would be promoted.

None of the factories reported using similar diagnostics in their promotion decisions prior to the management training. At the end of the session, managers were shown the scores for

[8]The intervention took place sequentially, as did the randomisation for the selection experiment. 26 of the initial factories were part of the International Labour Organizations BetterWork programme. For the BetterWork factories, the randomisation method was a stratified randomisation with rebalancing in two sequences. We used two stratification variables, 1) participation in a previous training project described in Macchiavello et al. (2020) and 2) a dummy for producing only knit garments. We re-balanced on three re-randomisation variables: 1) the number of workers in a factory, 2) the share of female workers and 3) the date of joining the BetterWork project. This information came from administrative data provided by BetterWork. We conducted the randomisation 1000 times in the first and 20 times for the second, smaller sequence. For each iteration, we identified the minimum p-value of the three two-sided t-tests of mean equality for the balancing variables between treatment and control (for the second sequence, the balance tests included all factories, including those assigned in the previous sequence). We re-randomised according to a clear criterion: Out of the assignments with a minimum p-value of >0.20, we randomly chose one. For the four participating non-BetterWork factories, we implemented a simple randomisation and allocated half to treatment and half to control. This, again, took place in two sequences. In Table 2.B.1 in the appendix, we show randomisation checks for the selection experiment for a wide range of factory characteristics and the average diagnostic scores of nominees. Out of 22 variables for which we test balance, we find one rejection using the p-value based on randomisation inference in the last column. This is fewer than one would expect to find by chance.

the seven diagnostics that we conducted with the nominees at the same time as the session with managers was run. Managers from factories that were randomised to the selection experiment were offered the opportunity to revise their rankings of the operators they had nominated for the supervisor training program.

For the factories that were in the control group for the selection experiment, the original ranking of nominees was changed only if nominees were ruled ineligible due to low scores on the literacy or numeracy tests (see Appendix 2.A). In that case, all nominees were moved up in the original order. The nominees with the best ranks in the final ranking were selected for the supervisory training, up to the number determined in advance with the factory.

2.3.2 Promotion training

The best-ranked operators represented the pool for the second intervention, which was aimed at training the selected operators to be line supervisors. The training program was an improved and streamlined version of the program developed initially by GIZ and also used in Macchiavello et al. (2020). For this project, the training was reduced to nine days, divided into four days of training on attitudes and soft skills and five days on apttitude or hard skills. The attitude training focused on stress management, assertive communication, and leadership, elements which earlier research identified as being particularly important for female supervisors (Macchiavello et al., 2020).[9] The aptitude training focused on the technical skills required for line supervisors, e.g. production processes, sewing machines, quality control, cutting, finishing, printing, embroidery, and the responsibilities of the supervisory role. Appendix 1.A of Chapter 1 presents more details.[10]

In each factory, we conducted a public lottery to randomly allocate trainees into one of three groups. The first received both the attitude and aptitude training immediately, while

[9]This part of the training also included sessions on understanding harassment, developing integrity and fairness, workers' rights and responsibilities, and human resources management, including types of management styles.

[10]Compared to the full six-weeks GIZ training programme, the attitude and aptitude training in this project puts relatively more weight on attitude training. We chose to emphasise the attitude training because earlier work showed that the biggest effect of the GIZ training was to increase the confidence level of female trainees (Macchiavello et al., 2020). Of course, it is possible that the increased confidence comes as a result of an increase in technical skills, but we chose to focus the attention on building confidence more directly.

the second group received only the attitude training immediately (and the aptitude training around six months later). The third did not receive any training initially and functions as our control group; they received both the attitude and aptitude training around six months later.[11] The training sessions were held on consecutive weeks at a local training centre on two days per six-day work week.[12]

In addition to the classroom-based training, all trainees were assigned to work as assistant supervisors on the line for two months from the start of the programme. We asked factories to choose a set of trial lines, and to assign the selected trainees to those lines. These choices were made before anyone was aware of which operators would be randomised into which of the three training groups. Hence, while the assignment of individual trainees to lines is not random, the assignment of training to the lines is random.

The standard practice in the factories is to train supervisors on the job by having them work for a period as assistant supervisors alongside an experienced supervisor. The assistant supervisors gradually take responsibility for an increasing number of sewing machine operators until they are managing a full section of the line. The group receiving both the soft- and hard-training sessions with six months delay thus mimicked the standard practice at factories.[13]

For logistical reasons, the 30 participating factories were divided into five sessions of roughly equal size, with the first beginning in November 2016 and the last in March 2017. Three factories dropped out of the project after we conducted the baseline survey, but before the promotion training began, leaving 27 factories that completed the program. We

[11] In Table 2.B.2 in the appendix, we show randomisation checks for the promotion training for a wide range of demographics, non-cognitive measures and diagnostic scores of the trainees. Out of 27 variables for which we test balance, we find one rejection using the p-value based on a wild percentile-t cluster bootstrap in the last column. This is fewer than one would expect to find by chance.

[12] At the end of the classroom training, all trainees, supervisors, line chiefs, and industrial engineering officers in charge of the respective trial lines participated in a team-building session at the factory. This was designed to increase collaboration among those working on the same line to ease the incorporation of the trainee in the management team. This was implemented in the same manner for all three treatment groups.

[13] To increase acceptance of the new female supervisors, on the first day each trainee began working as an assistant supervisor, the project trainers conducted an "onboarding training" that involved line chiefs and supervisors from the lines where the trainees were assigned to trial as assistant supervisors. Lower level managers were provided with a short training on the effectiveness of female line supervisors and the best way to support female line supervisors in succeeding in their new role. The onboarding exercise also included a session in which higher-level management introduced the trainee to the workers on the production line. The onboarding was implemented in the same manner for all three treatment groups.

use the 27 factories for the main analysis, but – where we have the data – show robustness to including the dropout factories in the appendix.

The initial training for the attitude and aptitude training and the attitude-only groups finished in May 2017. Attitude and aptitude training began for the control group, and aptitude training for the attitude-only group, in August 2017; all training was completed in October 2017. We conducted a follow-up survey in each factory just before the control group training started in each factory. The follow-up survey included samples of subordinate operators working on lines on which trainees were assigned to work as assistant supervisors. We use these data to conduct intent-to-treat regressions on the effectiveness of trainees as supervisors.

2.4 Results

2.4.1 Nominee rankings and baseline skills

We first evaluate the effect of the selection experiment on the rankings of the nominees and the initial skills of the selected trainees. These, and all other, outcome variables are defined in Table 2.B.3 in the appendix. As noted, in factories that were assigned to the selection experiment, we offered managers the chance to re-rank their nominees.[14] Managers in 11 of the 13 factories randomised into the treatment revised their lists after viewing the diagnostic scores of their nominees. We begin by testing whether the factories were implicitly taking into account the information contained in the diagnostic scores in their initial ranking of nominees. We do so by regressing the nomination rank of the candidates in all 27 factories before any potential re-ranking on the indices of nominees' initial attitude and aptitude. The indices are created as explained in Section 2.3. For clarity of exposition, we reversed the order of the rankings for the analysis such that a higher rank indicates a better position. Recall that the factories provided the nomination rankings before the nominated women even took the diagnostic tests, so none of the factories had seen the scores at this

[14] See Appendix Tables 2.B.5 and 2.B.6 for results of the re-ranking exercise for the sample including factories that dropped out. Those results are very similar to what we report here for the sample of the factories that continued through the full program.

Table 2.1: Nominee rankings and initial skills

	(1)	(2)
	Nomination rank	Movements in rank (after Selection experiment)
Aptitude (Index)	0.25	1.10**
	(0.41)	(0.05)
	[0.43]	[0.06]
Attitude (Index)	0.16	1.85***
	(0.66)	(0.00)
	[0.66]	[0.00]
Outcome mean	4.84	0.00
P-value (Soft=Hard skills)	0.87	0.30
Factory FE	Yes	Yes
Observations	243	138
Number factories	27	13
Sample	All factories	Selection experiment only

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in Column (1) and 8192 repetitions in Column (2) in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

point. The regression takes the following form for nominee i in factory f,

$$y_{if} = \alpha + \beta' \cdot \boldsymbol{S}_{if} + \theta_f + \epsilon_{if}, \tag{2.1}$$

where y is the nomination ranking as defined above[15] and $\boldsymbol{S}$ is a vector of the attitude and aptitude indices. We use within-factory variation by including factory fixed effects θ and cluster standard errors at the factory level.[16] Note that we are not arguing for a causal interpretation here, but test whether the initial skills predict the nomination rank.

The result of this regression is shown in the first column of Table 2.1. Both the aptitude and the attitude indices are positively related to a better nomination rank, with a one standard deviation higher index associated with a rank that is about a quarter and a sixth of a rank higher for aptitude and attitude, respectively. However, the coefficients are not significant and not significantly different from each other. In Table 2.B.4 in the appendix, we show the same regression but include all seven diagnostic scores (as fractions out of 1, where 1

[15]Because the number of nominees varies across factories, both the initial cardinal ranking and movements in the ranking may be disproportionately affected by a few factories with a large number of nominees. Regressions using percentile rankings, which reduces the influence of factories with more nominees (though arguably leads to excess influence of factories with a very small number of nominees), produce similar results.

[16]Standard errors from wild cluster percentile-t bootstraps are shown in squared brackets in this paper, to account for a potentially small number of clusters (following Cameron et al. (2008)).

represents a perfect score) separately as explanatory variables in the first column. We find positive but insignificant coefficients on all scores except numeracy, for which we find a negative and significant coefficient. Since this is likely due to the high positive correlations among the diagnostic scores, and especially among the two subgroups making up the attitude index and the aptitude index, we focus on the results for the indices.

This indicates that, at baseline, there was little correlation between the nominees' initial skills and the factories' preference for training them. We can think of several explanations for this. First, factories may have lacked good measures of these skills. Alternatively, they may not have thought these skills were important. Finally, this may reflect the highly selected nature of the pool of nominees to begin with, which could have made distinguishing between the candidates difficult. A typical factory selected only one woman for every two or three production lines, that is, around one in 40 female operators.

We now turn to ask whether the selection experiment changed how factories ranked their nominees for the promotion programme. This helps us to understand which initial skills presented in the results of the diagnostic tests, if any, the factories respond to. For this analysis, we only include the factories that were assigned to the selection experiment since only these factories were allowed to see the nominees' seven diagnostic scores and could then revise the ranking of their nominees. We capture how much nominees moved in the ranking by taking the difference between a nominee's ranking after the selection experiment and that nominee's initial ranking prior to the experiment. A positive difference indicates that a nominee moved to a better rank, whereas a negative difference indicates that a nominee moved to a worse rank. We again estimate equation 2.1, but with the nominees' movements in the rankings as defined above as the outcome variable. Thus, a positive coefficient implies that nominees with high initial attitude or aptitude were more likely to be moved up in the ranking. Note that the net movement as a result of re-ranking is zero – every move up is associated with an equivalent move down. Hence, a positive coefficient implies that managers placed a higher weight on the characteristic represented by that diagnostic. For this analysis, we include the 138 nominees in 13 factories that completed the selection experiment.

Column (2) in Table 2.1 shows the results. We find that nominees with both higher initial aptitude and attitude indices were moved to better positions in the rankings. On average, a one standard deviation higher index of aptitude or attitude is related to a movement up 1.10 ranks for aptitude and 1.85 ranks for attitude. We cannot reject that the coefficients for the aptitude and attitude indices are the same, however.[17] Note that the participating factories in the selection experiment only saw the seven diagnostic scores, not the attitude and aptitude indices aggregated from these scores. In Column (2) of Table 2.B.4 in the appendix, we include all seven diagnostic scores in the regression instead of the indices. Despite the correlation among the scores, we see that the numeracy, family support and confidence test scores are significantly associated with movements up the rankings.

The previous analysis suggests that factories re-ranked the nominees for the promotion programme based on their initial aptitude and attitude, once they were given the information. In the next step, we ask whether the selection experiment therefore leaves the participating factories with better trainees as measured by their initial skills. We address this in Table 2.2, where we report the results of a treatment effects regression of the form,

$$S_{if} = \alpha + \beta \cdot selection\_experiment_f + \gamma' \cdot \boldsymbol{X}_f + \epsilon_{if}, \tag{2.2}$$

where S is one of the indices of initial attitude and aptitude, $selection\_experiment$ indicates that factory f participated in the selection experiment, and $\boldsymbol{X}$ is a vector of variables used in the randomisation of the selection experiment. We again cluster standard errors at the factory level. The coefficient on the treatment variable tells us the difference between the average diagnostic scores of the selected trainees – those nominated at a sufficiently high enough level to be included in the training, after any re-ranking – in factories randomised to receive the selection experiment, compared with trainees in factories randomised out of the selection experiment.

Panel A of Table 2.2 reports the results for the diagnostics measuring aptitude – literacy,

[17] The regression does not control for the initial ranking of the nominee. When we add the initial ranking, we find that it is insignificant and does not change the magnitude or significance of the coefficients on either the indices or diagnostic measures.

Table 2.2: Trainees' initial skills

Panel A: Aptitude

	(1) Literacy	(2) Numeracy	(3) Processing speed	(4) Garment knowledge	(5) Aptitude (Index)
Selection experiment	0.06	0.04	0.01*	0.02*	0.29**
	(0.11)	(0.13)	(0.08)	(0.07)	(0.02)
	[0.16]	[0.22]	[0.10]	[0.16]	[0.05]
Control mean	0.55	0.45	0.32	0.53	0.10
Randomisation controls	Yes	Yes	Yes	Yes	Yes
Observations	199	199	199	199	199
Number factories	27	27	27	27	27

Panel B: Attitude

	(1) Family support	(2) Interest	(3) Confidence	(4) Attitude (Index)
Selection experiment	0.11***	0.15***	0.09***	0.62***
	(0.00)	(0.01)	(0.00)	(0.00)
	[0.01]	[0.04]	[0.00]	[0.00]
Control mean	0.68	0.68	0.68	−0.10
Randomisation controls	Yes	Yes	Yes	Yes
Observations	199	199	199	199
Number factories	27	27	27	27

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. The test of the cross-equation restriction that the coefficient for the Attitude and the Aptitude index is the same has a p-value of 0.02. * p<0.1, ** p<0.05, *** p<0.01

numeracy, fluid intelligence and technical knowledge related to garment production. All are defined as fractions out of 1, where 1 represents a perfect score. We see that factories participating in the selection experiment have trainees with marginally significantly higher fluid intelligence (as measured by processing speed) and technical knowledge, though the effect size is small and the coefficients are insignificant once we implement bootstrapping to account for a potential small number of clusters (p-values shown in square brackets). However, the effect on the aggregate aptitude index is larger at close to 0.3 SD and remains significant after bootstrapping. This indicates that the selected trainees in the factories that participated in the selection experiment have higher initial technical skills than the trainees in the non-participating factories.

Panel B of Table 2.2 shows the differences in average scores for the three diagnostics measuring attitudes as well as for the attitude index combining the three measures. The effects on the attitude diagnostics are larger than on any of the aptitude measures, and highly significant.[18] This is also true for the aggregate attitude index of the selected trainees, which is 0.6 SD higher in the factories who participated in the selection experiment than in the non-participating group. When testing the cross-equation restriction whether the coefficient is the same for the attitude and aptitude indices, we can reject equality with a p-value of 0.02.

Table 2.2 thus indicates that the factories completing the selection experiments selected trainees with both higher initial attitude and aptitude indices, though the effect is twice as large for attitude. The baseline balance shown in Appendix Table 2.B.1 indicates that the family support score is marginally imbalanced at baseline (p = 0.10), so some portion of the results on Table 2.2 may be driven by this imbalance. However, the combined results from Tables 2.1 and 2.2 do suggest that managers moved toward nominees with higher diagnostic scores following the selection experiment.

In sum, we find that factories value both the initial attitudes and aptitude of candidates for promotion, though they value attitude more. Factories who are given detailed information about the candidates' diagnostic scores choose trainees who have stronger initial attitude

[18] Fluid intelligence and technical knowledge have lower variances than the other five diagnostic scores, with standard deviations around 0.10 rather than 0.18-0.23 for the other five scores.

and aptitude.

2.4.2 Endline skills

The previous analysis showed that factories value the initial skills – and especially attitude measures – of their candidates for a promotion. We now investigate whether the initial skills that factories are selecting on are predictive of our outcomes of interest, and whether training in these skills can produce successful managers. We analyse this for the trainees' endline skills, promotions, as well as their performance according to evaluations of their subordinates and according to production data.

To measure these outcomes, we use a combination of survey and administrative data. Survey data collected just prior to the start of the delayed training sessions allows us to assess the short-term effect on endline skills and on official promotions. Regular telephone follow-up surveys during the initial training period allow us to measure the percentage of days the trainees worked as assistant supervisors rather than operators. Finally, surveys of operators working under the direction of the trainees and administrative data provide measures of trainee performance.

We start by investigating the trainees' endline skills, i.e. the skills that were enumerated at the end of the trainees' eight-week trial period as line supervisors. Recall that only the nominees that were ranked the highest by the factories – after a potential re-ranking in the factories that participated in the selection experiment and the removal of ineligible nominees based on the literacy and numeracy tests in all factories – were selected to receive training that aimed to prepare them for a promotion to supervisor. The selected operators were randomised into one of three groups, one receiving both the attitude and aptitude training immediately, one receiving only the attitude training immediately, and one receiving neither training immediately. All three groups were assigned to trial as assistant line supervisors for an eight-week trial period.

We use the randomisation to the different training regimes and the indices of the trainees' initial attitude and aptitude to investigate their impact on the skills trainees gained after

the trial period. We use a regression of the following form:

$$\begin{aligned} y_{ift=post} = \alpha &+ \beta \cdot attitudeonly\_training_{if} + \gamma \cdot aptitude\&attitude\_training_{if} \\ &+ \delta \cdot attitude_{ift=pre} + \mu \cdot aptitude_{ift=pre} \\ &+ \rho \cdot y_{ift=pre} + \omega' \boldsymbol{I}_{ift=pre} + \theta_f + \pi_e + \epsilon_{ift=post}, \end{aligned} \tag{2.3}$$

where y is a skill measure for nominee i in factory f at time t, which is post the trial period for $t = post$ and at baseline for $t = pre$. The variables $attitudeonly\_training$ and $aptitude\&attitude\_training$ indicate the random assignment to the different training groups, and $attitude$ as well as $aptitude$ are the indices measuring the initial conditions as defined above. We include controls for the one baseline imbalance we find (see Table 2.B.2 in the appendix), factory fixed effects θ, enumerator fixed effects π and cluster standard errors ϵ at the factory level.[19]

We are interested in the coefficients β and γ, which capture the effects of the trainings, as well as δ and μ, which show how the trainees' initial skills – that the factories selected on in the selection experiment – relate to the endline skills. Note that the randomisation of the training means that the variables capturing the treatment effects in expectation are orthogonal to the initial skills indices. The randomisation checks in Table 2.B.2 in the appendix confirm that this also holds in practice. The Analysis of Covariance (ANCOVA) specification in equation 2.3 also ensures that δ and μ capture the explanatory power of the initial attitude and aptitude indices once we have accounted for the outcomes' stability over time, i.e. for their gain in skills.[20]

As an alternative specification, we also show results of treatment effect regressions that include the assignment to the selection experiment as a dummy variable in equation 2.3, instead of the initial skill indices. We focus on the main specification in equation 2.3 since (1) this approach has greater statistical power compared to only relying on the across-factory

[19] Note that we do not have the statistical power to investigate the interactions between the training and initial skill levels.

[20] Table 2.B.7 in the appendix presents results from a non-ANCOVA specification, i.e. equation 2.3 estimated without including $y_{ift=pre}$. As can be expected, the coefficients for the initial aptitude and attitude indices are larger than in the ANCOVA specification.

Table 2.3: Effect on trainee skills

	(1) Garment knowledge	(2) Self-assessment	(3) Self-efficacy	(4) Internal LOC	(5) Stress
Attitude only training	−0.52	1.02**	0.12	0.35	0.88
	(0.77)	(0.03)	(0.10)	(0.20)	(0.47)
	[0.76]	[0.04]	[0.06]	[0.16]	[0.45]
Aptitude & Attitude training	1.50	0.52*	0.07	0.25	1.09
	(0.39)	(0.09)	(0.34)	(0.32)	(0.30)
	[0.34]	[0.07]	[0.31]	[0.29]	[0.28]
Aptitude (Index)	1.90	−0.05	−0.04	0.19**	−0.44
	(0.11)	(0.83)	(0.18)	(0.04)	(0.41)
	[0.11]	[0.82]	[0.28]	[0.04]	[0.36]
Attitude (Index)	−0.74	0.56***	0.14***	0.26	−0.17
	(0.32)	(0.01)	(0.00)	(0.10)	(0.77)
	[0.25]	[0.02]	[0.01]	[0.06]	[0.76]
Control mean	52.57	−1.68	3.27	4.18	11.30
P-value(Attitude=Aptitude&Attitude training)	0.19	0.22	0.45	0.75	0.86
Imbalance controls	Yes	Yes	Yes	Yes	Yes
Factory FE	Yes	Yes	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes	Yes	Yes
Observations	188	188	188	188	154
Number factories	27	27	27	27	25

Notes: ANCOVA specification. P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

variation of the assignment to the selection experiment, and (2) because this allows us to investigate the explanatory power of the initial attitude and aptitude indices separately.

Table 2.3 shows the results for five skills measures at endline. The first measure is the score in percentages on a test of knowledge about garment processes and production, measured after the trial period. This is the garments knowledge diagnostic that is described in Appendix 2.A, repeated after the trial period. The second column is a self-assessment of the trainee's expected performance as a supervisor. We asked trainees to rate both the typical supervisor and their own expected performance on a scale of one to ten. The self-assessment measure takes the difference between the own rating and the rating of the typical supervisor as reported by the trainee. A third measure is the self-efficacy score, which aggregates responses from the Generalized Self-Efficacy scale (Schwarzer and Jerusalem, 1995).[21] It captures an individual's belief in their own abilities to deal with new situations and to cope with any associated setbacks. The fourth is a measure of the trainees' internal locus of control, which we measure by asking respondents to choose one of two statements in each of seven different pairs of statements (based on Rotter (1966)). In each pair, the two

[21]Scholz et al. (2002) show that the questions produce reliable responses in 25 countries. Recent work by Laajaj et al. (2019) suggests that psychological measures may be measured with more noise in lower-income countries, and that there may also be concerns with response bias. The responses to the Generalized Self Efficacy scale have reasonable internal consistency, with Cronbach's alpha of 0.75 in our data.

statements either represent the view that an individual's life is controlled by their own actions (i.e. an internal locus of control) or that it is externally determined (i.e. an external locus of control). The final measure is the score on the Generalized Anxiety Disorder 7-item scale (Spitzer et al., 2006). This measures stress and anxiety over the preceding two weeks, where a higher score indicates higher stress.

Among these five measures, we find little effect on the trainees' garment knowledge at endline in column (1) of Table 2.3. While the coefficients of both the aptitude & attitude training and the aptitude index are positive, and the coefficients of the attitude only training and the attitude index are negative, they are not significant. In comparison, we do find some effects on the soft skills measures in columns (2) to (4). The supervisory trainings have a positive and significant effect on the trainee's self-assessment of their expected performance as a supervisor. Their effects on self-efficacy and the internal locus of control are also positive but insignificant. The initial skills indices, and especially the attitude measure, are significant predictors of the trainees' endline attitudes – over and above the baseline value of the outcome variable, which is also included in the regressions. For example, a one standard deviation increase in the initial attitude index is related to an increase in the trainees' self-assessment of about half a point, compared to a similar effect of half a point for the aptitude & attitude training and a whole point for the attitude only training. We do not find any significant relation of either the trainings or the initial skills measures on stress.

These results seem to suggest that both the initial skills that factories select on and the supervisory training matter for the new managers' gain in skills. Interestingly, there does not seem to be an additional value of the aptitude training, as we can in no case reject equality of the effects of attitude training only and of the training that focused on both attitude and aptitude.[22] That both training and initial skills matter is also confirmed if we estimate the alternative specification, by including a dummy for the selection experiment instead of the initial skills indices in equation 2.3.[23] We show the results in Table 2.B.8 in

[22]Of course, as in any training intervention, this could also mean that our skills training was not fit for purpose.

[23]Since the selection experiment was randomised on the factory level, we cannot include factory fixed ef-

the appendix. The variable for the selection experiment has a positive coefficient for all five outcomes, but the effect is only significant for self-efficacy. The coefficient's magnitude is in a similar ballpark as the supervisory training variables for the trainees' self-assessment, self-efficacy and internal locus of control.

2.4.3 Promotions

The next outcome of interest is whether the trainees complete the training programme and are promoted to supervisor. At one extreme, some operators dropped out of the program shortly after being selected as a participant; at the other extreme, some were officially promoted to a supervisory position by the time of the follow-up survey. Just over four-fifths (81 %) of the trainees reported working as an assistant supervisor at least once in weekly phone surveys. Table 2.4 reports the effects of the supervisory trainings and the coefficients of the initial skill indices on two outcome measures. The first is the percentage of days during the eight-week trial period that each trainee worked as an assistant supervisor. This data was collected in high-frequency phone surveys during the trial period.[24] The second outcome is whether trainees report having been officially promoted to a supervisory position. This was measured at two points in time, the follow-up survey at the end of the trial, and the household surveys described in more detail in Chapter 1. The household survey was conducted between January and March of 2018, six to ten months beyond the follow-up survey, and after all trainees had received the full training in attitude and aptitude.

We estimate equation 2.3 for these outcomes in Table 2.4, but do not include the baseline value in the regressions. In columns (1) and (2), we find that the percentage of workdays the trainee worked in a supervisory role is significantly higher among trainees receiving at least one of the formal supervisory training sessions when they began their trial as an assistant supervisor. This is particularly relevant given that the normal procedure in most

fects. Instead, we include the variables used in the randomisation as controls.

[24]Management often returned workers to machines during peak production periods, or when there was particular production pressure on a given day. Still, compliance in this regard was reasonably high: among those working at least one day as an assistant supervisor, 75 % worked more than half of the available days, and one-third worked every day in a supervisory capacity.

Table 2.4: Effect on promotions

	Share of trial days as SV		Promoted (at end of trial)		Promoted (after control training)	
	(1)	(2)	(3)	(4)	(5)	(6)
Attitude only training	0.22***	0.23***	0.03	0.03	0.17*	0.17*
	(0.00)	(0.00)	(0.48)	(0.42)	(0.06)	(0.05)
	[0.00]	[0.00]	[0.45]	[0.39]	[0.04]	[0.03]
Aptitude & Attitude training	0.18**	0.19**	0.10*	0.10*	0.06	0.07
	(0.03)	(0.02)	(0.08)	(0.07)	(0.47)	(0.41)
	[0.03]	[0.01]	[0.06]	[0.06]	[0.42]	[0.36]
Aptitude (Index)		−0.02		0.00		−0.02
		(0.57)		(0.83)		(0.74)
		[0.56]		[0.82]		[0.77]
Attitude (Index)		0.10***		0.02		0.10*
		(0.00)		(0.29)		(0.06)
		[0.01]		[0.27]		[0.09]
Control mean	0.51	0.51	0.00	0.00	0.19	0.19
P-value(Attitude=Aptitude&Attitude training)	0.44	0.41	0.15	0.15	0.25	0.28
P-value(Attitude=Aptitude)		0.00		0.61		0.20
Imbalance controls	Yes	Yes	Yes	Yes	Yes	Yes
Factory FE	Yes	Yes	Yes	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	175	175	188	188	178	178
Number factories	27	27	27	27	27	27

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

factories is to start the trial period without any formal training. The second column of Table 2.4 adds the initial attitude and aptitude indices. We find no evidence that trainees with higher technical scores worked more often as an assistant supervisor, but the attitude index is very strongly associated with working as a supervisor. Nevertheless, the magnitude of the training effect is larger, since the coefficients on the training variables are roughly equivalent to increasing the attitude index by two standard deviations.

The results in columns (3) to (6) in Table 2.4 indicate that the supervisory training increased the likelihood that the participating women were offered a promotion and were promoted. However, the difference to the control group is significant only for those in the aptitude & attitude training group when we measure promotions at the follow-up survey in column (3). Recall that at this point, the control group had not yet received any training. By the time the household survey was conducted, when we measured promotions for columns (5) and (6), trainees in all groups had completed training. The data show that around two-thirds of the trainees in each of the three randomisation groups were offered a promotion at that point. However, only those in the group that received early attitude only training

were more likely to have accepted the promotion and hence be working as a supervisor at the time of the household surveys, compared to the delayed group.[25]

The initial skill indices do not appear to be predictive of promotions at the end of the trial in column (4), though the attitude measure again is predictive for promotions measured in the longer term in column (6). This is confirmed in Table 2.B.9 in the appendix, when we regress the promotion outcomes of interest on the supervisory training variables and the selection experiment. The selection experiment in column (5) leads to a 14 percentage point increase in promotions in the longer term, though this effect is not significant after bootstrapping. This magnitude is nearly identical to the significant effect of the early attitude only training of 15 percentage points. Table 2.B.9 in the appendix also provides further evidence that the effect of the selection experiment indeed operates through factories selecting trainees with a higher initial attitude index, since the magnitude of the coefficient of the selection experiment reduces by two thirds once we add the initial attitude and aptitude indices to the regression in column (6).

Similar to the results on endline skills, overall we find that both the supervisory training and the initial skills are related to higher promotion rates, though the results are sometimes not significant. Again, the initial attitude measure seems to be more predictive than the aptitude measure, and we do not discern a difference between training focusing on attitudes only and teaching additional hard skills.

2.4.4 Performance as supervisor

Do the formal training or the initial skills affect the performance of trainees as supervisors?[26] We use two sources of data: production data – further discussed below – and surveys with 708 subordinates of the trainees, i.e. operators working on lines where the trainees were assigned to work during the trial period. These surveys were conducted at follow-up, when the early training was completed but the late training not yet started. To

[25]Though we cannot reject that the coefficients for the attitude only and the aptitude & attitude training are equal.

[26]We note at the start that the design of the project is better suited to answer the question of continued interest and promotion than performance, given that the experimentally generated variation lasts only a short time after the normal trial period ends. Only one fifth of the trainees had been offered a promotion to full supervisor at the time of the follow-up, though an additional 40 % were still working as assistant supervisors.

obtain a performance evaluation for each trainee as viewed by their subordinates, we ask the operators to rate, on a scale of one to ten, a typical supervisor in the factory, and then to rate the trainee on the same scale. We then subtract the typical supervisor rating from the rating for the trainee to obtain the subordinate rating for each trainee. This construction is analogous to the self-assessment measure in Table 2.3. In addition to this performance evaluation, we also investigate how the subordinates' wellbeing is affected. We use a well-being index, which combines the GAD-7 diagnostic discussed above (recoded such that higher numbers mean less stress), questions about verbally and physically abusive behaviour on the line, aspirations to be a supervisor and a question on general happiness over the previous two weeks. These questions are detailed in Table 2.B.3 in the appendix.

The regression specification we estimate is similar to equation 2.3, except that this time the unit of observation is subordinate s on production line l for the two outcomes y just discussed:

$$\begin{aligned} y_{slft=post} = \alpha &+ \beta \cdot attitudeonly\_training_{lf} + \gamma \cdot aptitude\&attitude\_training_{lf} \\ &+ \delta \cdot attitude_{lft=pre} + \mu \cdot aptitude_{lft=pre} \\ &+ \omega' \boldsymbol{I}_{lft=pre} + \theta_f + \pi_e + \epsilon_{slft=post}. \end{aligned} \tag{2.4}$$

The other variables are defined as above. Note that β and γ capture the effect of the trainee's assignment to a training group on the outcomes of the subordinates working on the production line to which the trainee was assigned to work as assistant supervisor during the trial period. Similarly, δ and μ capture how the trainee's initial attitude and aptitude measures relate to the outcomes from the subordinate surveys.

We report on these intent-to-treat regressions in Table 2.5.[27] In the first two columns of Table 2.5, we show results for the ratings the subordinates gave the new supervisors. On

[27] There are two types of non-compliance relevant for interpreting the results. First, as we have noted, almost one fifth of the trainees never work as assistant supervisors. For the survey questions, we therefore begin by asking operators if they recall the trainee working as a supervisor. Those responding "not at all" or reporting that the trainee never worked in a supervisory capacity on their line, we ask for a generic comparison between a typical supervisor and a "typical female supervisor." Just over one third of operators (38 %) answer the generic question. Second, some of the trainees work on a line other than the line where they were assigned. Of course, these movements may be endogenous, so we present the intent-to-treat regressions.

Table 2.5: Effect on subordinate outcomes

	Subordinate ratings		Subordinate wellbeing	
	(1)	(2)	(3)	(4)
Attitude only training	0.18	0.22	0.08	0.08
	(0.42)	(0.35)	(0.34)	(0.38)
	[0.44]	[0.37]	[0.34]	[0.39]
Aptitude & Attitude training	0.07	0.06	0.17*	0.17*
	(0.73)	(0.77)	(0.07)	(0.06)
	[0.73]	[0.77]	[0.06]	[0.06]
Aptitude (Index)		0.14		−0.04
		(0.14)		(0.34)
		[0.26]		[0.36]
Attitude (Index)		0.07		0.09**
		(0.40)		(0.03)
		[0.43]		[0.02]
Control mean	−0.50	−0.50	−0.01	−0.01
P-value(Attitude=Aptitude&Attitude training)	0.46	0.34	0.51	0.45
P-value (Attitude=Aptitude)		0.61		0.04
Imbalance controls	Yes	Yes	Yes	Yes
Factory FE	Yes	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes	Yes
Observations	707	707	708	708
Number factories	26	26	26	26

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

average, trainees are rated about a half point (one-third of a standard deviation) lower than typical supervisors. We see that both the variables capturing the training effects as well as the initial skill indices are positively related to the ratings, but they are far from statistical significance. When we investigate the subordinate wellbeing index in columns (3) and (4), we find positive coefficients of the supervisory training, which is significant only for the aptitude & attitude training group. However, we cannot reject the equality of coefficients for the training groups. When we add the two initial skill indices in column (4), we find that the initial attitude score is significantly associated with higher subordinate well-being, and find an insignificant negative coefficient of the aptitude index. We can also reject equality between the effects of the initial attitude and aptitude indices.

These patterns are confirmed in Table 2.B.10 in the appendix, where we include the selection experiment indicator instead of the initial skills indices. The coefficients on this variable are positive and insignificant for the subordinate ratings in column (1) and the wellbeing index in column (3). Similar to the results on promotions, we find that the treat-

ment effect of the selection experiment becomes smaller once we control for the initial skills of the trainees in the even columns of Table 2.B.10. There, we see that the initial aptitude and attitude indices are marginally significant predictors of the subordinate ratings, and that only the initial attitude index is a highly significant predictor for subordinate well-being. These results from the subordinate surveys suggest that both the initial skills and training matter, to an extent.

In addition to the survey data, we use administrative records from the factories measuring daily productive efficiency, quality defect rates (in percent) and absenteeism (in percent) at the production-line level. These measures are also described in 2.B.3. We report how the trainings and the initial skills indices relate to these performance measures in Table 2.6.[28] We estimate an intent-to-treat ANCOVA regression for production line l for the outcomes y in the eight-week trial period as follows:

$$\begin{aligned} y_{lfdt=post} = \alpha &+ \beta \cdot attitudeonly\_training_{lf} + \gamma \cdot aptitude\&attitude\_training_{lf} \\ &+ \delta \cdot attitude_{lft=pre} + \mu \cdot aptitude_{lft=pre} \\ &+ \rho \cdot \bar{y}_{lft=pre} + \omega' \boldsymbol{I}_{lft=pre} + \theta_f + \tau_d + \epsilon_{lfdt=post}. \end{aligned} \tag{2.5}$$

By including factory-fixed effects θ and date fixed effects τ, we only rely on variation within each factory and day. By including the mean of the outcome for each line in the pre-trial period, $\bar{y}$, we ensure that a potential assignment of the trainees to more or less productive lines is not driving results.[29]

As Table 2.6 shows, the data give little indication that the training or the initial skills predict the production outcomes. The coefficients for both the supervisory trainings and the initial skill indices are always insignificant and small compared to the control mean, though the

[28] Note that, in addition to the two sources of non-compliance discussed above with regard to the operator opinions, there is a third measurement issue that is relevant in interpreting these results. Our productivity measures are made at the production line level. More complex products are produced on lines that typically have two or even three line supervisors, each responsible for only a part of the line. The line-level measures will therefore reflect the combined effort of more than one supervisor. Moreover, even on lines with only a single supervisor, our trainees were almost always working as an assistant supervisor, and hence responsible for only a part of the line.

[29] Note that we only keep lines that were working on specific days, and also drop the few lines where several trainees were assigned to trial with different trial start dates or different training assignments.

Table 2.6: Effect on production outcomes

	Efficiency		Alteration rates (%)		Absenteeism (%)	
	(1)	(2)	(3)	(4)	(5)	(6)
Attitude only training	0.20	0.23	0.15	0.16	0.30	0.30
	(0.89)	(0.88)	(0.64)	(0.63)	(0.17)	(0.19)
	[0.89]	[0.88]	[0.65]	[0.64]	[0.14]	[0.17]
Aptitude & Attitude training	0.37	0.37	−0.29	−0.29	0.29	0.27
	(0.73)	(0.73)	(0.37)	(0.36)	(0.28)	(0.32)
	[0.73]	[0.73]	[0.38]	[0.39]	[0.27]	[0.31]
Aptitude (Index)		0.15		0.11		0.15
		(0.82)		(0.59)		(0.12)
		[0.82]		[0.61]		[0.33]
Attitude (Index)		0.07		−0.16		−0.03
		(0.90)		(0.28)		(0.88)
		[0.90]		[0.23]		[0.93]
Control mean	52.21	52.21	6.85	6.85	4.65	4.65
P-value(Attitude=Aptitude&Attitude training)	0.89	0.91	0.25	0.23	0.96	0.89
P-value (Attitude=Aptitude)		0.93		0.38		0.50
Imbalance controls	Yes	Yes	Yes	Yes	Yes	Yes
Factory FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treatment line mean	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5769	5769	5911	5911	5680	5680
Number factories	17	17	23	23	17	17

Notes: Intent-to-treat ANCOVA comparison within the 8 weeks trial period. P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

initial attitude index does have signs in the direction of improvement for all three measures.[30] The lack of an effect may not be surprising given that the trainees on most lines represented one of two or three supervisors on each line managing 20-80 line operators. Also, only a small percentage of the trainees are working as full supervisors at the time of the follow-up, and even those would have accumulated little experience at the time.[31]

[30]When we include the selection experiment variable instead of the initial skill indices in the regression in Table 2.B.11 in the appendix, we find that it is associated with significantly worse outcomes for quality defect rates. However, since we do not find these associations for the initial skill indices in Table 2.6, we posit that this is primarily due to not being able to control for factory fixed effects in Table 2.B.11 in the appendix.

[31]When we compare the production outcomes of lines to which trainees were assigned with outcomes of lines with no assigned trainee during the factory's trial period (again controlling for factory fixed effects, date fixed effects, and baseline means of outcomes), we find that lines with trainees have significantly lower absenteeism, insignificantly higher efficiency and insignificantly lower alteration rates. These results seem to suggest that the trainees – compared to typical lines in the factory – positively contribute to their lines' productivity, and support the conclusions we drew from the subordinate survey outcomes. They are also in line with evidence from a similar context, the garment sector in India, for which Adhvaryu et al. (2019) show that managerial skills are important determinants of line productivity.

2.5 Conclusion

After four decades of rapid growth, the garment sector in Bangladesh now represents around one eighth of GDP and more than 80 % of exports. Importantly, the sector remains the primary source of employment for women working full time outside agriculture. The 2017 Bangladesh Labor Force Survey indicates that among women without tertiary education and working full time, 40 % of those employed outside agriculture are employed in the garment sector. However, women's role in the sector is limited almost exclusively to production-worker positions; fewer than one in ten factory managers are women. Facing new pressures from several sources, factories are increasingly interested in promoting women to supervisory positions. The shift toward increased interest in promoting women represents an important cultural shift in the factories.

The results in this paper should be viewed in the context of this recent interest in promoting women. We document four results. First, factories do not take into account the initial aptitude and attitudes of candidates when they nominate women for a promotion training. Second, the factories that are provided information about their candidates' skills and are able to change their selection as part of an experiment react to this information. They select candidates who have higher baseline skills, especially on the attitude and soft skills dimension. Third, we find that the initial skills – and especially attitudes and soft skills – matter for the outcomes of the promotion programme. They are related to higher endline skills, to the promotion rates to official line supervisor, to higher wellbeing of their subordinates and – to some degree – to higher performance evaluations as judged by subordinates. Fourth, training in these attitudes and aptitude supervisory skills leads to improvements in some of these same outcomes, though the results are not always significant. In none of our results do we find that training in technical skills has an additional effect when compared to a training only targeting attitudes and soft skills.

Returning to our initial question, we therefore find that both factories learning how to choose supervisors and women learning how to lead matters. For the new female low-level managers that are the focus of this paper, attitudes and soft skills appear to be partic-

ularly important. The initial attitudes and soft skills are predictive of the trainees' success, and training in these attitudes and soft skills also enables trainees to be more successful. The results of the selection experiment suggest that factories had some idea that attitudes and soft skills would matter for their trainees' success, but the little experience in choosing female supervisors likely meant that they lacked the ability to convincingly measure these skills. This is also important from the perspective of cost-effectiveness, since we find effects of similar magnitude from a short, cheap intervention with higher-level management that aims to base manager selection on formal skills tests as compared to a longer-term, expensive training programme for new female line supervisors. In the case of female career advancement, this paper therefore not only indicates that training in non-cognitive skills is a promising avenue to increase the participation of women in managerial positions, but also provides evidence for the importance of hard measures of soft skills and attitudes.

Acknowledgements The paper has benefited greatly from discussions with and comments by Abi Adams, Elwyn Davies, Kate Orkin, and from audiences at the EDI Annual Conference 2020, the Monash Conference on RCTs 2019, and OxFLO 2017-18. The authors thank Gunjita Gupta and Ferran Vega for research assistance, and the International Finance Corporation, the Urban Services Initiative (USI) at J-PAL, and the Economic Development and Institutions research programme (EDI), funded by UK aid from the UK government, for financial support. Woodruff's time was supported by ERC Advanced Grant RMGPP. The surveys and interventions were supported by the Bangladesh office of Innovations for Poverty Action (IPA). The project would not have been possible without the support and cooperation of participating factories and workers, which is greatly appreciated. This study, together with Chapter 1, was registered with the AER RCT Registry as AEARCTR-0001843, and received ethical approval from the University of Oxford and Innovations for Poverty Action.

Appendices

2.A Description of diagnostic tests

After the factory nominated workers, the IPA team visited the factory to assess the abilities of the nominees. Nominees were only disqualified based on numeracy and literacy scores, see below. The areas that were tested and their respective tests are detailed below. For each of the seven diagnostics, the score is calculated as a fraction out of 1, where 1 is a perfect score.

1. **Literacy**: Multiple choice questions testing reading comprehension, vocabulary, basic grammar, paragraph/letter structure, and writing. The maximum is 20 points.
2. **Numeracy**: Multiple choice questions testing calculations, fractions, percentages, number patterns, angles, and visual patterns. The maximum is 20 points.
3. **Processing speed**: This score consisted of two tests of fluid intelligence, detailed below.
    (a) Coding: This test is modelled after the Wechsler Adult Intelligence Scale is called 'Digit Symbol' (WAIS-R), 'Digit-Symbol-Coding' (WAIS-III), 'Coding' (WAIS-IV), also known as the Digit symbol substitution test. It is a neuropsychological test sensitive to brain damage, dementia, age and depression. It consists of digit-symbol pairs followed by a list of digits. Under each digit the subject should write down the corresponding symbol as fast as possible. The number of correct symbols within the allowed time is measured. The maximum is 133 points.
    (b) Symbol search: Symbol Search is a subtest of the Wechsler Adult Intelligence Scale (WAIS). The Symbol Search subtest is designed to assess information pro-

cessing speed and visual perception. High scores require rapid and accurate processing of nonverbal visual information. During Symbol Search, the examinee is asked to mark either the yes or no checkbox with a pencil in response to as many items as possible within 2 min. The maximum is 60 points.

4. **Garments Knowledge**: Contains multiple choice, and open-ended questions testing what machine to use for what operation, names of processes, cause of mechanical issues, cause of quality issues, identifying quality issues in photographs, identifying working condition issues in photographs, and understanding an operation breakdown. The maximum is 84 points.

5. **Family Support**: Gives 5 statements about family support, and asks the respondent to respond on a four-point scale from agree to disagree. An additional three questions ask about the level of support given to other women in the family who work in garment factories. The answers to all questions are recoded so that higher numbers represent more support, and are summed to give a maximum of 24 points.

6. **Interest**: The survey instrument includes 2 questions about whether the nominee would want to be promoted to supervisor or line chief. In addition, four questions indirectly probe whether nominees are interested in the supervisor position, and asks the respondent to respond on a four-point scale from agree to disagree. The answers to all questions are recoded so that higher numbers represent more interest, and are summed to give a maximum of 18 points.

7. **Confidence**: Consists of asking how they would rate their performance compared to a typical supervisor on a 5-point scale. In addition, three questions indirectly ask whether the respondent is confident. The respondent is asked to choose between two statements. One statement that says "I am confident" using various words, and a dummy statement about the factory. The answers are recoded so that higher numbers represent more confidence and are summed to give a maximum of 7 points.

Disqualification rule: Nominees were only disqualified based on the scores in the numeracy and literacy tests. The rule is:

Table 2.A.1: Overview of diagnostic test scores

Testing area	Maximum points
Literacy	20
Numeracy	20
Processing speed	193
Garment knowledge	84
Family support	24
Interest	18
Confidence	7

- If the nominee scores below 25 % on both literacy and numeracy tests, she fails.
- If the nominee scores 0 % in either literacy or numeracy tests, she fails irrespective of her score in the other test.

2.B Additional tables

Table 2.B.1: Factory balance, Selection experiment

	Control		Selection		Mean equality	Mean equality
	Obs	(Mean)	Obs	(Mean)	p-value	p-value (RI)
Participated in prev project	14	0.36	13	0.46	0.60	0.27
Produces knit garments	14	0.50	13	0.46	0.85	0.84
Number of workers	14	2332.21	13	2860.77	0.46	0.34
Share of female workers	14	0.57	13	0.59	0.69	0.63
Date of joining programme	14	20346.43	13	20386.15	0.72	0.62
Number of supervisors	14	48.14	13	52.69	0.78	0.77
Share of female supervisors	14	0.04	13	0.06	0.44	0.43
Number of lines	14	24.43	13	30.46	0.49	0.44
(A)PM's age (avg)	14	37.94	13	37.14	0.59	0.59
(A)PM's education years (avg)	14	10.40	13	10.45	0.92	0.90
(A)PM's spouse works (avg)	14	0.08	13	0.06	0.81	0.81
(A)PM exposure (avg)	14	0.98	13	0.97	0.81	0.76
(A)PM nr. male SV are better (avg)	14	4.68	13	5.03	0.55	0.50
Actual to calculated trainees	14	3.94	11	4.12	0.89	0.90
SV/Line Chief IAT (avg)	14	−0.25	13	−0.29	0.58	0.59
Literacy score	14	0.49	13	0.54	0.22	0.22
Numeracy score	14	0.42	13	0.43	0.80	0.80
Processing speed score	14	0.31	13	0.33	0.19	0.21
Garments knowledge score	14	0.54	13	0.55	0.36	0.38
Family support score	14	0.69	13	0.74	0.15	0.10*
Interest score	14	0.70	13	0.74	0.39	0.33
Confidence score	14	0.69	13	0.74	0.21	0.20

Notes: p-value from a regression with robust standard errors in penultimate column, and from randomisation inference with 10,000 permutations in last column. Scores are the average by factory for all operators nominated by that factory. Missing values in Actual to calculated trainees are due to calculated number of trainees being 0 for 2 factories. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.2: Balance, Promotion training

	Control (N=65) (Mean)	**Attitude only (N=67)** (Mean)	**Aptitude & Attitude (N=67)** (Mean)	**Joint test p-value**	**p-value** (bootstrapped)
Age	25.26	25.88	25.79	0.28	0.30
Married	0.77	0.76	0.72	0.79	0.81
Household members	3.18	3.51	3.12	0.22	0.24
Household head	0.22	0.25	0.22	0.89	0.90
Migrant	0.63	0.69	0.60	0.55	0.55
Education years	8.38	8.36	8.42	0.98	0.98
Experience in garment sector	5.60	5.84	6.36	0.12	0.15
Tenure	3.14	3.40	3.84	0.12	0.13
Nr. of factories worked in	1.28	1.34	1.31	0.96	0.96
Exposure to female SV	0.68	0.46	0.51	0.04**	0.06*
Nr. male SV are better	3.34	3.37	3.57	0.78	0.79
Internal locus of control	4.43	4.57	4.45	0.82	0.83
Grit	2.90	2.98	2.92	0.66	0.67
Self-efficacy	3.58	3.66	3.60	0.46	0.50
Emotional competence	3.05	3.15	3.11	0.17	0.19
Multi-factor Leadership	3.67	3.72	3.78	0.19	0.21
Life satisfaction	7.72	7.33	7.52	0.45	0.46
Numeracy	1.12	0.87	1.19	0.12	0.15
Self-assessment	0.12	0.25	−0.15	0.31	0.35
Ambition	2.35	2.18	2.12	0.30	0.31
Numeracy score	0.49	0.47	0.46	0.57	0.59
Literacy score	0.59	0.55	0.60	0.44	0.48
Processing speed score	0.34	0.32	0.33	0.37	0.40
Garment knowledge score	0.55	0.53	0.56	0.21	0.23
Family support score	0.74	0.72	0.74	0.91	0.91
Interest score	0.74	0.77	0.73	0.47	0.50
Confidence score	0.75	0.71	0.71	0.24	0.26

Notes: p-value of the joint test Attitude only = Aptitude & Attitude = 0 from a regression with standard errors clustered on factory level in penultimate column, and p-value of the same test from wild percentile-t cluster bootstrap with 10,000 repetition in the last column. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.3: Outcome variables

Variable	Definition
Nomination rank	Rank of nominated woman in first ranking provided by factories, after removal of ineligible candidates but before a potential re-ranking. A higher number indicates a better rank.
Movements in rank	Number of ranks that a nominee moved between the first ranking and the final ranking after a potential re-ranking. Calculated as (nominee's final rank after a potential re-ranking - nominee's nomination rank). A positive number indicates a movement to a better rank. Only defined for nominees in factories that participated in the selection experiment.
Literacy score	Score out of 1 on the diagnostic tests detailed in Appendix 2.A, calculated as (points achieved divided by maximum points possible).
Numeracy score	
Processing speed score	
Garment knowledge score	
Family support score	
Interest score	
Confidence score	
Aptitude (Index)	Standardised index of the Literacy, Numeracy, Processing speed, and Garment knowledge scores, created following Anderson (2008).
Attitude (Index)	Standardised index of the Family support, Interest and Confidence scores, created following Anderson (2008).
Self-assessment	The variable is created from two survey questions. Respondents were first asked to rate the overall supervisor ability of a typical supervisor in their factory on a scale from 1 to 10, where a higher rating is better. Second, they were then asked how they think they perform or would perform as supervisor on the same scale. Self-assessment is calculated as (own rating - typical supervisor rating). A higher rating indicates a better assessment compared to a typical supervisor.
Self-efficacy	Self-efficacy score of 10 items following Schwarzer and Jerusalem (1995). After re-coding reverse-coding items, the score is calculated as the mean of all 10 items. A higher scores indicates higher self-efficacy.
Internal LOC	Internal locus of control score of 7 items that are a subset of Rotter (1966). The scores is calculated as the number of items on which the respondent chooses the internal option.
Stress	Stress is measured by the Generalized Anxiety Disorder 7-item scale (Spitzer et al., 2006), which respondents completed on paper with emojis representing response options from 1 to 4. The stress score is calculated as the total of all 7 responses. A higher number represents higher reported stress.

(Table 2.B.3 continued.)

Share of trial days as supervisor	Variable is derived from weekly phone surveys during the trial period. For each weekly phone survey, we calculate the share of work days that the respondent worked as supervisor or assistant supervisor (taking into account holidays and training days). For each respondent, the share of trial days as supervisor is calculated as the average across the completed phone surveys of the share of work days that the trainee worked as supervisor or assistant supervisor.
Promoted	Dummy variable indicating whether respondent accepted a promotion to line supervisor since the on-boarding session (if measured at end of trial) or since December 2016 (if measured after control training).
Subordinate wellbeing	Standardised index created following Anderson (2008), capturing respondent wellbeing and composed of • Stress: Created as defined above, but the score was reversed such that a higher number captures less stress. • Aspiration: Dummy variable indicating whether respondents would someday accept an offer for a promotion to line supervisor. • No verbal abuse: Dummy variable indicating that, in a phone survey, respondents replied "Not at all" to the question "Please think about the other operators on your line. Over the last two weeks, how often have they needed to put up with shouting or abusive language at work?" • No physical abuse: Dummy variable indicating that, in a phone survey, respondents replied "Not at all" to the question "Some people experience situations at work that make them feel uncomfortable. I am going to read some examples: People staring persistently or winking, coming very close to them or calling them to get close, flirting, singing, making sounds or whistling, making gestures, bumping into or rubbing against them, grabbing their hand or other parts of their bodies, tickling, etc. Please think about the other operators on your line and answer this question. Over the last two weeks, how often did at least one of these things happen to them at work?" • Very happy: Dummy variable indicating that, in a phone survey, respondents reported that they were very happy over the last two weeks.

(Table 2.B.3 continued.)

Subordinate ratings	Defined in analogy to the self-assessment variable. The variable is created from two survey questions. Respondents were first asked to rate the overall supervisor ability of a typical supervisor in their factory on a scale from 1 to 10. Second, they were then asked how they think the trainee assigned to trial on their production line performs or would perform as supervisor on the same scale. Subordinate ratings is calculated as (trainee rating - typical supervisor rating). If the respondents did not remember the trainee well or very well, they were asked to rate a female line supervisor in their factory, and this rating then replaced the trainee rating.
Efficiency	Measures how the daily output of each line in garment units, adjusted for the standard minute value of the garment, compares to the total output possible that day with the workers present on the line and the hours worked. Calculated as $$efficiency = \frac{(Pieces\,produced \cdot SMV)}{Workers\,on\,the\,line \cdot Hours\,operated \cdot 60} \cdot 100. \quad (2.6)$$
Alteration rate	Percentage of all garments that needs to be altered out of all produced units
Absenteeism rate	Percentage of workers absent amongst all workers (present and absent) on the line

Table 2.B.4: Nominee rankings and initial diagnostic scores

	(1)	(2)
	Nomination rank	Movements in rank (after Selection experiment)
Literacy	1.08	1.94
	(0.45)	(0.50)
	[0.45]	[0.49]
Numeracy	−3.16 **	5.17**
	(0.03)	(0.03)
	[0.04]	[0.07]
Processing speed	2.43	−5.26
	(0.53)	(0.22)
	[0.54]	[0.27]
Garment knowledge	4.83	4.76
	(0.12)	(0.33)
	[0.13]	[0.32]
Family support	1.19	3.56*
	(0.52)	(0.08)
	[0.50]	[0.19]
Interest	0.09	0.64
	(0.97)	(0.68)
	[0.97]	[0.67]
Confidence	0.05	5.70**
	(0.96)	(0.01)
	[0.96]	[0.02]
Outcome mean	4.84	0.00
Factory FE	Yes	Yes
Observations	243	138
Number factories	27	13
Sample	All factories	Selection experiment only

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in Column (1) and 8192 repetitions in Column (2) in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.5: Nominee rankings, and initial skills and scores – including dropout factories

	Nomination rank		Movements in rank (after Selection experiment)	
	(1)	(2)	(3)	(4)
Aptitude (Index)	0.25		1.03*	
	(0.37)		(0.05)	
	[0.40]		[0.06]	
Attitude (Index)	0.16		1.82***	
	(0.62)		(0.00)	
	[0.63]		[0.00]	
Literacy		0.92		1.72
		(0.50)		(0.54)
		[0.50]		[0.53]
Numeracy		−2.91 **		5.21**
		(0.03)		(0.03)
		[0.04]		[0.06]
Processing speed		2.62		−5.39
		(0.48)		(0.21)
		[0.48]		[0.25]
Garment knowledge		4.54		4.60
		(0.13)		(0.34)
		[0.15]		[0.32]
Family support		1.00		3.57*
		(0.58)		(0.07)
		[0.56]		[0.16]
Interest		0.36		0.45
		(0.87)		(0.75)
		[0.86]		[0.75]
Confidence		0.02		5.67**
		(0.99)		(0.01)
		[0.99]		[0.02]
Outcome mean	4.76	4.76	0.01	0.01
Factory FE	Yes	Yes	Yes	Yes
Observations	257	257	143	143
Number factories	28	28	13	13
Sample	All factories	All factories	Selection experiment only	Selection experiment only
Includes dropout factories	Yes	Yes	Yes	Yes

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in Columns (1)-(2) and 8192 repetitions in Columns (3)-(4) in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.6: Trainees' initial skills – including dropout factories

Panel A: Aptitude

	(1) Literacy	(2) Numeracy	(3) Processing speed	(4) Garment knowledge	(5) Aptitude (Index)
Selection experiment	0.06	0.03	0.01	0.02**	0.28**
	(0.11)	(0.19)	(0.19)	(0.04)	(0.02)
	[0.18]	[0.30]	[0.23]	[0.11]	[0.06]
Control mean	0.54	0.45	0.32	0.54	0.13
Randomisation controls	Yes	Yes	Yes	Yes	Yes
Observations	233	233	233	233	233
Number factories	29	29	29	29	29

Panel B: Attitude

	(1) Family support	(2) Interest	(3) Confidence	(4) Attitude (Index)
Selection experiment	0.11***	0.14**	0.08***	0.60***
	(0.00)	(0.01)	(0.00)	(0.00)
	[0.01]	[0.07]	[0.00]	[0.00]
Control mean	0.67	0.67	0.68	−0.11
Randomisation controls	Yes	Yes	Yes	Yes
Observations	233	233	233	233
Number factories	29	29	29	29

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. The test of the cross-equation restriction that the coefficient for the Attitude and the Aptitude index is the same has a p-value of 0.01. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.7: Effect on trainee skills – no ANCOVA

	(1) Garment knowledge	(2) Self-assessment	(3) Self-efficacy	(4) Internal LOC	(5) Stress
Attitude only training	−0.45	1.03**	0.16*	0.35	1.22
	(0.80)	(0.02)	(0.06)	(0.22)	(0.30)
	[0.79]	[0.03]	[0.04]	[0.17]	[0.26]
Aptitude & Attitude training	1.58	0.50	0.09	0.24	1.27
	(0.35)	(0.10)	(0.22)	(0.34)	(0.21)
	[0.30]	[0.08]	[0.19]	[0.30]	[0.18]
Aptitude (Index)	2.33**	−0.07	−0.02	0.19*	−0.55
	(0.02)	(0.77)	(0.40)	(0.07)	(0.30)
	[0.00]	[0.76]	[0.45]	[0.09]	[0.24]
Attitude (Index)	−0.72	0.60***	0.20***	0.30**	−0.20
	(0.34)	(0.00)	(0.00)	(0.04)	(0.71)
	[0.29]	[0.01]	[0.00]	[0.02]	[0.70]
Control mean	52.57	−1.68	3.27	4.18	11.30
P-value(Attitude=Aptitude&Attitude training)	0.19	0.20	0.35	0.73	0.96
Imbalance controls	Yes	Yes	Yes	Yes	Yes
Factory FE	Yes	Yes	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes	Yes	Yes
Observations	188	188	188	188	154
Number factories	27	27	27	27	25

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.8: Effect on trainee skills – with Selection experiment

	(1) Garment knowledge	(2) Self-assessment	(3) Self-efficacy	(4) Internal LOC	(5) Stress
Attitude only training	−1.16	0.88**	0.08	0.33	0.93
	(0.46)	(0.03)	(0.28)	(0.22)	(0.37)
	[0.47]	[0.04]	[0.26]	[0.20]	[0.36]
Aptitude & Attitude training	0.55	0.49*	0.05	0.22	1.07
	(0.76)	(0.09)	(0.49)	(0.37)	(0.27)
	[0.75]	[0.08]	[0.49]	[0.37]	[0.28]
Selection experiment	0.24	0.40	0.14**	0.19	0.25
	(0.91)	(0.38)	(0.03)	(0.34)	(0.68)
	[0.92]	[0.55]	[0.06]	[0.43]	[0.71]
Control mean	52.57	−1.68	3.27	4.18	11.30
P-value(Attitude=Aptitude&Attitude training)	0.29	0.27	0.69	0.68	0.90
Imbalance controls	Yes	Yes	Yes	Yes	Yes
Randomisation controls	Yes	Yes	Yes	Yes	Yes
Factory FE	No	No	No	No	No
Enumerator FE	Yes	Yes	Yes	Yes	Yes
Observations	188	188	188	188	154
Number factories	27	27	27	27	25

Notes: ANCOVA specification. P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.9: Effect on promotions – with Selection experiment

	Share of trial days as SV		Promoted (at end of trial)		Promoted (after control training)	
	(1)	(2)	(3)	(4)	(5)	(6)
Attitude only training	0.26***	0.28***	0.04	0.05	0.15**	0.16**
	(0.00)	(0.00)	(0.24)	(0.18)	(0.04)	(0.02)
	[0.00]	[0.00]	[0.24]	[0.17]	[0.03]	[0.01]
Aptitude & Attitude training	0.18**	0.19***	0.10*	0.10*	0.04	0.04
	(0.02)	(0.01)	(0.07)	(0.06)	(0.61)	(0.55)
	[0.01]	[0.01]	[0.07]	[0.06]	[0.58]	[0.51]
Selection experiment	0.01	−0.07	−0.05	−0.08 *	0.14	0.04
	(0.91)	(0.37)	(0.26)	(0.08)	(0.22)	(0.68)
	[0.92]	[0.43]	[0.29]	[0.12]	[0.35]	[0.73]
Aptitude (Index)		0.00		0.01		0.03
		(0.95)		(0.52)		(0.67)
		[0.95]		[0.54]		[0.69]
Attitude (Index)		0.14***		0.04*		0.14***
		(0.00)		(0.05)		(0.00)
		[0.00]		[0.05]		[0.03]
Control mean	0.51	0.51	0.00	0.00	0.19	0.19
P-value(Attitude=Aptitude&Attitude training)	0.14	0.08	0.25	0.30	0.15	0.13
P-value(Attitude=Aptitude)		0.00		0.42		0.21
Imbalance controls	Yes	Yes	Yes	Yes	Yes	Yes
Randomisation controls	Yes	Yes	Yes	Yes	Yes	Yes
Factory FE	No	No	No	No	No	No
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	175	175	188	188	178	178
Number factories	27	27	27	27	27	27

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.10: Effect on subordinate outcomes – with Selection experiment

	Subordinate ratings		Subordinate wellbeing	
	(1)	(2)	(3)	(4)
Attitude only training	0.15	0.19	0.07	0.06
	(0.48)	(0.39)	(0.43)	(0.50)
	[0.49]	[0.41]	[0.44]	[0.51]
Aptitude & Attitude training	0.05	0.04	0.15	0.16*
	(0.77)	(0.82)	(0.11)	(0.08)
	[0.78]	[0.83]	[0.10]	[0.08]
Selection experiment	0.08	−0.05	0.01	−0.03
	(0.66)	(0.77)	(0.83)	(0.66)
	[0.76]	[0.82]	[0.84]	[0.67]
Aptitude (Index)		0.15*		−0.06
		(0.10)		(0.10)
		[0.20]		[0.13]
Attitude (Index)		0.12*		0.10***
		(0.07)		(0.01)
		[0.11]		[0.01]
Control mean	−0.50	−0.50	−0.01	−0.01
P-value(Attitude=Aptitude&Attitude training)	0.51	0.35	0.50	0.39
P-value (Attitude=Aptitude)		0.77		0.01
Imbalance controls	Yes	Yes	Yes	Yes
Randomisation controls	Yes	Yes	Yes	Yes
Factory FE	No	No	No	No
Enumerator FE	Yes	Yes	Yes	Yes
Observations	707	707	708	708
Number factories	26	26	26	26

Notes: P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.11: Effect on production outcomes – with Selection experiment

	Efficiency		Alteration rates (%)		Absenteeism (%)	
	(1)	(2)	(3)	(4)	(5)	(6)
Attitude only training	0.09	0.21	0.19	0.25	0.18	0.17
	(0.95)	(0.88)	(0.64)	(0.55)	(0.50)	(0.54)
	[0.95]	[0.88]	[0.68]	[0.61]	[0.51]	[0.56]
Aptitude & Attitude training	0.11	0.13	−0.35	−0.32	0.23	0.21
	(0.91)	(0.91)	(0.21)	(0.26)	(0.45)	(0.49)
	[0.92]	[0.91]	[0.21]	[0.26]	[0.46]	[0.50]
Selection experiment	−0.88	−1.30	1.13**	1.07**	−0.45	−0.40
	(0.51)	(0.30)	(0.02)	(0.01)	(0.38)	(0.41)
		[0.35]		[0.05]		[0.44]
Aptitude (Index)		0.69		0.25		0.07
		(0.32)		(0.39)		(0.51)
		[0.36]		[0.55]		[0.57]
Attitude (Index)		0.18		−0.04		−0.30
		(0.68)		(0.81)		(0.14)
		[0.68]		[0.82]		[0.14]
Control mean	52.21	52.21	6.85	6.85	4.65	4.65
P-value(Attitude=Aptitude&Attitude training)	0.98	0.95	0.17	0.15	0.73	0.84
P-value (Attitude=Aptitude)		0.60		0.37		0.17
Imbalance controls	Yes	Yes	Yes	Yes	Yes	Yes
Factory FE	No	No	No	No	No	No
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Randomisation controls	Yes	Yes	Yes	Yes	Yes	Yes
Pre-treatment line mean	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5769	5769	5911	5911	5680	5680
Number factories	17	17	23	23	17	17

Notes: Intent-to-treat ANCOVA comparison within the 8 weeks trial period. P-values clustered at the factory level in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. * p<0.1, ** p<0.05, *** p<0.01

Chapter 3

Who Loses and Who Benefits from Large-scale Infrastructure Projects? Evidence from the Impact of India's Golden Quadrilateral Project on Incumbent Manufacturing Firms

Abstract India's Golden Quadrilateral was an ambitious infrastructure project that improved the quality and width of highways connecting four major Indian cities. This paper analyses the project's short-term impact on incumbent manufacturing firms using firm-level panel data. Treatment is defined as being located in a city that is less than 5 km away from the Golden Quadrilateral. The results show large negative effects on employment and assets of incumbent firms close to the Golden Quadrilateral. These aggregate effects mask substantial heterogeneity. The results are primarily driven by firms which at baseline were less profitable, less productive, smaller and young. For firms with these characteristics, I find significant and very large negative effects on sales, gross profits, and employment. In contrast, firms that were initially more successful saw their sales, gross profits and labour productivity increase, though not significantly so. These results, viewed together with existing evidence pointing towards large entry effects, provide suggestive evidence that the Golden Quadrilateral project increased competition along the improved highways even in the short term, and precipitated a reallocation of production towards more successful firms.

3.1 Introduction: Infrastructure, firms and development

It is often argued that firms depend on well-functioning systems of transport, electricity, water and communications to source inputs, produce goods or services, reach output markets and access productivity-improving ideas (Banerjee et al., 2020). Yet, insufficient infrastructure is a severe constraint to firm operations and growth in many developing countries (see e.g. Bigsten and Söderborn, 2006; Gollin and Rogerson, 2014; The World Bank, 2004). Infrastructure projects that aim to ameliorate these deficiencies are therefore high on the political agenda of both developing countries' governments and donors (see e.g. The World Bank, 2007).

However, empirically evaluating the causal effects of these infrastructure projects is difficult. The primary challenge is the often endogenous nature of infrastructure placement. Persuasive identification strategies are required to convincingly argue that observed outcomes are the result of infrastructure projects and dispel the suspicion that (potential) outcomes may have triggered infrastructure placement (Banerjee et al., 2020; Datta, 2012).

The existing empirical literature that has employed such identification strategies, discussed in detail in Section 3.3, has largely found positive impacts of transportation infrastructure on aggregate economic activity (Redding and Turner, 2015). It has also demonstrated, however, that improvements in infrastructure can affect economic activity through different mechanisms. Most directly, enhancing transport infrastructure can decrease firms' shipping costs for inputs and outputs. Indirectly, the improved market access can also lead to entry of new firms, exit of existing firms, productivity improvements in existing firms, and a reallocation of inputs and output between firms (Yang, 2018).

This paper focuses on the latter two potential mechanisms by investigating productivity improvements and a potential reallocation of resources for existing firms. I analyse the short-term impact of a large-scale infrastructure intervention, India's Golden Quadrilateral (GQ) project, on incumbent manufacturing firms. The Golden Quadrilateral project was an ambitious highway improvement programme implemented in the years 1999 to 2012 that aimed to better connect India's largest four cities – Delhi, Kolkata, Chennai, and Mumbai.

For the empirical analysis, I construct a panel of 674 manufacturing firms using data from the "Firm Analysis and Competitiveness Survey of India" (FACS) in 2002 and 2005. I employ a difference-in-differences approach to analyse the impact of the GQ project on the incumbent firms' output, inputs, profits, and productivity. In order to have a clear pre-intervention and post-intervention period, I define treatment as being located in a city that is less than 5 km away from the parts of the Golden Quadrilateral completed in between the two survey waves, from 2002 to 2004. 674 firms are included in the analysis, which are located outside the agglomeration areas of the targeted four cities. This strategy is called the "inconsequential units approach" by Redding and Turner (2015), and was introduced to the literature by Chandra and Thompson (2000) and Michaels (2008). The identifying assumption is thus that the GQ project is exogenous to time-varying characteristics of treated firms located in cities in between the four metropolises. I thus exploit that the GQ project explicitly aimed to improve connectivity for firms located in the four metropolises and that more productive firms may have moved to areas close to the GQ project.

I find that the Golden Quadrilateral project had a pronounced negative short-term impact on input usage of the incumbent firms, with employment and fixed assets declining by a quarter to a third. There are some indications of a reduction in sales and profits, though these are insignificant. I implement robustness checks with different specifications, different variable definitions, and different treatment definitions, which support the main findings. Moreover, I combine the difference-in-differences approach with instrumentation to account for a potential endogenous placement of the infrastructure improvements. I use proximity to straight lines connecting the four metropolises as an instrumental variable, an approach pioneered by Banerjee et al. (2020). This instrumental variable approach confirms the large negative impacts on firms' labour and capital usage.

These aggregate results mask substantial heterogeneity. Splitting the sample by median baseline profits demonstrates that the effects are primarily driven by firms which were less successful according to this measure.[1] These initially less profitable firms saw their

[1] I show that this heterogeneity also holds if splitting the sample by other baseline measures, e.g. median labour productivity, median number of employees, and being below four years of age. As can be expected, these measures are positively correlated with each other.

sales, profits, and labour inputs significantly decline by close to or more than half. In contrast, for the initially more profitable firms, I find suggestions that sales, profits, and labour productivity increased. Though these latter results are not individually significant, I can reject that the effects for the two groups are the same for sales and profits. Both groups appear to have shrunk in terms of labour and capital usage, but the initially more profitable firms were able to sustain this downsizing while at the same time increasing sales and net sales, and thus labour productivity.

Though data limitations preclude me from making definitive statements, the evidence in this paper, read in conjunction with the existing literature on the GQ project,[2] is consistent with a narrative in which the infrastructure improvements lead to increased competition in the now better connected areas. This competition results from the entry of new firms and an expansion of more successful existing firms, both brought about by improved market access. This increased competition also leads to a reallocation of production away from the initially less successful firms. Rather than demonstrating that infrastructure improvements destroy firms, I therefore interpret the evidence in this paper as providing a snapshot of the process of reallocation towards more successful firms when areas become better connected.

The paper is organised as follows. Section 3.2 briefly explains the Golden Quadrilateral project. Section 3.3 discusses this paper's contribution to the literature, and the existing evidence on the impact of the GQ project. Section 3.4 presents the empirical methodology and identification strategy used. Section 3.5 describes the data I use and Section 3.6 explains the construction of the main outcomes. Section 3.7 presents and discusses the results. Section 3.8 concludes.

3.2 The Golden Quadrilateral project

The Golden Quadrilateral project was an ambitious infrastructure development project designed to address deficiencies in India's transport system in the 1990s. High rates of economic growth in this decade combined with insufficient public investments had increased demand for transport services, especially for roads, that was not met by supply (Mohan,

[2]See Section 3.3 for a detailed discussion.

2004; The World Bank, 1995). At the end of the 1990s, a large share of India's extensive road network was characterised by insufficient lane capacity and/or poor surface quality, resulting in a deficient transport system that imposed high costs on firms (The World Bank, 2002).

The GQ project aimed to improve the connections between Delhi, Kolkata, Chennai and Mumbai, four major Indian cities. For 5,846 km of national highway, the project intended to improve the quality and to extend the capacity to four lanes (National Highways Authority of India, 2002). In addition to connecting the four above-mentioned metropolises, it also passed a number of major Indian cities, e.g. Bangalore, Kanpur and Viskhapatnam.[3]

The Golden Quadrilateral project was launched in 1999 as part of phase I of the National Highway Development Project (NHDP) and was originally scheduled to be completed in December 2004 (National Highways Authority of India, 1999).[4] Due to issues with land acquisition and re-negotiation of contracts, progress was significantly delayed. Only 1,063 km or 18.2 % of the improvement works had been completed by 31 March 2002 (National Highways Authority of India, 2002). These included improvement works that had already begun before the GQ project was launched, and which were later counted as part of the project.

The length of completed highway stretches had increased to 4,697 km or 80.3 % by the end of March 2005 (National Highways Authority of India, 2005). The dataset used in this paper, which surveys the years 2001 and 2004, therefore covers the main construction period. The entire project was, however, not declared complete until January 2012.[5] I demonstrate the different phases of construction in Figure 3.1 and discuss the existing evidence on the impact of the GQ project on the Indian economy in Section 3.3.

[3]The project also allowed the private infrastructure companies contracted on a build, operate, transfer model to collect tolls for a concessionary period after construction.

[4]Phase I of the NHDP also included the construction of North-South and East-West corridors, totalling 7,300 km. Thereof, 106 km are common between the GQ project and the North-South and East-West corridors (National Highways Authority of India, 2002). The progress on these corridors was, however, severely delayed. Until 31 March 2005, less than 10 % of the entire length of the North-South and East-West corridors had been completed, which partially coincide with the GQ (National Highways Authority of India, 2005).

[5]http://archive.indianexpress.com/news/govt-declares-golden-quadrilateral-complete/896873/, last accessed 12 February 2016.

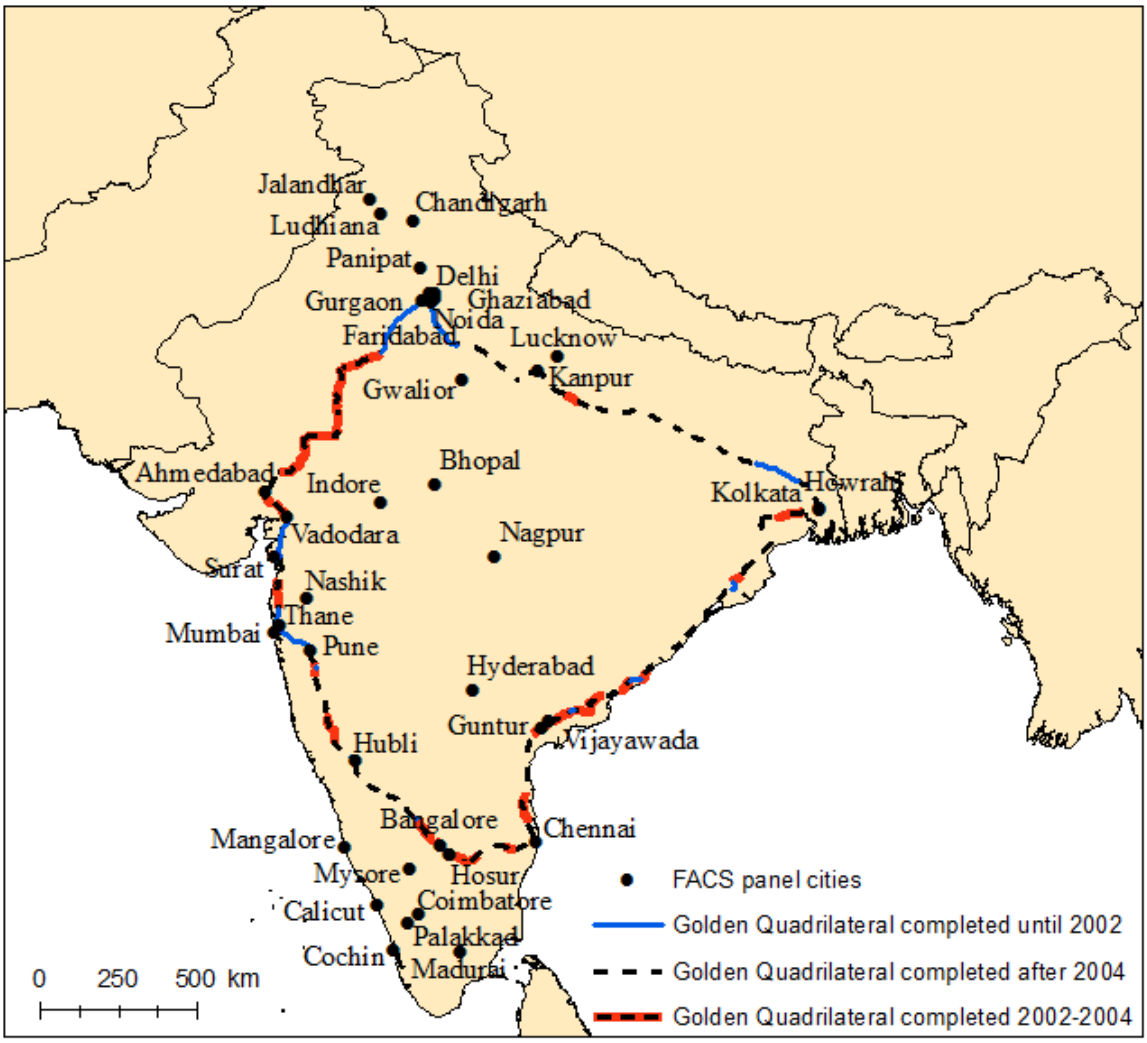


Figure 3.1: The Golden Quadrilateral project and surveyed panel cities

3.3 Contribution to the literature

This paper contributes to a large literature that evaluates the impact of transportation infrastructure on economic outcomes.[6] For this paper, analyses of major inter-city transportation infrastructure, such as highways and railways, are most relevant. Redding and Turner (2015) point out that most papers in this strand of literature have found positive impacts of transportation infrastructure on aggregate economic activity. For example, in high-income countries, Chandra and Thompson (2000) and Michaels (2008) find improvements in historical per-capita earnings in areas close to the newly built US interstate highway system. The picture is similar for low- and middle-income countries. Donaldson (2018) for India and Jedwab and Moradi (2016) for Sub-Saharan Africa both estimate positive effects of colonial railroads on aggregate income close to the network. Similarly, Banerjee et al. (2020) find that an expansion of Chinese road and rail infrastructure led to higher income levels, though not to higher growth rates.

Two rare exceptions to this positive pattern are Faber (2014), which estimates a negative impact on GDP in counties in proximity to new Chinese highways, and Khanna (2016). The latter analyses the Golden Quadrilateral project and finds that the strong negative relationship between luminosity (as a proxy for economic activity) and distance from the GQ areas in the 1990s becomes much weaker during and after the improvements of the highway system. This suggests that the highway improvements may have caused a relative shift of economic activity to areas further away from the Golden Quadrilateral.

The impacts of infrastructure interventions on aggregate economic activity just discussed can result from different potential mechanisms, such as entry of new firms, exit of existing firms, reallocation between continuing firms, and productivity improvements (Yang, 2018). The analyses of firm-level data attempt to clarify which of these mechanisms are at play, and have so far provided most support for the entry of new firms (e.g. Gibbons et al. (2019) for the UK, Yang (2018) for China, Melo et al. (2010) for Portugal). There is also some evidence for productivity improvements in existing firms, for example via improved

[6] See Redding and Turner (2015) for an excellent review.

supply or inventory practices (e.g. Gibbons et al. (2019), Holl (2016) for Spain, Li and Li (2013) for China).

The evidence on the impact of the Golden Quadrilateral so far suggests an increase in aggregate economic output for the formal sector, which is driven largely by new entry, some growth among established incumbents, and improvements in allocative efficiency. Ghani et al. (2016) use data from the Annual Survey of Industries between 1994-2007 that they aggregate at the district level. Their results point to an expansion of total manufacturing output in districts less than ten kilometres away from the GQ project. The authors show that this expansion is due both to higher rates of new firm entry as well as to growth in the number of workers and plants for firms of at least ten years of age at the start of the project. For the informal manufacturing sector, however, evidence in Ghani et al. (2013) shows no significant impact of the GQ project in either rural or urban areas. Additional evidence of the GQ effect from Datta (2012) – using the same survey data as this paper – indicate that manufacturing firms in cities close to the improved highways significantly reduced their average input inventory holdings and, in some specifications, were more likely to change their suppliers. Asturias et al. (2019), using a model of internal trade with variable mark-ups, investigate potential welfare gains from the GQ project through different channels. Using data from the Annual Survey of Industries and the National Sample Survey, they find that gains in allocative efficiency account for 7.4 % of the real income gains in the manufacturing sector.

Similar to Asturias et al. (2019), this paper focuses on documenting the short-term process of reallocation FACS, which allows me to construct a panel of firms to follow over time. I focus on analysing the surviving incumbents, evaluate potential productivity improvements in existing firms and a potential reallocation of production in the short term. As Asturias et al. (2019), I observe that resources move from less successful to more successful firms in the process of reallocation. Thanks to the panel structure of the firm-level data, this paper provides a more micro-level snapshot of who these less successful firms are, and how the reallocation takes place.

3.4 Empirical strategy

I employ a two-period difference-in-differences approach to estimate the short-term impact of the Golden Quadrilateral project on incumbent firms. The main analysis focuses on those parts of the GQ project that were completed between the two survey waves of the FACS in order to have a clear pre-intervention and post-intervention period. The analysis therefore evaluates the short-term impacts of the GQ project as it applies to the years 2002 - 2004. Figure 3.1 demonstrates the intervention, which constitutes more than 60 % of the entire GQ project. In Section 3.7.3, I show that my results are robust to using different definitions of the intervention.

I employ the mentioned inconsequential units approach (Redding and Turner, 2015) by only including firms that were located outside the four metropolitan areas that were targeted by the Golden Quadrilateral project in the analysis.[7] Firms in the agglomeration areas of up to 50 km distance around the centroids of Delhi, Kolkata, Chennai and Mumbai are therefore excluded from the analysis. I call cities in these agglomeration areas "nodal" and those outside these agglomeration areas "non-nodal" cities, following Datta (2012).

I define the treatment group as those firms located in non-nodal cities that are in up to 5 km geographical distance from the parts of the GQ project completed between 2002 and 2004. This definition is primarily driven by data considerations. Since the FACS presents city names as the only location information for each firm, the data provide limited variation with respect to distance from the improved highways. More specifically, there are a number of cities within 5 km distance to the improved highways, but no cities within 5-15 km. Therefore, results do not change when re-defining treatment as being located in cities that are, for example, in up to 8 km, 10 km, 12 km or 15 km distance from the newly improved highways. Firms in non-nodal cities that are more than 5 km distance from any part of the GQ project constitute the control group. This means that firms close to parts of the GQ project that were improved before 2002 and after 2004 are excluded from the main

[7] This identification strategy was introduced by Chandra and Thompson (2000) and Michaels (2008), and also adopted by Datta (2012) and Ghani et al. (2016, 2013) when analysing the GQ project.

analysis.[8] As the firms in the treatment group, the firms in the control group are located in large cities that are connected to main – but not improved – roads.

I estimate the impact of the Golden Quadrilateral project with a standard difference-in-differences model using first differences. The main specification is

$$\Delta y_{it} = \beta_0 + \beta_1 \cdot nearGQ_i + \boldsymbol{\gamma}' \cdot \mathbf{X}_{it-1} + \Delta\eta_{it}. \tag{3.1}$$

y_{it} is an outcome for firm i in period t, $t = 0$ in 2001 and $t = 1$ in 2004, $nearGQ_i$ is the treatment variable defined as above, $\mathbf{X}_{it-1}$ is a column vector of control variables and η_{it} is an idiosyncratic error term. $\mathbf{X}_{it-1}$ for the main analysis consists of dummy variables for industry, state, type of ownership and legal status, as well as firm age, distance from the nearest nodal city, ln(Sales), ln(Permanent employees), and ln(Fixed assets at net book value).[9]

Errors are clustered at the city level to account for intra-city correlation. I also implement wild cluster percentile-t bootstrap procedures for the coefficient of interest, following Cameron et al. (2008) and Cameron and Miller (2015). In the main analysis, observations are weighted according to their inverse probability of selection into the sample to account for attrition, which is discussed in Section 3.5. This approach accounts for non-random attrition driven by observables, as suggested by Wooldridge (2002), and gives observations with a low probability of being in the sample a larger weight in the estimation.

β_1 in equation 3.1 is the coefficient of interest that captures the impact of the GQ project on firm outcomes. The key identifying assumption for the coefficient of interest to identify a causal effect of the intervention is that, conditional on the control variables, treatment and control groups would have followed a common time trend in the absence of treatment. This assumption implies that, conditional on control variables, treatment is exogenous with regards to time-varying firm characteristics.

[8] In the main analysis, I exclude these firms to be able to isolate the short-term effect from longer-term effects and anticipation effects. Section 3.7.3 presents results including these firms.

[9] To include control variables with missing values, I set missing values to zero and add dummy variables indicating a missing value to the set of controls.

By focusing on firms in cities that were incidentally treated because they lay alongside existing highways that connected the four targeted metropolitan areas of the Golden Quadrilateral project, this identification strategy exploits the fact that the project specifically aimed at improving the connectivity for firms in the four metropolises. The project may thus have been triggered by, for example, expected higher growth of firms in these four cities. This identification strategy also allows for spatial re-allocation on the basis of firms' time-invariant characteristics, e.g. that more productive or more innovative firms may have moved to areas near the GQ project in anticipation of the highway improvements.

A common trend is more likely if treatment and control groups are similar at baseline, such that the control group represents a credible counterfactual for the treatment group in the absence of treatment. Baseline balance is therefore investigated in Section 3.5. Using recall values, I visually investigate time trends of outcome variables for treatment and control groups in pre-intervention periods in Section 3.7.1.

Even when excluding nodal cities, identification would be threatened if the project were endogenously placed in order to reach non-nodal cities that were either expected to grow or in decline, both of which would violate the common trend assumption for non-nodal firms. Since the GQ project primarily consisted of improvement works and hence followed previously existing highway routes, this consideration is of a lesser concern in this analysis than for newly constructed infrastructure. However, as a robustness check, I combine a difference-in-differences approach with instrumentation to address this concern. I employ an instrumental variable introduced by Banerjee et al. (2020) who analyse the macroeconomic effects of transportation infrastructure in China. The authors use the distance to straight lines connecting historical cities as instrumental variables to obtain plausibly exogenous variation for the improvement of infrastructure.

Similarly, I draw a quadrilateral of straight lines from Delhi to Kolkata, Kolkata to Chennai, Chennai to Mumbai and from Mumbai to Delhi in order to obtain the bird's-eye shortest connections between these cities. In the estimation of equation 3.1, I instrument $nearGQ_i$ with a dummy to indicate whether firms are up to 250 km away from this straight line

instrumental variable. Since the Golden Quadrilateral project did not deviate substantially from the straight lines connecting the four cities, the instrument is relevant. Its strength is discussed in Section 3.7.3.

3.5 Data

I use two sources of data. The firm data comes from the "Firm Analysis and Competitiveness Survey of India" in 2002 and 2005, a joint undertaking between the World Bank and the Confederation of Indian Industry.[10] The two waves survey the years 2001 and 2004, respectively.[11] The questionnaires in both waves consisted of two parts: one part about general firm information to be answered by the manager and one part covering financial statements and employment information to be answered by the accountant and the personnel manager, respectively. Recall values for the two years preceding 2001 and 2004 were also recorded for key variables, though suffer from significant non-response.[12] Importantly, the name of the city is the only information about firms' location included in the survey, which limits the possibilities of spatial analysis.

To obtain the necessary geographical information about proximity to the GQ project, I use the Natural Earth data set (version 2.0.0). With highway maps from the official National Highways Authority of India (NHAI) website[13] and information about the construction progress from NHAI's annual reports, I traced out the entire GQ project as well as the parts that were completed between 2002 and 2004 in ArcGIS. I subsequently obtained the geographical distances in kilometres of each city included in the survey from the highways

[10] I used this dataset for manufacturing firms for two main reasons. First, manufacturing firms are arguably more likely to rely on transportation infrastructure for input sourcing and output marketing than e.g. services firms. Therefore, I expected the short-term effects which are the focus in this paper to be more pronounced for manufacturing firms. Possible effects on structural transformation, such as a shift from manufacturing to services, may take longer to materialise. Second, the dataset was available free of charge for a graduate student.

[11] The 2002 sample was drawn using stratified systematic sampling from sample frames containing establishments with at least ten full-time employees across 40 cities and 12 manufacturing industries, in addition to the software industry (The World Bank, 2004; Ferrari and Dhingra, 2008). The stratification took place at the state level according to each state's share in national manufacturing gross domestic product in 1998/99. Because the data do not contain sampling weights, I am unable to correct for these in my analysis to achieve representativeness.

[12] Therefore, a total of six years of financial statements and employment information from 1999 to 2004 are available for the subset of firms for which these recall values were successfully collected. The recall values are used in Section 3.7.2 to investigate the key identifying assumption of parallel trends.

[13] http://www.nhai.org/ (last accessed 22 March 2014)

Table 3.1: Incidence of attrition (in %)

	Non-attriter	Attriter	Total
Control	62.92	37.08	100.00
Treated	78.38	21.62	100.00
Total	66.67	33.33	100.00

Notes: The table shows the incidence of attrition across treatment and control. Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral project completed in 2002-2004, control as being more than 5 km geographical distance away from any part of the Golden Quadrilateral. Nodal cities are excluded.

forming the entire GQ project, the GQ project finished between 2002 and 2004, the straight-line instrumental variable explained in Section 3.4, and each of the nodal cities.

Following the empirical strategy in Section 3.4, I combine the two datasets to define those manufacturing firms as treated that in the first survey wave were located in non-nodal cities which are up to 5 km away from the parts of the GQ completed between 2002 and 2004. The control group are those manufacturing firms in non-nodal cities which are more than 5 km away from any part of the Golden Quadrilateral. In wave 1, the FACS surveyed 1,068 firms that are part of either the treated or the control group. 712 of these firms were re-interviewed for the second wave in 2005, an overall attrition rate of 33 %. After I drop 38 firms for which many variables containing financial information were missing in 2005, I obtain 674 firms that were interviewed in both waves. This is the sample of firms on which I focus for the main analysis.

Table 3.1 shows that attrition differs between treatment and control. Whereas the control group has a rate of attrition of 37 %, only 22 % of the treatment group attrited. Unfortunately, the survey documentation does not provide any information about the reasons for this attrition, which precludes me from making an argument about firm exit. However, a substantial amount of this attrition is explained by observable firm characteristics, as seen in Table 3.2. The table shows probit regressions of a dummy variable indicating attrition on the treatment variable $nearGQ$ and baseline firm characteristics, for the treated and control firms surveyed in the first wave. As seen in column (2), the coefficient of the treatment

Table 3.2: Dependent variable: Attrition

	(1) Probit	(2) Probit
nearGQ	−0.16 **	−0.10
	(0.02)	(0.12)
Firm age (in decades)		−0.04 ***
		(0.00)
ln(Sales)		0.00
		(0.88)
ln(Net Sales)		−0.01
		(0.43)
ln(Permanent employees)		−0.01
		(0.48)
ln(Fixed assets)		0.01*
		(0.09)
Exporter (dummy)		0.02
		(0.54)
Importer (dummy)		0.06*
		(0.09)
ln(Investment)		0.00
		(0.78)
ln(R&D)		0.01
		(0.33)
Manager experience		0.00
		(0.14)
Controls:		
Industry	No	Yes
Legal status	No	Yes
States	No	Yes
Distance to nearest nodal city	No	Yes
Observations	1068	1068
Pseudo R-squared	0.02	0.08

Notes: The probit regressions test whether attrition is significantly related to being near the Golden Quadrilateral. The table presents average marginal effects for continuous variables and marginal effects for a discrete change for discrete variables. P-values in parentheses, with standard errors clustered at the city level. All non-categorical variables are winsorised at 1 % in each tail. For variables with missing values, I set missing values to zero and include indicator dummies for missing values. NBV = net book value. * p<0.1, ** p<0.05, *** p<0.01

variable is reduced by more than a third when introducing firm baseline controls, and loses significance.[14] As suggested by Wooldridge (2002), I use inverse probability weighting in the analysis in section 3.7 to account for this attrition. I do so by first calculating the predicted probability of attrition from column (2) of Table 3.2. I then obtain the predicted probability of selection into the sample as the complement of the probability of attrition. Finally, I obtain the inverse probability weights as the reciprocal of the probability of selection into the sample and winsorize the weights at the first and 99th percentile to reduce the influence of outliers. As compared to an unweighted regression, this procedure gives more weight to observations with a low probability of selection into the sample, to account for potential biases introduced by non-random attrition.

I next present summary statistics and balance tests for the sample of 674 firms included in the main analysis. As Table 3.3 shows, 196 of these firms are in the treatment group. They are located in the cities of Guntur, Vijayawada, Vadodara, Ahmedabad and Hosur in the states of Andhra Pradesh, Gujarat and Tamil Nadu. The control group consists of 478 firms in 19 cities and 11 states. The map in Figure 3.1 demonstrates the location of these cities in relation to the GQ project.

Table 3.4 shows the breakdown across manufacturing industries and tests of mean-equality across treatment and control, with p-values obtained from regressions weighted by the inverse probability of selection. The mean for the whole sample in column (1) indicates that no industry is dominant, with no industry representing more than 15 % of the sample. However, the separate means for the control group in column (2) and the treatment group in column (3) demonstrate that there are some differences between the groups. The treatment group is composed of significantly fewer firms in the garments and leather sector, and of significantly more firms in the pharmaceutical sector compared to the control group.

The median and mean of 22 crucial firm characteristics of the sample are presented in columns (1) and (2) in Table 3.5. These show that the sample primarily consists of small and

[14] Note that attrition is positively related to fixed assets and being an importer, though negatively related to firm age. These patterns suggest that attrition is not simply driven by firm exit.

Table 3.3: Distribution of sample firms (N = 674) across non-nodal cities at baseline in 2001

City	State or Union Territory	Number of control firms	Number of treated firms
Hyderabad	Andhra Pradesh	54	
Vijayawada	Andhra Pradesh		42
Guntur	Andhra Pradesh		12
Chandigargh	Chandigargh	17	
Ahmedabad	Gujarat		58
Vadodara	Gujarat		65
Surat	Gujarat	35	
Panipat	Haryana	22	
Mysore	Karnataka	24	
Mangalore	Karnataka	27	
Palakkad	Kerala	14	
Calicut	Kerala	12	
Cochin	Kerala	16	
Bhopal	Madhya Pradesh	25	
Gwalior	Madhya Pradesh	25	
Indore	Madhya Pradesh	36	
Nashik	Maharashtra	17	
Nagpur	Maharashtra	23	
Ludhiana	Punjab	36	
Jalandhar	Punjab	20	
Coimbatore	Tamil Nadu	38	
Hosur	Tamil Nadu		19
Madurai	Tamil Nadu	20	
Lucknow	Uttar Pradesh	17	
Total		**478**	**196**

Note: Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral built between 2002 and 2004, control as being more than 5 km geographical distance away from any part of the Golden Quadrilateral.

Table 3.4: Sample firm (N = 674) industry composition and mean-equality tests at baseline in 2001

	Whole sample (Mean)	**Control** (Mean)	**Treatment** (Mean)	**Mean equality p-value**
Garments	0.09	0.14	0.02	0.07*
Textiles	0.13	0.13	0.12	0.92
Drugs and pharma	0.12	0.07	0.24	0.00***
Electronics	0.05	0.06	0.07	0.90
Electrical appliances	0.08	0.07	0.08	0.96
Machine tools	0.06	0.04	0.08	0.34
Auto components	0.12	0.11	0.16	0.40
Leather/leather products	0.02	0.02	0.00	0.05**
Sugar	0.00	0.00	0.01	0.39
Food processing	0.11	0.13	0.05	0.09*
Chemicals	0.14	0.13	0.11	0.45
Metals	0.07	0.08	0.06	0.50

Notes: Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral project completed in 2002-2004, control as being more than 5 km geographical distance away from any part of the Golden Quadrilateral. Nodal cities are excluded. For each industry, the p-value for mean equality is obtained from a regression of each industry dummy on treatment, with standard errors clustered at the city level. For the mean-equality tests, observations are weighted by their inverse probability of selection into the sample. * p<0.1, ** p<0.05, *** p<0.01

medium enterprises with a median number of permanent employees of 15 and a median age of 12 years. Distributions of permanent employees, annual sales (as an approximation for output), annual net sales (as an approximation for profits) and capital defined as fixed assets at net book value are positively skewed. The majority of firms are non-listed limited liability companies with private, domestic owners.[15]

Out of the 22 variables for which I test mean equality in columns (3) to (5) in Table 3.5, I find significant differences in means of the treatment and control group for two variables. This is in line with what one would expect to find by chance. Treated firms are about three years older than control firms, and they are one percentage point more likely to be owned by the government. I also observe that treated firms are larger in terms of sales, net

[15]Similar to other low- and middle-income countries, a large share of the manufacturing sector in India consists of micro, small and medium enterprises. Data from representative surveys of both firms larger than ten employees (from the Annual Survey of Industries) and firms smaller than this size (from the Unorganized Manufacture Survey of the National Sample Survey Organisation) in Nataraj (2010) indicates that firms larger than ten employees accounted for approximately 80 % of output, 20 % of employment and 1 % of all firms in 1999. The industry composition of firms with more than ten employees presented in Nataraj (2010) appears similar to the industry composition in the FACS sample, though this is difficult to confirm with certainty because the surveys use different industry classifications.

Table 3.5: Sample firm characteristics and mean-equality tests at baseline in 2001

	Whole sample		Control	Treated	Mean equality
	Median	Mean	Mean	Mean	p-value
	(N = 674)		(N = 478)	(N = 196)	
Firm age	12.00	15.11	13.65	16.82	0.01***
Sales	4993.07	76953.46	54434.98	158882.26	0.29
Net sales	1622.36	34459.50	24960.62	71423.45	0.37
Permanent employees	15.00	59.35	51.09	75.72	0.31
Fixed assets at NBV	1889.23	25045.98	19013.56	48571.79	0.35
Capital-labour ratio	104.66	208.74	219.55	200.68	0.76
Capital-output ratio	0.32	0.56	0.52	0.64	0.46
Labour-output ratio	0.00	0.01	0.01	0.01	0.15
Investment	0.00	1494.80	1588.24	1877.32	0.87
Investment (dummy)	0.00	0.44	0.42	0.49	0.42
R&D spending	0.00	259.74	166.97	616.27	0.15
Ownership (dummies)					
private, domestic	1.00	0.98	0.98	1.00	0.11
private, foreign	0.00	0.00	0.00	0.00	0.14
governmental	0.00	0.01	0.01	0.00	0.05**
Legal status (dummies)					
listed LLC	0.00	0.15	0.14	0.15	0.90
Non-listed LLC	1.00	0.57	0.52	0.63	0.45
Other legal status	0.00	0.28	0.34	0.22	0.44
Manager experience (years)	6.00	8.64	8.64	8.23	0.66
Exporter (dummy)	0.00	0.26	0.26	0.27	0.95
Importer (dummy)	0.00	0.16	0.19	0.15	0.61
Unionisation	0.00	6.25	5.74	6.71	0.76
Distance to nodal city	426.43	415.12	436.24	385.13	0.26

Notes: Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral project completed in 2002-2004, control as being more than 5 km geographical distance away from any part of the Golden Quadrilateral. Nodal cities are excluded. For each variable, the p-value for mean equality is obtained from a regression of each industry dummy on treatment, with standard errors clustered at the city level. For the mean-equality tests, observations are weighted by their inverse probability of selection into the sample. All non-categorical variables are winsorised at 1 % in each tail. All monetary values are in real terms and thousands Indian rupees. Net sales are sales net of employment and material cost. NBV = net book value. LLC = limited liability company. R&D = research and development. * p<0.1, ** p<0.05, *** p<0.01

sales, employees, fixed assets and spending on research and development, though these differences are not statistically significant.

I use a first-differenced estimator to control for potential differences in outcomes arising from differences in time-invariant characteristics and include a wide range of baseline control variables to account for potentially different time trends resulting from these differences in firm characteristics.

Table 3.6: Definition of main outcome variables

Variable	Definition
Sales	Annual sales in thousand Indian Rupees (INR), deflated using the wholesale price index (WPI) for the manufacturing sector with a base year of 1993-94.†
Net Sales	Sales minus annual total employee and material costs in thousand INR. Sales is defined as above. Annual total employee costs are the total cost of labour, deflated using the consumer price index (CPI) with a base year of 2005.‡ Material costs are the total purchases of materials, deflated using the WPI for the manufacturing sector with a base year of 1993-94.†
Permanent employees	Number of permanent employees. In the first wave calculated as number of total employees minus number of temporary employees, enumerated directly in the second wave.§
Fixed assets at net book value	Net book value of property, plant and equipment in thousand INR, deflated using the WPI for the manufacturing sector with a base year of 1993-94.†
Capital-labour ratio	Ratio of Fixed assets at net book value and Permanent employees.
Output-labour ratio	Ratio of Sales and Permanent employees.

Notes: All variables are winsorised at 1 % in each tail across all years after the appropriate transformation (log for the main analysis, inverse hyperbolic sine for robustness).
† I obtained the WPI data from the website of the Office of the Economic Adviser to the Government of India, http://www.eaindustry.nic.in/Download_Data_9394.html (last accessed 20 March 2014).
‡ The CPI data comes from the World Development Indicators, version of 18 December 2013.
§ The 2002 wave enquired about the number of workers *at the end of the fiscal year 2001*, the 2005 questionnaire pertained to the number of workers *as an average over the year 2004*. I argue that these differences will not have a significant bearing on my results (i) because of limited seasonality in manufacturing businesses and (ii) because employing a difference-in-differences approach will efface these differences if treatment and control group are equally affected.

3.6 Main outcomes of interest

The data described in the previous section allow me to investigate the impact of the Golden Quadrilateral project on potential productivity improvements of incumbent firms, and a potential reallocation of production. Due to data limitations, I am unable to investigate firm entry and exit. My main outcomes are six measures of firm output, inputs, profitability, and productivity. These are defined in Table 3.6. Specifically, I analyse the impact on the firms' annual sales as a proxy for output and their annual net sales as a proxy for profits. Net sales are defined as sales minus employment and material costs.[16]

I use the number of permanent employees as a measure of labour input, and fixed assets

[16] I am unable to obtain gross or net profit figures since the data do not provide measures of cost of goods sold and total costs, respectively, and therefore employ this proxy.

at net book value (NBV) as a measure of capital inputs.[17] The capital-labour ratio, defined as fixed assets at NBV divided by the number of permanent employees, is a proxy for the capital intensity of the firms' production. As a measure of labour productivity, I calculate the output-labour ratio by dividing sales by the number of permanent employees.

All monetary outcomes are deflated as described in Table 3.6. All outcomes are winsorised at 1 % in each tail of the distribution. It should be noted that the FACS relies on data that is self-reported by the firms participating in the survey, which is likely to introduce some measurement error. Nevertheless, financial statements and employment information were answered by the accountant and the personnel manager of each firm, which makes it more likely that the survey respondents had the accurate information at hand when responding.

3.7 Results

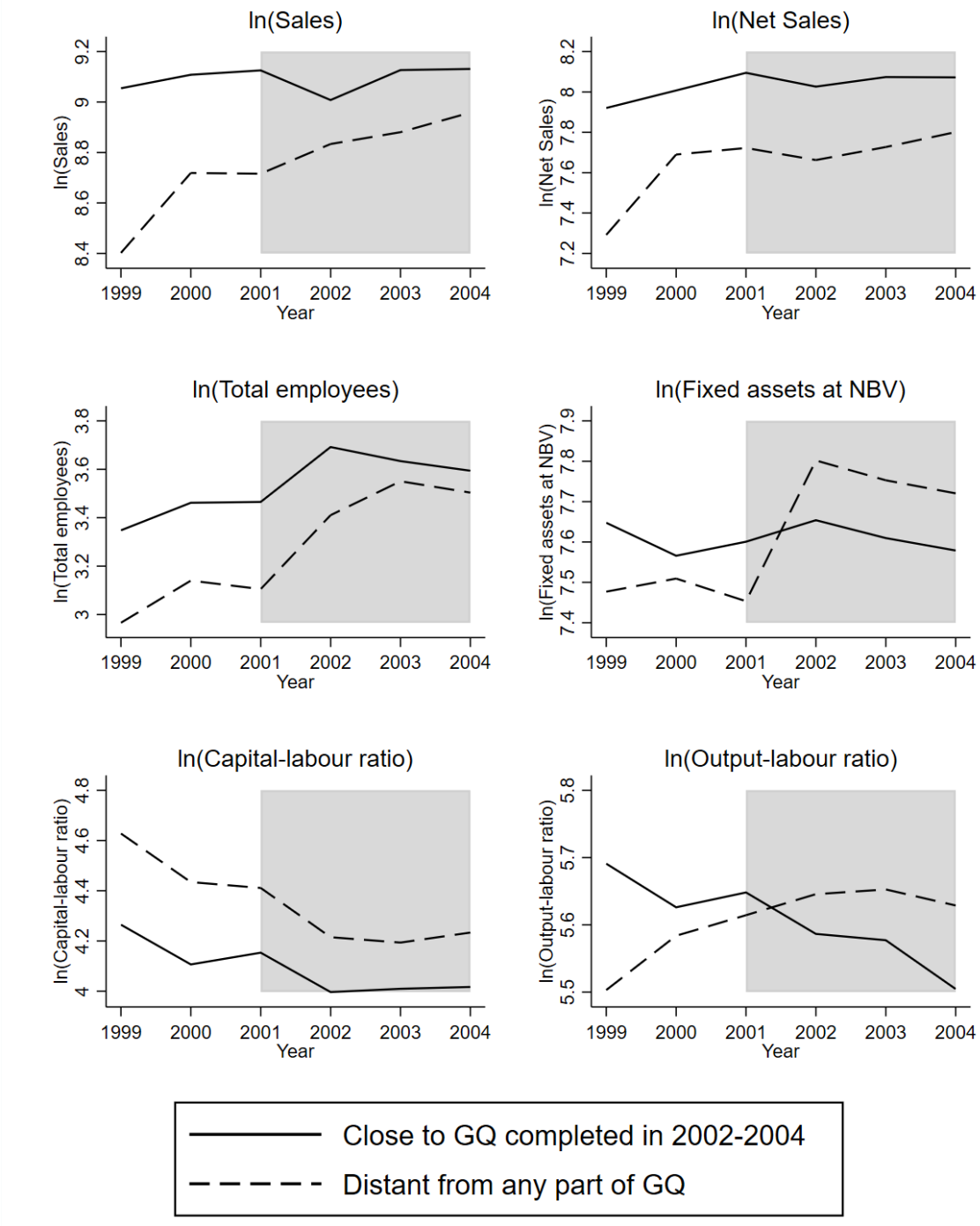
3.7.1 Visual inspection

I first visually inspect the main outcomes and evaluate the common trend assumption. Using recall values, Figure 3.2 illustrates trends for the mean of the main outcomes over the period 1999-2004 for treatment and control groups separately. The recall values suffer from frequent non-response, which means that not all firms are included in each year. Note that I use the number of total employees to illustrate trends for labour inputs for the figure since the number of permanent employees is not available for the recall years before 2001. The treatment period is shaded in grey.

Not conditioning on any control variables, Figure 3.2 provides supportive evidence for the common trend assumption in pre-treatment periods for most of the outcomes. Total employment and the capital-labour ratio appear to be following very similar trends in the treatment and control groups in the pre-treatment period, and the trends differ slightly for sales, net sales and fixed assets. The difference in trends for the output-labour ratio is, however, evident. The formal analysis below will therefore control for a number of baseline variables to account for potential differential trends.

[17] Because I lack a time series of investment, I am unable to estimate the capital stock using the perpetual inventory method (cf. Berlemann and Wesselhöft, 2014).

Figure 3.2: Trends for main outcomes



Notes: The figure shows time series for the mean of key variables for treatment and control groups amongst sample firms excluding nodal cities (N = 674). Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral project completed in 2002-2004. The pre-treatment period is thus 1991-2001. Values in 2001 and 2004 are from surveys for the respective years, values for other years are recall values. Note that not all firms are included in each year due to missing values.

Table 3.7: Main results

	(1) ln(Sales)	(2) ln(Net sales)	(3) ln(Permanent employees)	(4) ln(Fixed assets)	(5) ln(Capital-labour ratio)	(6) ln(Output-labour ratio)
nearGQ	−0.21	−0.09	−0.26 ***	−0.29 ***	0.11	0.05
	(0.37)	(0.71)	(0.00)	(0.01)	(0.45)	(0.83)
	[0.58]	[0.81]	[0.03]	[0.05]	[0.55]	[0.91]
Observations	671	499	604	505	458	601
Adjusted R-sq	0.19	0.16	0.34	0.32	0.37	0.29

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes. Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral project defined at the bottom of each column, control as being more than 5 km geographical distance away from any part of the Golden Quadrilateral. Baseline controls included are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. Observations are weighted by inverse probability of non-attrition. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. NBV = net book value. * p<0.1, ** p<0.05, *** p<0.01

Figure 3.2 also suggests that the GQ project had a pronounced impact on some of the outcomes. Specifically, the treatment group exhibits a steeper downward trend in employment in the treated period than the control group, as well as a less pronounced increase in fixed assets. I also observe some indication of smaller growth in sales and net sales of the treatment group compared to the control group. The trends of the capital-labour ratio and the output-labout ratio appear largely unaffected. Because of large numbers of missing values in the recall periods, the subsequent analysis focuses on a two-period difference-in-differences estimation between 2001 and 2004.

3.7.2 Main results

The results in Table 3.7 demonstrate the main findings from the estimation of equation 3.1. For the incumbent manufacturing firms in the sample, being close to the Golden Quadrilateral project had a pronounced negative effect on input usage. As columns (3) and (4) indicate, the effects on the number of permanent employees and fixed assets are large at -0.26 and -0.29 log points, respectively. These effects are highly significant, and remain so when I implement the wild percentile-t bootstrap to account for small number of clusters (bootstrapped p-values are shown in square brackets). These results suggest that, in terms of input usage, the incumbent manufacturing firms in the sample shrank by close to a third

Table 3.8: Other outcomes

Panel A

	(1) New hires	(2) Dismissals	(3) ln(Search time for production worker)	(4) ln(Search time for non-production worker)	(5) ln(Search time for manager)
nearGQ	1.89	1.20***	−0.03	−0.07	0.18
	(0.57)	(0.00)	(0.76)	(0.29)	(0.24)
	[0.63]	[0.03]	[0.82]	[0.42]	[0.44]
Observations	476	430	451	388	284
Adjusted R-sq	0.09	0.05	0.09	0.16	0.14

Panel B

	(1) ln(Search time for technician)	(2) Investment (dummy)	(3) Disinvestment (dummy)	(4) Exporter (dummy)	(5) Importer (dummy)
nearGQ	0.15*	0.13	0.09*	0.10***	−0.05 **
	(0.09)	(0.24)	(0.07)	(0.00)	(0.04)
	[0.21]	[0.43]	[0.18]	[0.03]	[0.04]
Observations	391	492	421	653	647
Adjusted R-sq	0.12	0.07	0.14	0.05	0.07

Panel C

	(1) ln(Materials cost (Share in sales))	(2) ln(Employment cost (Share in sales))	(3) ln(Employment cost per employee)	(4) ln(Main material price)	(5) ln(Main product price)
nearGQ	0.11*	−0.03	−0.06	0.74	0.34
	(0.07)	(0.86)	(0.71)	(0.57)	(0.69)
	[0.23]	[0.88]	[0.76]	[0.66]	[0.75]
Observations	654	598	540	208	216
Adjusted R-sq	0.04	0.14	0.26	0.04	0.06

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes. Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral project defined at the bottom of each column, control as being more than 5 km geographical distance away from any part of the Golden Quadrilateral. Baseline controls included are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. Observations are weighted by inverse probability of non-attrition. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. NBV = net book value. * p<0.1, ** p<0.05, *** p<0.01

of their previous size.

The other four main outcomes are not significantly affected, as shown in the remaining columns of Table 3.7. The effect on sales is negative and at -0.21 log points similar to the effect size on inputs, but the estimate is too noisy to be significant. In comparison, the impacts on net sales, the capital-labour ratio and the output-labour ratio are smaller and insignificant.

To understand the impacts of the GQ project better, I investigate additional outcomes derived from the survey data in Table 3.8. Note that some of these variables have frequent non-response, but their examination is nevertheless instructive to understand the observed

main results. Panel A and column (1) of Panel B show additional variables related to labour inputs, such as new hires made by the firms, dismissals, and the time it took firms to fill a vacancy for different groups of workers. The estimates indicate that firms were both more likely to dismiss workers (Panel A, column (2)), and that the search time for skilled, but not unskilled, workers increased (Panel B, column (1) and Panel A, column (3)-(5)). Note that the point estimate of the impact on new hires is positive and larger in absolute terms than on worker dismissals, but it is insignificant. This points to potentially heterogeneous impacts, which are discussed further below. Overall, these variables supports the finding that employment decreased as a result of the GQ project for the incumbent firms in this sample.

Columns (2) and (3) of Panel B present additional outcomes regarding firms' capital accumulation, namely whether firms reported that they were investing or disinvesting. Both coefficients are insignificant after accounting for the small number of clusters, though their size could suggest that firms close to the GQ projects may be adjusting their capital stock more than firms located further away by both investing and disinvesting.

The next two columns – Panel B, columns (4) and (5) – show that the firms' international trading behaviour significantly changed. I find that the GQ project leads to a higher likelihood of firms exporting their products by 0.10 percentage points, and to a lower likelihood of directly importing inputs by 0.05 percentage points. These results could suggest that the improved transport links enabled firms to participate in international trade as an exporter. The negative impact on importing fits the overall pattern of firms reducing the inputs employed which I observed in Table 3.7.

Lastly, Panel C of Table 3.8 investigates outcomes related to firms' costs. Column (1) suggests that the share of material costs in sales could have increased. Though this is not robust to correcting for the small number of clusters, this result – and the negative impacts on labour and capital – could indicate that firms are sourcing more inputs outside of the firm rather than producing them internally. In contrast, I don't find any impact on the share of labour costs in sales or the average costs per employee in columns (2) and (3), respectively. The last two columns present estimates for the prices of the main material input

Table 3.9: Main results for firms with no missing values for outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Sales)	ln(Net sales)	ln(Permanent employees)	ln(Fixed assets)	ln(Capital-labour ratio)	ln(Output-labour ratio)
nearGQ	−0.02	−0.11	−0.32 **	0.01	0.28	0.33*
	(0.94)	(0.69)	(0.02)	(0.96)	(0.17)	(0.05)
	[0.97]	[0.80]	[0.05]	[0.96]	[0.37]	[0.28]
Observations	348	348	348	348	348	348
Adjusted R-sq	0.12	0.18	0.40	0.32	0.33	0.37

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes. The sample is restricted to firms for which the outcomes in Table 3.7 are not missing. Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral project defined at the bottom of each column, control as being more than 5 km geographical distance away from any part of the Golden Quadrilateral. Baseline controls included are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. Observations are weighted by inverse probability of non-attrition. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. NBV = net book value. * p<0.1, ** p<0.05, *** p<0.01

and the firms' main output, though note that this information is only available for about a third of the sample. The coefficients are both large and suggest that both prices could have increased, but the small sample size prevents me from drawing reliable conclusions.

3.7.3 Robustness checks

A variety of robustness checks largely confirm the main findings of large negative effects on employment and assets of the incumbent firms. I start by showing the main results for the subsample of firms for which none of the main outcome variables are missing in Table 3.9.[18] As can be seen in column (3), the negative effect on employment is even more pronounced at -0.32 log points for this subsample than for the full sample in Table 3.7. Nevertheless, for this subsample, I do not observe a decline in fixed assets or sales. I do, however, find a large positive coefficient on the output-labour ratio as a proxy for labour productivity, though the effect is not significant after correcting for a small number of clusters. This exercise suggests that there is heterogeneity in the effects of the GQ project, and that some firms might have been able to reduce the size of their labour force without im-

[18]Table 3.A.1 in the appendix shows the results for the additional outcomes for the subsample of firms for which none of the main outcome variables are missing. The patterns are largely similar to Table 3.8. However, the point estimate for new hires in this subsample is negative and large at -3.55 hires (though insignificant), and the positive estimate for the share of material costs in sales is much smaller and insignificant.

pacting their output by increasing productivity. It also indicates that the non-response to questions likely is non-random. The subsample of 348 firms with no missing outcomes could potentially be the more productive or better managed firms. I will discuss this potential heterogeneity in detail below.

I also check the robustness of the large negative effects on employment and assets to different specifications, different definitions of the outcome variables, different definitions of the treatment, and the instrumental variable strategy defined in Section 3.4. I show the results of these checks in Table 3.10, for the outcomes of permanent employees in Panel A and fixed assets at net book value in Panel B. In column (1) of both panels, I repeat the headline result from the main specification in Table 3.7 and present the results with no baseline controls and no weighting of observations in column (2). Column (3) presents results from the inverse hyperbolic sine transformation instead of a logarithmic transformation, with baseline controls and inverse probability weights.

In column (4), I use a different definition of treatment but include weights and controls. Instead of focusing on the parts of the GQ project that were completed in 2002-2004, as has been the case so far, in this column firms in non-nodal cities within 5 km of *any* part of the GQ project are defined as the treatment group, and firms in non-nodal cities further away from any part of the GQ project are the control group. To the treatment group as defined before are therefore added firms that had already received access to improved highways at baseline in 2001 and those that were yet to receive access in 2004, which increases the sample size. Finally, column (5) presents the results of the instrumentation approach, where I instrument $nearGQ$ in equation 3.1 with a dummy variable indicating whether firms were located less than 250 km from the straight lines connecting the four nodal cities.

As Panel A of Table 3.10 shows, the negative effect on the number of permanent employees is highly robust. Compared to the main specification in column (1), the coefficient only varies by 0.05 log points across columns (2) to (5), and remains significant. The estimate from the instrumental strategy in column (5) also supports the large negative effect on labour inputs, though with a much larger negative effect of -0.58. Note that the F-statistic of 13.20 from the first-stage regression indicates a sufficiently, though not exceedingly, strong

Table 3.10: Robustness checks

Panel A: Permanent employees

	(1) FD	(2) FD	(3) FD	(4) FD	(5) FD IV
nearGQ	−0.26 ***	−0.22 *	−0.31 ***	−0.21 **	−0.58 ***
	(0.00)	(0.06)	(0.00)	(0.01)	(0.00)
	[0.03]	[0.12]	[0.02]	[0.06]	[0.10]
Baseline controls	Yes	No	Yes	Yes	Yes
Inverse probability weights	Yes	No	Yes	Yes	Yes
Transformation	Ln	Ln	IHS	Ln	Ln
First-stage F stat					13.20
Intervention	GQ 02-04	GQ 02-04	GQ 02-04	Entire GQ	GQ 02-04
Observations	604	604	632	703	604
Adjusted R-sq	0.34	0.01	0.65	0.33	0.33

Panel C: Fixed assets at net book value

	(1) FD	(2) FD	(3) FD	(4) FD	(5) FD IV
nearGQ	−0.29 ***	−0.21	−0.26 **	−0.42 ***	−0.15
	(0.01)	(0.38)	(0.02)	(0.00)	(0.43)
	[0.05]	[0.42]	[0.10]	[0.01]	[0.50]
Baseline controls	Yes	No	Yes	Yes	Yes
Inverse probability weights	Yes	No	Yes	Yes	Yes
Transformation	Ln	Ln	IHS	Ln	Ln
First-stage F stat					10.25
Intervention	GQ 02-04	GQ 02-04	GQ 02-04	Entire GQ	GQ 02-04
Observations	505	505	508	566	505
Adjusted R-sq	0.32	0.00	0.32	0.35	0.32

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes. Baseline controls are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. IHS = inverse hyperbolic sine transformation. * p<0.1, ** p<0.05, *** p<0.01

Table 3.11: Heterogeneity of main results with respect to baseline profits

	(1) ln(Sales)	(2) ln(Net sales)	(3) ln(Permanent employees)	(4) ln(Fixed assets)	(5) ln(Capital-labour ratio)	(6) ln(Output-labour ratio)
nearGQ · below median profits	−0.53 *	−0.73 **	−0.42 ***	−0.07	0.33	−0.13
	(0.10)	(0.04)	(0.01)	(0.79)	(0.13)	(0.60)
	[0.12]	[0.07]	[0.04]	[0.80]	[0.21]	[0.63]
nearGQ · above median profits	0.27	0.44	−0.09	−0.28	0.02	0.41
	(0.43)	(0.14)	(0.49)	(0.33)	(0.93)	(0.27)
	[0.69]	[0.41]	[0.60]	[0.47]	[0.94]	[0.46]
P-value(equal coefficients)	0.08	0.04	0.23	0.63	0.41	0.20
Observations	543	499	492	408	374	490
Adjusted R-sq	0.15	0.18	0.38	0.31	0.31	0.29

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes of the following specification: $\Delta y_{it} = \alpha_1 \cdot nearGQ02/04_i \cdot c_{it-1} + \alpha_2 \cdot nearGQ02/04_i \cdot (1 - c_{it-1}) \quad + \alpha_3 \cdot c_{it-1} + \alpha_4 \cdot (1 - c_{it-1}) + \boldsymbol{\delta}' \cdot \mathbf{X}_{it-1} + \Delta\epsilon_{it}$. c_{it-1} here is a dummy that equals 1 if the firm is below median net sales in 2001. Baseline controls included are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. Observations are weighted by inverse probability of non-attrition. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. The P-value(equal coefficients) tests $\alpha_1 = \alpha_2$, from wild cluster percentile-t bootstraps with 10000 repetitions. * p<0.1, ** p<0.05, *** p<0.01

instrument.

Panel B demonstrates that the negative impact on firms' fixed assets varies slightly more depending on the robustness check undertaken. Across columns (2) to (5), the coefficient varies between -0.15 and -0.42 and is insignificant in three out of four cases once I correct for the potential small number of clusters. Overall, these checks support a large negative and significant impact on firms' labour inputs. I continue to find negative impacts on firms' capital, though these are somewhat less robust.

3.7.4 Heterogeneity analysis

The previous results have already hinted at some heterogeneity in impacts, which is investigated further in this section. The large negative effects on employment and, partially, assets are primarily driven by firms that were less profitable at baseline. Pairwise correlations show that these are also the firms which were less productive, smaller and younger. I investigate this heterogeneity by dividing the sample into those firms with above and below median net sales in 2001, and interacting the resulting dummy variable with the

treatment variable, $nearGQ$, in the following specification:

$$\begin{aligned}\Delta y_{it} = \quad & \alpha_1 \cdot nearGQ_i \cdot c_{it-1} + \alpha_2 \cdot nearGQ_i \cdot (1 - c_{it-1}) \quad + \\ & \alpha_3 \cdot c_{it-1} + \alpha_4 \cdot (1 - c_{it-1}) + \boldsymbol{\delta}' \cdot \mathbf{X}_{it-1} + \Delta\epsilon_{it} \end{aligned} \tag{3.2}$$

c_{it-1} is the dummy that equals 1 if the firm has below median net sales in 2001, and the coefficients of interest are α_1 and α_2. α_1 captures the impact of treatment on firms with below median profits in 2001, as compared to control firms with below median profits. α_2 captures the treatment effect for firms with above median profits.

Table 3.11 presents the estimates for α_1 and α_2 for the main outcomes, and shows that the effects of the GQ project on these two groups of firms differ widely. Firms that were initially less profitable saw their sales, net sales and labour inputs significantly decline by close or more than half, with point estimates of -0.53, -0.73 and -0.42 log points, respectively. In comparison, the point estimates for the firms that were more profitable at baseline actually suggest that their sales, net sales and the output-labour ratio as a measure of labour productivity *increased*, with point estimates of 0.27, 0.44 and 0.41, respectively. Though none of the coefficients for the more profitable group are individually significant, I can reject that the effects across the two groups are the same for sales and net sales.

Interestingly, the initially more profitable group also decreased their input usage of permanent employees and fixed assets, though with insignificant point estimates of -0.09 and -0.28. While both groups therefore appear to have shrunk in terms of input usage, the group with higher baseline profits was able to sustain this downsizing while at the same time increasing their sales and net sales, and thus labour productivity.

Tables 3.A.2 to 3.A.4 in the appendix present additional heterogeneity results when splitting the sample by baseline median labour productivity, median employment, and being a young firm (below five years of age). While the heterogeneity of impacts using these characteristics is not as stark as the heterogeneity by baseline profitability presented here, the results support the observation that less profitable, less productive, smaller and younger firms at baseline were significantly negatively affected by the GQ project. In contrast, ini-

tially more established, successful firms decreased their input usage, but saw little impact or even positive impacts on their bottom line.

Viewed overall, these results for incumbent and surviving manufacturing firms indicate a significant reallocation and some productivity improvements as a result of the Golden Quadrilateral project. Previous evidence on the GQ project discussed in Section 3.3 showed increased entry of new firms in areas close to the improved highways, which could have increased competition. The incumbent manufacturing firms, which are the focus of this paper, decreased their use of labour and capital inputs as a result. However, only the initially less profitable firms lost sales and net sales, whereas I find indications that firms with higher profits at baseline saw increases in sales, net sales and labour productivity.

3.8 Conclusion

In this paper, I used a two-period firm-level panel to analyse the short-term impact of the Golden Quadrilateral project, a large-scale highway improvement programme, on incumbent and surviving Indian manufacturing firms. I find robust significant negative effects on the firms' usage of labour and capital inputs. These results are primarily driven by initially less successful firms, as measured by profits, size and profitability. For this group of firms, I observe large decreases in sales, profits, and labour inputs. These effects significantly differ from the impacts on more successful firms. For these firms, I find suggestive evidence that sales, profits and labour productivity increased, though the coefficients are not individually significant.

I argue that the observed patterns, combined with existing evidence, are consistent with a narrative in which the Golden Quadrilateral project increased market access along the improved highways, leading to higher firm entry and higher competition for existing firms. Out of the existing and surviving manufacturing firms that are analysed in this paper, only the initially more successful firms appear to have benefited from the improved market access. They produced more output and profits with fewer inputs. In contrast, firms that were initially less profitable and less productive, smaller and young, shrank significantly. The results in this paper therefore seem to present a snapshot of the process of reallocation

towards more successful firms, which was initiated by the increased market access that infrastructure improvements provide.

Acknowledgements I thank Christopher Woodruff and Simon Quinn for invaluable advice and support. The paper has benefited greatly from discussions with and comments by Sam Asher, Jose Asturias, Steve Bond, Stefano Caria, Elwyn Davies, Stefan Dercon, James Fenske, Doug Gollin, Justice Tei Mensah, Peter Neary, an anonymous examiner, and from audiences at the CSAE Conference 2016, the Oxford Gorman Workshop 2016, and the Oxford Firms and Development Research Group 2016.

Appendices

3.A Additional tables

Table 3.A.1: Other outcomes for firms with no missing values for main outcomes

Panel A

	(1) New hires	(2) Dismissals	(3) ln(Search time for production worker)	(4) ln(Search time for non-production worker)	(5) ln(Search time for manager)
nearGQ	−3.55	1.04***	−0.01	0.01	0.12
	(0.18)	(0.00)	(0.92)	(0.90)	(0.54)
	[0.38]	[0.01]	[0.95]	[0.91]	[0.62]
Observations	253	211	251	218	151
Adjusted R-sq	0.06	0.04	0.08	0.18	0.19

Panel B

	(1) ln(Search time for technician)	(2) Investment (dummy)	(3) Disinvestment (dummy)	(4) Exporter (dummy)	(5) Importer (dummy)
nearGQ	0.18*	0.11	−0.01	0.08	−0.03
	(0.06)	(0.50)	(0.90)	(0.15)	(0.44)
	[0.20]	[0.66]	[0.92]	[0.31]	[0.39]
Observations	215	265	227	339	334
Adjusted R-sq	0.23	0.08	0.18	0.04	0.07

Panel C

	(1) ln(Materials cost (Share in sales))	(2) ln(Employment cost (Share in sales))	(3) ln(Employment cost per employee)	(4) ln(Main material price)	(5) ln(Main product price)
nearGQ	0.02	−0.20	0.01	0.10	0.32
	(0.83)	(0.15)	(0.95)	(0.93)	(0.78)
	[0.86]	[0.42]	[0.94]	[0.96]	[0.78]
Observations	348	343	343	96	99
Adjusted R-sq	0.12	0.27	0.42	−0.11	−0.05

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes. The sample is restricted to firms for which the outcomes in Table 3.7 are not missing. Treatment is defined as being up to 5 km geographical distance away from the parts of the Golden Quadrilateral project defined at the bottom of each column, control as being more than 5 km geographical distance away from any part of the Golden Quadrilateral. Baseline controls included are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. Observations are weighted by inverse probability of non-attrition. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. NBV = net book value. * p<0.1, ** p<0.05, *** p<0.01

Table 3.A.2: Heterogeneity of main results with respect to baseline labour productivity

	(1) ln(Sales)	(2) ln(Net sales)	(3) ln(Permanent employees)	(4) ln(Fixed assets)	(5) ln(Capital-labour ratio)	(6) ln(Output-labour ratio)
nearGQ · below median labour productivity	−0.45	−0.47	−0.42 ***	−0.33	0.09	−0.04
	(0.11)	(0.23)	(0.00)	(0.17)	(0.69)	(0.86)
	[0.14]	[0.30]	[0.00]	[0.15]	[0.72]	[0.87]
nearGQ · above median labour productivity	0.09	0.24	−0.10	−0.01	0.14	0.19
	(0.80)	(0.41)	(0.36)	(0.96)	(0.52)	(0.50)
	[0.88]	[0.62]	[0.40]	[0.96]	[0.58]	[0.69]
P-value(equal coefficients)	0.21	0.18	0.04	0.45	0.92	0.58
Observations	602	450	604	459	458	601
Adjusted R-sq	0.19	0.17	0.40	0.33	0.37	0.32

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes of the following specification: $\Delta y_{it} = \alpha_1 \cdot nearGQ02/04_i \cdot c_{it-1} + \alpha_2 \cdot nearGQ02/04_i \cdot (1 - c_{it-1}) \quad + \alpha_3 \cdot c_{it-1} + \alpha_4 \cdot (1 - c_{it-1}) + \boldsymbol{\delta}' \cdot \mathbf{X}_{it-1} + \Delta\epsilon_{it}$. c_{it-1} here is a dummy that equals 1 if the firm is below median sales per permanent employee in 2001. Baseline controls included are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. Observations are weighted by inverse probability of non-attrition. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. The P-value(equal coefficients) tests $\alpha_1 = \alpha_2$, from wild cluster percentile-t bootstraps with 10000 repetitions. * p<0.1, ** p<0.05, *** p<0.01

Table 3.A.3: Heterogeneity of main results with respect to baseline employment

	(1) ln(Sales)	(2) ln(Net sales)	(3) ln(Permanent employees)	(4) ln(Fixed assets)	(5) ln(Capital-labour ratio)	(6) ln(Output-labour ratio)
nearGQ · below median employment	−0.41 *	−0.68 *	−0.56 ***	−0.39 *	0.19	0.19
	(0.10)	(0.08)	(0.00)	(0.05)	(0.31)	(0.31)
	[0.13]	[0.13]	[0.01]	[0.07]	[0.43]	[0.43]
nearGQ · above median employment	−0.01	0.27	−0.04	0.00	0.06	0.00
	(0.97)	(0.42)	(0.60)	(0.98)	(0.70)	(1.00)
	[0.98]	[0.67]	[0.65]	[0.99]	[0.75]	[1.00]
P-value(equal coefficients)	0.20	0.03	0.02	0.24	0.64	0.39
Observations	602	450	604	459	458	601
Adjusted R-sq	0.19	0.18	0.41	0.33	0.37	0.32

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes of the following specification: $\Delta y_{it} = \alpha_1 \cdot nearGQ02/04_i \cdot c_{it-1} + \alpha_2 \cdot nearGQ02/04_i \cdot (1 - c_{it-1}) \quad + \alpha_3 \cdot c_{it-1} + \alpha_4 \cdot (1 - c_{it-1}) + \boldsymbol{\delta}' \cdot \mathbf{X}_{it-1} + \Delta\epsilon_{it}$. c_{it-1} here is a dummy that equals 1 if the firm is below median employment in 2001. Baseline controls included are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. Observations are weighted by inverse probability of non-attrition. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. The P-value(equal coefficients) tests $\alpha_1 = \alpha_2$, from wild cluster percentile-t bootstraps with 10000 repetitions. * p<0.1, ** p<0.05, *** p<0.01

Table 3.A.4: Heterogeneity of main results with respect to baseline firm age

	(1) ln(Sales)	(2) ln(Net sales)	(3) ln(Permanent employees)	(4) ln(Fixed assets)	(5) ln(Capital-labour ratio)	(6) ln(Output-labour ratio)
nearGQ · below 4 years age	−0.70 *	−0.19	−0.22	−1.16 *	0.11	−0.13
	(0.08)	(0.63)	(0.23)	(0.08)	(0.78)	(0.76)
	[0.13]	[0.65]	[0.27]	[0.15]	[0.81]	[0.78]
nearGQ · above 4 years age	−0.13	−0.08	−0.26 ***	−0.18	0.10	0.10
	(0.62)	(0.76)	(0.00)	(0.17)	(0.50)	(0.65)
	[0.75]	[0.84]	[0.03]	[0.24]	[0.60]	[0.79]
P-value(equal coefficients)	0.30	0.78	0.85	0.27	0.99	0.66
Observations	671	499	604	505	458	601
Adjusted R-sq	0.20	0.15	0.39	0.33	0.37	0.32

Notes: The table presents results from difference-in-difference regressions of the impact of the Golden Quadrilateral project on firm outcomes of the following specification: $\Delta y_{it} = \alpha_1 \cdot nearGQ02/04_i \cdot c_{it-1} + \alpha_2 \cdot nearGQ02/04_i \cdot (1 - c_{it-1}) \quad + \alpha_3 \cdot c_{it-1} + \alpha_4 \cdot (1 - c_{it-1}) + \boldsymbol{\delta}' \cdot \mathbf{X}_{it-1} + \Delta\epsilon_{it}$. c_{it-1} here is a dummy that equals 1 if the firm is below four years of age in 2001. Baseline controls included are industry, state, type of ownership and legal status dummies, as well as firm age, distance from nearest nodal city, ln(Sales), ln(Permanent employees), ln(Fixed assets at NBV). To include control variables with missing values, I set missing values to zero and include dummies indicating the missing values. Observations are weighted by inverse probability of non-attrition. P-values clustered at the city level are given in parentheses, p-values from wild cluster percentile-t bootstraps with 10000 repetitions in squared brackets. All monetary values are in real terms, and all non-categorical variables are winsorised at 1 % in each tail. The P-value(equal coefficients) tests $\alpha_1 = \alpha_2$, from wild cluster percentile-t bootstraps with 10000 repetitions. * p<0.1, ** p<0.05, *** p<0.01

Concluding remarks

This thesis aimed to enhance our understanding of private sector development in South Asia, by investigating constraints related to gender, the importance of skills and selection of managers, and infrastructure. The contributions and conclusions discussed in each chapter open up new avenues for future research, of which I highlight a few here.

The findings of Chapter 1 suggest that relaxing constraints to female career advancement may be a promising lever to empower women in the home – both for promoted women and for the much larger number of women who could work under their supervision. Since the chapter presents the first evidence on this question, replication of the results in other contexts would give the policy implications additional credibility, particularly in light of the discussed recent evidence in Folke and Rickne (2020), which finds higher rates of divorce if women get promoted in Sweden. Additional data with larger sample sizes from such replications would also be helpful to structurally estimate the household's bargaining problem, and could thus be instructive to elucidate the mechanisms further.

Two conclusions from Chapter 2 warrant further investigation. The first is the importance of women's soft skills for fostering female career advancement. This suggests that developing these skills early in a woman's career might increase the pool of qualified female candidates for management positions. Together with Atonu Rabbani and Chris Woodruff, I am currently involved in a project in the Bangladeshi garment industry that intervenes to build these skills shortly after women enter the sector and that analyses the effects on attachment to the sector and career plans. In addition, Chapter 2 indicates that firms in their recruitment and promotion practices could benefit from learning which worker characteristics are predictive of later performance and how to measure these, especially for decisions with which firms have little prior experience. These characteristics may vary for

different positions or even for different types of applicants, which could be the focus of future work.

Chapter 3 provides evidence that increasing allocative efficiency as a result of relaxed infrastructure constraints can produce losers – in the form of shrinking or exiting firms, and workers who are displaced from their current jobs. This aligns with the literature on removals of international trade barriers, which finds that the benefits from expanded trade are accompanied by substantial adjustment costs and distributional consequences (Autor et al., 2016; Dell et al., 2019). Extending the analysis to the costs and distributional effects of removing internal trade barriers would be a valuable contribution.

As elaborated in the introduction, the three constraints addressed in this thesis are not easily captured in the standard theory of the firm, as is true for many other constraints. While various strands of literature have incorporated some of the interdependencies between firms, households, and governments into economic theory, we lack an overarching theory of firm development that is able to account for the internal and external constraints that businesses, especially in low- and middle-income countries, face. Developing such a theory in a parsimonious yet useful manner is a formidable challenge, but could be pivotal in identifying and addressing binding constraints to private sector development in low- and middle-income countries.

References

Abadie, A., Athey, S., Imbens, G., and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? *NBER Working Paper No. 24003*.

Abadie, A., Drukker, D., Herr, J. L., and Imbens, G. W. (2004). Implementing Matching Estimators for Average Treatment Effects in Stata. *The Stata Journal*, 4(3):290–311.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. (2019). Managerial Quality and Productivity Dynamics. *Working Paper*.

Almås, I., Armand, A., Attanasio, O., and Carneiro, P. (2018). Measuring and Changing Control: Women's Empowerment and Targeted Transfers. *The Economic Journal*, 128(612):F609–F639.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Anderson, S. and Eswaran, M. (2009). What Determines Female Autonomy? Evidence from Bangladesh. *Journal of Development Economics*, 90(2):179–191.

Antecol, H., Eren, O., and Ozbeklik, S. (2014). The Effect of Teacher Gender on Student Achievement in Primary School. *Journal of Labor Economics*, 33(1):63–89.

Asadullah, M. N. and Wahhaj, Z. (2016). Missing from the Market: Purdah Norm and Women's Paid Work Participation in Bangladesh. *IZA Discussion Paper No. 10463*.

Asgari, S. and Dasgupta, N. (2010). When Does Contact with Successful Ingroup Members Change Self-Stereotypes? A Longitudinal Study Comparing the Effect of Quantity vs. Quality of Contact with Successful Individuals. *Social Psychology*, 41(3):203–211.

Ashraf, N., Bandiera, O., Davenport, E., and Lee, S. S. (2020a). Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services. *American Economic Review*, 110(5):1355–1394.

Ashraf, N., Bau, N., Low, C., and McGinn, K. (2020b). Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment. *The Quarterly Journal of Economics*, 135(2):1095–1151.

Ashraf, N., Karlan, D., and Yin, W. (2010). Female Empowerment: Impact of a Commitment Savings Product in the Philippines. *World Development*, 38(3):333–344.

Asian Center for Development (2015). Garment Workers in Bangladesh. Social Impact of the Garment Industry. *Technical Report*.

Asturias, J., Garciá-Santana, M., and Ramos, R. (2019). Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India. *Journal of the European Economic Association*, 17(6):1881–1940.

Atkin, D. (2009). Working for the Future: Female Factory Work and Child Health in Mexico. *Working Paper*.

Autor, D. and Scarborough, D. (2008). Does Job Testing Harm Minority Workers? Evidence from Retail Establishments. *The Quarterly Journal of Economics*, 123(1):219–277.

Autor, D. H., Dorn, D., and Hanson, G. H. (2016). The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade. *Annual Review of Economics*, 8(1):205–240.

Avolio, B. J. and Bass, B. M. (2014). Multifactor Leadership Questionnaire (MLQ). *Statistic Solutions*.

Bandiera, O., Prat, A., Hansen, S., and Sadun, R. (2020). CEO Behavior and Firm Performance. *Journal of Political Economy*, 128(4):1325–1369.

Banerjee, A., Duflo, E., and Qian, N. (2020). On the Road: Access to Transportation Infrastructure and Economic Growth. *Journal of Development Economics*, 145(February):102442.

Banks, J., Blundell, R., and Lewbel, A. (1997). Quadratic Engel Curves and Consumer Demand. *The Review of Economics and Statistics*, 79(4):527–539.

Basu, K. (2006). Gender and Say: A Model of Household Behaviour with Endogenously Determined Balance of Power. *The Economic Journal*, 116(1980):558–580.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., and Topalova, P. (2009). Powerful Women: Does Exposure Reduce Bias? *The Quarterly Journal of Economics*, 124(4):1497–1540.

Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. *Science*, 335(6068):582–586.

Beaman, L., Keleher, N., and Magruder, J. (2018). Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi. *Journal of Labor Economics*, 36(1):121–157.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):0305–9049.

Belloni, A., Chernozhukov, V., and Hansen, C. (2013). Inference for High-dimensional Sparse Econometric Models. *Advances in Economics and Econometrics: 10th World Congress*, 3:245–195.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014a). High-dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28(2):29–50.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014b). Inference on Treatment Effects after Selection among High-dimensional Controls. *Review of Economic Studies*, 81:608–650.

Belloni, A., Chernozhukov, V., Hansen, C., and Kozbur, D. (2016). Inference in High-Dimensional Panel Models With an Application to Gun Control. *Journal of Business and Economic Statistics*, 34(4):590–605.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, 57(1):289–300.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive Linear Step-up Procedures that Control the False Discovery Rate. *Biometrika*, 93(3):491–507.

Benson, A., Li, D., and Shue, K. (2019). Promotions and the Peter Principle. *The Quarterly Journal of Economics*, 134(4):2085–2134.

Berlemann, M. and Wesselhöft, J.-E. (2014). Estimating Aggregate Capital Stocks Using the Perpetual Inventory Method. A Survey of Previous Implementations and New Empirical Evidence for 103 Countries. *Review of Economics*, 65(1):1–34.

Bertrand, M., Black, S. E., Jensen, S., and Lleras-Muney, A. (2019). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labor Market Outcomes in Norway. *Review of Economic Studies*, 86:191–239.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender Identity and Relative Income within Households. *The Quarterly Journal of Economics*, 130(2):571–614.

Bettinger, E. P. and Long, B. T. (2005). Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students. *American Economic Review: Papers & Proceedings*, 95(2):152–157.

Bigsten, A. and Söderborn, M. (2006). What Have We Learned from a Decade of Manufacturing Enterprise Surveys in Africa? *World Bank Research Observer*, 21(2):241–265.

Blattman, C. and Dercon, S. (2018). The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia. *American Economic Journal: Applied Economics*, 10(3):1–38.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does Managment Matter? Evidence from India. *The Quarterly Journal of Economics*, 128(1):1–51.

Bloom, N., Fischer, G., Rasul, I., Rodriguez-Clare, A., Suri, T., Udry, C., Verhoogen, E., Woodruff, C., and Zane, G. (2014a). IGC Evidence Paper: Firm Capabilities and Economic Growth. *Technical Report*.

Bloom, N., Lemos, R., Sadun, R., Scur, D., and Van Reenen, J. (2014b). JEEA-FBBVA Lecture 2013: The New Empirical Economics of Management. *Journal of the European Economic Association*, 12(4):835–876.

Boudreau, L. (2019). Multinational Enforcement of Labor Law: Experimental Evidence from Bangladesh's Apparel Sector. *Working Paper*.

Browning, M., Chiappori, P.-A., and Weiss, Y. (2014). *Economics of the Family*. Cambridge University Press, Cambridge.

Bruhn, M., Karlan, D., and Schoar, A. (2018). The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Rrial in Mexico. *Journal of Political Economy*, 126(2):635–687.

Cameron, A. C. and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372.

Cameron, C. A., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90(3):414–427.

Campos, F., Frese, M., Goldstein, M., Iacovone, L., Johnson, H. C., McKenzie, D., and Mensmann, M. (2017). Teaching Personal Initiative Beats Traditional Training in Boosting Small Business in West Africa. *Science*, 357(6357):1287–1290.

Cantril, H. (1965). *The Pattern of Human Concerns*. Rutgers University Press, New Brunswick, NJ.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *The Quarterly Journal of Economics*, 125(3):1101–1144.

Chandra, A. and Thompson, E. (2000). Does Public Infrastructure Affect Economic Activity? Evidence from the Rural Interstate Highway System. *Regional Science and Urban Economics*, 30(4):457–490.

Chong, A. and La Ferrara, E. (2009). Television and Divorce: Evidence from Brazilian Novelas. *Journal of the European Economic Association*, 7(2-3):458–468.

Dale-Olsen, H., Schøne, P., and Verner, M. (2013). Diversity among Norwegian Boards of Directors: Does a Quota for Women Improve Firm Performance? *Feminist Economics*, 19(4):110–135.

Datta, S. (2012). The Impact of Improved Highways on Indian Firms. *Journal of Development Economics*, 99(1):46–57.

Dee, T. S. (2005). A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review: Papers & Proceedings*, 95(2):158–165.

Deke, J. and Dragoset, L. (2012). Statistical Power for Regression Discontinuity Designs in Education: Empirical Estimates of Design Effects Relative to Randomized Controlled Trials. *Mathematica Policy Research Working Paper*.

Dell, M., Feigenberg, B., and Teshima, K. (2019). The Violent Consequences of Trade-Induced Worker Displacement in Mexico. *American Economic Review: Insights*, 1(1):43–58.

DellaVigna, S. and La Ferrara, E. (2015). Economic and Social Impacts of the Media. In Anderson, S. P., Waldfogel, J., and Strömberg, D., editors, *Handbook of Media Economics*, pages 723–768. Elsevier.

Deszö, C. and Ross, D. G. (2012). Does Female Representation in Top Management Improve Firm Performance? A Panel Data Investigation. *Strategic Management Journal*, 33:1072–1089.

Dharmalingam, A. and Morgan, S. P. (1996). Women's Work, Autonomy, and Birth Control: Evidence From Two South Indian Villages. *Population Studies*, 50:187–201.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5):899–934.

Doss, C. (2013). Intrahousehold Bargaining and Resource Allocation in Developing Countries. *World Bank Research Observer*, 28(1):52–78.

Duckworth, A. L. and Quinn, P. D. (2009). Development and Validation of the Short Grit Scale. *Journal of Personality Assessment*, 91(2):166–174.

Duflo, E. (2012). Women Empowerment and Economic Development. *Journal of Economic Literature*, 50(4):1051–1079.

Eswaran, M. and Malhotra, N. (2011). Domestic Violence and Women's Autonomy in Developing Countries: Theory and Evidence. *Canadian Journal of Economics*, 44(4):1222–1263.

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies*, 81(3):1046–1070.

Fafchamps, M. and Woodruff, C. (2017). Identifying Gazelles: Expert Panels vs. Surveys as a Means to Identify Firms with Rapid Growth Potential. *World Bank Economic Review*, 31(3):670–686.

Ferrari, A. and Dhingra, I. S. (2008). *India's Investment Climate. Voices of Indian Businesses*. The World Bank Group.

Flabbi, L., Macis, M., Moro, A., and Schivardi, F. (2019). Do Female Executives Make a Difference? The Impact of Female Leadership on Gender Gaps and Firm Performance. *The Economic Journal*, 129(2012):2390–2423.

Folke, O. and Rickne, J. (2020). All the Single Ladies: Job Promotions and the Durability of Marriage. *American Economic Journal: Applied Economics*, 12(1):260–287.

Friedemann-Sanchez, G. (2006). *Assembling Flowers, Cultivating Homes*. Lexington Books, Lanham, MD.

Getahun, T. D. and Villanger, E. (2018). Labour-Intensive Jobs for Women and Development: Intra-household Welfare Effects and Its Transmission Channels. *Journal of Development Studies*, 54(7):1232–1252.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2016). Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing. *The Economic Journal*, 126(591):317–357.

Ghani, E., Grover Goswami, A., and Kerr, W. R. (2013). The Golden Quadrilateral Highway Project and Urban/Rural Manufacturing in India. *World Bank Policy Research Working Paper 6620*.

Gibbons, R. and Henderson, R. (2013). What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises. In Gibbons, R. and Roberts, J., editors, *The Handbook of Organizational Economics*, pages 680–731. Princeton University Press.

Gibbons, S., Lyytikäinen, T., Overman, H. G., and Sanchis-guarner, R. (2019). New Road Infrastructure: The Effects on Firms. *Journal of Urban Economics*, 110:35–50.

Glaub, M. E., Frese, M., Fischer, S., and Hoppe, M. (2014). Increasing Personal Initiative in Small Business Managers or Owners Leads to Entrepreneurial Success: A Theory-based Controlled Randomized Field Intervention for Evidence-based Management. *Academy of Management Learning and Education*, 13(3):354–379.

Gollin, D. and Rogerson, R. (2014). Productivity, Transport Costs and Subsistence Agriculture. *Journal of Development Economics*, 107:38–48.

Hanna, R., Mullainathan, S., and Schwartzstein, J. (2014). Learning through Noticing: Theory and Evidence from a Field Experiment. *The Quarterly Journal of Economics*, 129(3):1311–1353.

Heath, R. (2014). Women's Access to Labor Market Opportunities, Control of Household Resources, and Domestic Violence: Evidence from Bangladesh. *World Development*, 57:32–46.

Heath, R. and Jayachandran, S. (2018). The Causes and Consequences of Increased Female Education and Labor Force Participation in Developing Countries. In Averett, S. L., Argys, L. M., and Hoffman, S. D., editors, *The Oxford Handbook of Women and the Economy*. Oxford University Press, New York, NY.

Heath, R. and Mobarak, A. M. (2015). Manufacturing Growth and the Lives of Bangladeshi Women. *Journal of Development Economics*, 115:1–15.

Heckman, J. J. and Kautz, T. (2012). Hard Evidence on Soft Skills. *Labour Economics*, 19(4):451–464.

Hoffman, M., Kahn, L. B., and Li, D. (2018). Discretion in Hiring. *The Quarterly Journal of Economics*, 133(2):765–800.

Holl, A. (2016). Highways and Productivity in Manufacturing Firms. *Journal of Urban Economics*, 93:131–151.

Hussam, R., Rigol, N., and Roth, B. (2020). Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design In The Field. *Harvard Business School Working Paper 20-082*.

Imbens, G. (2015). Matching Methods in Practice. Three Examples. *The Journal of Human Resources*, 50(2):373–419.

Imbens, G. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79(3):933–959.

Jedwab, R. and Moradi, A. (2016). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. *Review of Economics and Statistics*, 98(2):268–284.

Jensen, R. (2012). Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India. *The Quarterly Journal of Economics*, 127(2):753–792.

Jensen, R. and Oster, E. (2009). The Power of TV: Cable Television and Women's Status in India. *The Quarterly Journal of Economics*, 124(3):1057–1094.

Kabeer, N., Mahmud, S., and Tasneem, S. (2018). The Contested Relationship between Paid Work and Women's Empowerment: Empirical Analysis from Bangladesh. *European Journal of Development Research*, 30(2):235–251.

Khanna, G. (2016). Road Oft Taken: The Route to Spatial Development. *Working Paper*.

La Ferrara, E. (2016). Mass Media and Social Change: Can we Use Television to Fight Poverty? *Journal of the European Economic Association*, 14(August):791–827.

La Ferrara, E., Chong, A., and Duryea, S. (2012). Soap Operas and Fertility: Evidence from Brazil. *American Economic Journal: Applied Economics*, 4(4):1–31.

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., Rubiocodina, M., and Vakis, R. (2019). Challenges to Capture the Big Five Personality Traits in Non-WEIRD Populations. *Science Advances*, 5(7):eaaw5226.

Labowitz, S. and Baumann-Pauly, D. (2015). Bangladesh's Forgotten Apparel Workers. *Technical Report*.

Lazear, E. P., Shaw, K. L., and Stanton, C. T. (2015). The Value of Bosses. *Journal of Labor Economics*, 33(4):823–861.

Li, H. and Li, Z. (2013). Road Investments and Inventory Reduction: Firm Level Evidence from China. *Journal of Urban Economics*, 76:43–52.

Luke, N. and Munshi, K. (2011). Women as Agents of Change: Female Income and Mobility in India. *Journal of Development Economics*, 94(1):1–17.

Macchiavello, R., Menzel, A., Rabbani, A., and Woodruff, C. (2020). Challenges of Change: An Experiment Promoting Women to Managerial Roles in the Bangladeshi Garment Sector. *Working Paper*.

Majlesi, K. (2016). Labor Market Opportunities and Women's Decision Making Power within Households. *Journal of Development Economics*, 119:34–47.

Marinova, J., Plantenga, J., and Remery, C. (2016). Gender Diversity and Firm Performance: Evidence from Dutch and Danish Boardrooms. *The International Journal of Human Resource Management*, 27(15):1777–1790.

McKelway, M. (2019). Experimental Evidence on the Effects of Women's Employment. *Working Paper*.

McKenzie, D. and Sansone, D. (2017). Man vs. Machine in Predicting Successful Entrepreneurs: Evidence from a Business Plan Competition in Nigeria. *World Bank Policy Research Working Paper 8271*.

McKenzie, D. and Woodruff, C. (2014). What are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World? *World Bank Research Observer*, 29(1):48–82.

McKenzie, D. and Woodruff, C. (2017). Business Practices in Small Firms in Developing Countries. *Management Science*, 63(9):2967–2981.

Melo, P., Graham, D., and Noland, R. (2010). Impact of transport infrastructure on firm formation. *Transportation Research Record*, 2163:133–143.

Menzel, A. and Woodruff, C. (2019). Gender Wage Gaps and Worker Mobility: Evidence from the Garment Sector in Bangladesh. *NBER Working Paper No. 25982*.

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *Review of Economics and Statistics*, 90(4):683–701.

Mikolajczak, M., Brasseur, S., and Fantini-Hauwel, C. (2014). Measuring Intrapersonal and Enterpersonal EQ: The Short Profile of Emotional Competence (S-PEC). *Personality and Individual Differences*, 65:42–46.

Miller, A. R. (2018). Women and Leadership. In Averett, S. L., Argys, L. M., and Hoffman, S. D., editors, *The Oxford Handbook of Women and the Economy*. Oxford University Press, New York, NY.

Mohan, R. (2004). Infrastructure Development in India: Emerging Challenges. In Bourguignon, F. and Pleskovic, B., editors, *Annual World Bank Conference on Development Economics, 2004: Accelerating Development*, pages 35–71. The World Bank Group.

Nanda, R. (2016). Financing High-potential Entrepreneurship. *IZA World of Labor*, 252:1–10.

Nataraj, S. (2010). *Productivity and Firm Size Distribution: Evidence from India's Organized and Unorganized Manufacturing Sectors*. PhD thesis, University of California, Berkeley.

National Highways Authority of India (1999). *Annual Report 1998-1999*. National Highways Authority of India.

National Highways Authority of India (2002). *Annual Report 2001-2002*. National Highways Authority of India.

National Highways Authority of India (2005). *Annual Report 2004-2005*. National Highways Authority of India.

O'Brien, P. (1984). Procedures for Comparing Samples with Multiple Endpoints. *Biometrics*, 40(4):1079–1087.

Pande, R. and Ford, D. (2011). Gender Quotas and Female Leadership. *World Development Report 2012 Background Paper*.

Pitt, M. M., Khandker, S. R., and Cartwright, S. R. (2006). Empowering Women with Micro Finance: Evidence from Bangladesh. *Economic Development and Cultutal Change*, 54(4):791–831.

Qian, N. (2008). Missing Women and the Price of Tea in China: The Effect of Sex-specific Earnings on Sex Imbalance. *The Quarterly Journal of Economics*, 123(3):1251–1285.

Quinn, S. and Woodruff, C. (2019). Experiments and Entrepreneurship in Developing Countries. *Annual Review of Economics*, 11(1):225–248.

Rahman, L. and Rao, V. (2004). The Determinants of Gender Equity in India: Examining Dyson and Moore's Thesis with New Data. *Population and Development Review*, 30(2):239–268.

Ramos, A. (2018). Household Decision Making with Violence: Implications for Conditional Cash Transfer Programs. *Working Paper*.

Redding, J. and Turner, M. A. (2015). Transportation Costs and the Spatial Organization of Economic Activity. In Duranton, G., Henderson, J. V., and Strainge, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5, chapter 20, pages 1339–1398. Elsevier B.V.

Rotter, J. B. (1966). Generalized Expectancies for Internal versus External Control of Reinforcement. *Psychological Monographs: General and Applied*, 80(1):1–27.

Sabatier, M. (2015). A Women's Boom in the Boardroom: Effects on Performance? *Applied Economics*, 47(26):2717–2727.

Schochet, P. Z. (2008). Technical Methods Report: Statistical Power for Regression Discontinuity Designs in Educational Evaluations. *NCEE Working Paper 2008-4026*.

Scholz, U., Gutiérrez Doña, B., Sud, S., and Schwarzer, R. (2002). Is General Self-Efficacy a Universal Construct? Psychometric Findings from 25 Countries. *European Journal of Psychological Assessment*, 18(3):242–251.

Schwarzer, R. and Jerusalem, M. (1995). Generalized Self-efficacy Scale. In Weinman, J., Wright, S., and Johnston, M., editors, *Measures in Health Psychology: A User's Portfolio. Causal and Control Belief*, pages 35–37. NFER-NELSON, Windsor, England.

Spitzer, R. L., Kroenke, K., Williams, J. B., and Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, 166(10):1092–1097.

The World Bank (1995). *India. Transport Sector. Long Term Issues. Report No. 13192-IN*. The World Bank Group.

The World Bank (2002). *India's Transport Sector: The Challenges Ahead. Volume 1: Main Report*. The World Bank Group.

The World Bank (2004). *India. Investment Climate Assessment 2004. Improving Manufacturing Competitiveness*. The World Bank Group.

The World Bank (2007). *A Decade of Action in Transport. An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005*. The World Bank Group.

The World Bank (2008). Whispers to Voices. Gender and Social Transformation in Bangladesh. *Bangladesh Development Series Paper No. 22*.

The World Bank (2012). *World Development Report 2012. Gender Equality and Development*. The World Bank Group.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

World Health Organization (2001). Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence against Women. *Technical Report*.

Yang, Y. (2018). Transport Infrastructure, City Productivity Growth and Sectoral Reallocation: Evidence From China. *IMF Working Paper WP/18/276*.