DOES ASSIGNING MORE WOMEN TO MANAGERIAL POSITIONS ENHANCE FIRM PRODUCTIVITY? EVIDENCE FROM SWEDEN

Yoshihiro Sato
Michihito Ando
EIJS, Stockholm School of Economics

Postal address: P.O. Box 6501, S-113 83 Stockholm, Sweden.
Office address: Saltmätargatan 19C Telephone: +46 8 736 93 60
E-mail: japan@hhs.se Internet: http://www.hhs.se/eijs

Preface

Gender inequality in the labor market has become a serious concern in Japan. Although the last few decades have seen several pieces of legislation passed in this policy area and there have been some improvements in the situations of female workers in general and working mothers in particular, there are still many obstacles to gender equality in the workplace. Japan was recently ranked 111th out of the 144 countries assessed in the World Economic Forum's Global Gender Gap Report 2016.

In April 2013, Prime Minister Abe vowed that women would be an integral part of his economic expansion plan ("Abenomics") and set goals of having women hold 30% of leadership positions in corporations by 2020 and increasing female labor participation by 5%.

As the discussion of gender equality has intensified within academic as well as policy debates in Japan, the question of the relationship between gender equality and firm performance has attracted attention. Does gender equality in the workplace contribute to higher productivity?

We have examined this question using data from Swedish companies in the following study.

Does Assigning More Women to Managerial Positions Enhance Firm Productivity? Evidence from Sweden[*]

Yoshihiro Sato[†]   Michihito Ando[‡]

January 9, 2017

Abstract

We analyze whether gender composition at non-board managerial levels has any impact on firm productivity and other related outcomes in the service sector using a linked employer-employee dataset from Sweden. Exploiting within-firm variation, we apply a difference-in-differences propensity score matching method to address an endogeneity issue. Our results suggest no significant effects on productivity but significant positive effects on firms' growth in terms of value added and labor inputs when a firm "replaces" a male manager with a woman. We do not observe any impact when a firm "appoints" a woman instead of a man to a new managerial position.

[†]Corresponding author: European Institute of Japanese Studies, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm, Sweden, +46-8-736-9323, yoshihiro.sato@hhs.se

[‡]National Institute of Population and Social Security Research, Hibiya Kokusai Building 6th Floor, 2-2-3 Uchisaiwaicho, Chiyoda-ku, Tokyo, 100-0011, Japan, +81-3-3595-2984, andou-michihito@ipss.go.jp

1

1 Introduction

Gender inequality and gender discrimination in labor markets are persistent problems in both developed and developing countries. They are also well-studied topics in economics, and various studies have revealed diverse aspects of these issues (Altonji and Blank 1999; Olivetti and Petrongolo 2016). It has been shown, for example, that, although female labor participation rates increase in developed countries, women suffer from lower wages, lower managerial status, and less representation on boards of directors even in the Nordic countries, whose gender-equality indices are among the highest in the world.

What remains relatively unclear, however, is whether employing more women or assigning them to higher positions such as managers and directors results in higher or lower firm productivity. This is an important question, because if it results in lower productivity then at least some part of gender inequality in labor markets, such as wage or income disparity, can be explained by this productivity difference. If this is not the case, however, then persistent gender inequality in management-level positions must be explained by some other socio-economic factors.

Several recent studies examine whether the share of top female directors on boards or their interaction with female workers has any impact on firm productivity or other firm characteristics. The meta analysis of this topic in Pletzer et al. (2015) demonstrates that the share of female directors does not have a significant impact on firms' productivity. Ahern and Dittmar (2012) investigates new legislation that requires publicly traded Norwegian firms to have boards made up of least 40 percent female directors and find that this constraint leads to a significant decline in Tobin's Q over the following years, whereas a more recent study, Eckbo et al. (2016), which also examines the same reform in Norway, finds no such effect. Another related area of research is experimental studies on the effect of gender composition on labor markets and other related outcomes. For a discussion of relavent publications see the literature review in Azmat and Petrongolo (2014).

Nonetheless, it is still largely unknown whether gender composition at non-board levels affects firms' productivity and related outcomes and whether this effect differs from the effect of board-level gender composition

and gender effects obtained from experimental evidence. Because most daily business activities are conducted at non-board levels and the number of non-board workers is considerably larger than that of directors, it is important to examine how gender composition at non-board levels affects firm productivity.

To address the lack of studies in this field, we investigate the effect of gender composition at the non-board managerial level on firm productivity and other related outcomes in service sectors, using a Swedish matched employer–employee dataset. The advantage of our register-based dataset is that it covers a large part of Swedish firms and employees and it is also a form of panel data with which we can exploit not only between-firm variation but also within-firm variation. Another advantage of this dataset is that it contains information on each employee's occupational role, including their manager or non-manager status. This enables us to identify gender composition of managers at the firm level.

The endogenous determination of the share of non-board female managers and firm productivity is of great interest in determining the direction of causation. Unlike the case of board-level female ratio investigated in Ahern and Dittmar (2012), it is hard to find a reform that causes an exogenous variation in gender composition of managers. We therefore adopt a selection-on-observable approach to reveal the causal link between the gender ratio of managers and firms' productivity and related outcomes. At the same time, because we can utilize a rich panel dataset with matched employer-employee information, we can effectively control for unobserved time-invariant fixed effects by focusing on the changes in, rather than the levels of, gender composition and firms' outcomes.

More concretely, our identification strategy is based on a difference-in-differences (DID) matching method in which we compare outcomes between the firms that increased the number of female managers during a certain period (treated) and the firms that are similar to the treated firms but did not change their gender composition (control). Because it is likely that firms' pre-treatment characteristics and/or expected future business performance determine their decisions to change (or not change) the gender composition of managers, we need to control for possible confounding factors by utilizing a matching method.

3

Although our DID strategy eliminates any omitted variables bias that arises from the confounding effects of unobserved time-invariant factors on the *level* of an outcome variable, we cannot control for the effect of unobserved factors (either time-invariant or time-varying) on the *change* of the outcome. We then use placebo tests to indirectly test our identifying assumption.

Our estimation results suggest that there is no statistically significant effect of the gender composition of managers on firm productivity. This implies that either the assignment of managerial positions is realized basically on the basis of merit-based criteria and gender itself does not affect firm productivity, or that any positive or negative gender-related effect that exists is canceled out by some other factors.

On the other hand, we observe that a change in gender composition of managers has an effect on the size of firms' inputs and outputs. This effect also differs depending on the firm's situation when the number of female managers increases. First, we find a positive impact of the increase on firm growth in terms of labor and output when an existing manager position is reassigned from a man to women (i.e. a positive "replacement effect"). Second, we observe no such effect when a new managerial position is assigned to a woman rather than a man. Based on previous studies, we explore potential mechanisms behind these effects.

The rest of the paper is organized as follows. In Section 2 we explore possible mechanisms of gender effects found in previous studies and explain gender inequality in the Swedish labor market. Section 3 describes the dataset used for our empirical analysis. In Section 4 we discuss our empirical strategies and related empirical issues. Section 5 presents estimation results, including placebo trials and robustness checks. Section 6 concludes.

2 Background

2.1 Female managers and firm productivity

Previous studies suggest that a change in gender composition of the top executives or employees may affect firm productivity through various channels

(for reviews, see Post and Byron (2015) and Pletzer et al. (2015)). First, Jackson et al. (1995) and Robinson and Dechant (1997) argue that a diversified workforce in terms of education, age, ethnicity, and gender may provide greater width of information sources, skill sets, problem-solving abilities and socio-cognitive perspectives. For instance, male and female employees have different experiences in their daily lives because of their genders, and this difference may provide valuable insight into the different needs and preferences of male and female customers (Nkomo and Cox 1999). Hambrick et al. (1996) suggests that a mixture of different perspectives will stimulate discussion of current strategies and provoke a search for potential alternatives. In addition, labor diversity can also enhance innovation and creativity and lead to more knowledge spillover (Osborne 2000; Hong and Page 2001; Rauch and Casella 2003; Hong and Page 2004; Alesina and Ferrara 2005; Berliant and Fujita 2008). Lazear (1999) points out that positive effects may prevail as long as organization members' areas of competence complement each other.

Imagining an opposite case may help us understand the potential benefits of labor diversity. For example, a homogeneous group may often be keen on holding fast with shared world views and common decision-making routines, which is not always effective in finding creative solutions, especially in a fast-changing and complex environment. Under such circumstances, recruiting new members to the group may force the incumbents to explicitly articulate their ideas, to make the decision-making process more transparent, and to open up critical discussions (Farrell and Hersch 2005; Carter et al. 2010).

Nevertheless, diversity can also potentially undermine the benefits that homogeneity generates, such as team coherence and behavioral integration (Carpenter 2002; Bassett-Jones 2005). Poor communication among co-workers who do not share the same values can induce misunderstanding, conflict and uncooperative behavior. These negative effects become substantial if workers are prejudiced and the resulting conflicts incur costs on the workplace (Becker 1971). Koenig et al. (2011) argue that women are especially subject to discrimination-related conflicts among male-dominated colleagues and might not be able to realize their full potential, a state of affairs that reduces firm productivity.

5

Second, promoting more female managers may contribute to firm productivity through gender-specific cognitive abilities. The literature on gender variation of psychological traits points to systematic differences between men and women in risk preference, attitudes towards competition and social preferences (for a review of this research, see Bertrand (2011)). Several experimental studies report that women in general are more risk averse (Eckel and Grossman 2008b; Croson and Gneezy 2009; Dohmen and Falk 2011; Dohmen et al. 2011). Datta Gupta et al. (2005) and Niederle and Vesterlund (2007) find that women have less taste for competition. It has also been argued that women are more socially minded than men and exhibit stronger redistributive and altruistic preferences (Eckel and Grossman 2008a; Alesina and Giuliano 2009; Croson and Gneezy 2009; Funk and Gathmann 2015). Borghans et al. (2006) point out that women in general have higher interpersonal skills.

It is not clear, however, whether these cognitive differences affect firm productivity positively or negatively. If risk preferences are related to risk premiums for investment returns, risk-averse decisions and a lower preference for competition may lead to worse financial performance in the long run. At the same time, these traits can induce caution in taking action and lead to well-founded decisions, which is not necessarily negative for firm productivity. When it comes to social preferences, women's altruistic values are often associated with less financially attractive outcomes for themselves (Fortin 2008). On the other hand, managers that prioritize other workers' well-being are more likely to create worker-friendly workplaces and enhance worker cohesion. In addition, interpersonal skills, which women are more likely than men to possess, have become more valuable for firms in the context of current technological and organizational changes (Borghans et al. 2006).

The third channel in the link between female managers and firm productivity is a recruitment pool with higher human capital. Underrepresentation of discriminated groups in managerial positions implies a waste of resources for the society in question if the group is, on average, as well educated as other groups. According to Becker (1971)'s classic model of employer discrimination, firms that employ both genders equally obtain the advantage of acquiring access to a skilled labor force with lower, dis-

6

criminatory wages, while discriminatory firms will fail to survive in the long run under a competitive labor market. These argument on the basis of discriminatory wages is unlikely to apply in Sweden, where the gender wage gap is fairly small among workers with similar roles (See Section 2.2). But non-discriminatory firms may still obtain the advantage of getting access to a recruitment pool of educated workers without competition from discriminatory firms.

The last channel is a signaling effect. Increasing the number of female managers is often driven by gender-focused human resource policies[1] in the firm. Signaling theory (Spence 1973; 2002) predicts that, under information asymmetry between firms and job applicants, observable actions from the firm side, such as employing more female managers, are perceived as a sign that the firms are actively promoting gender diversity and supporting their employees. If successful, the signal draws the attention of job applicants from both genders and makes the competition for positions at the firm in question stronger, thereby attracting more competent workers not only to managerial positions but also to other positions within the firm (Ali 2016). The signaling effect can thus raise the level of human capital in the whole organization, which in turn may positively contribute to firm productivity.

The potential gender effects described above are concerned with "soft" factors, such as managerial skills, human resource management, leadership, inter-organizational communication and market exploitation. We assume these factors are more crucial and prevalent in service industries than in manufacturing industries. In contrast, productivity of manufacturing firms is more likely to be affected by "hard" factors, such as the technological innovation of products and effectiveness of production processes. In these cases, potential gender effects might be diluted and more difficult to observe. We therefore focus on the service sector in the current study.

2.2 Female Managers in Sweden

The Swedish labor market is often regarded as one of the most gender-equal among the OECD countries. The labor force participation rate for

---

[1]This refers to such policies that provide active support to achieve a gender-diverse workforce rather than simply ensuring equal opportunities (Shen et al. 2009; Ali 2016)

Figure 1: Share of female managers from 2001 to 2012
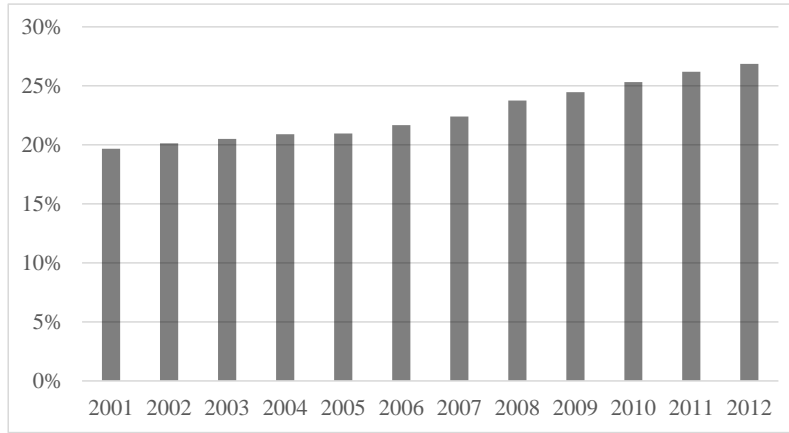
Notes: The share of female managers in both the private and public sectors. These figures are based on the longitudinal integrated database for labor market research, provided by Statistics Sweden. For details, see Section 3.

women between 15 and 64 years old in 2011 was 77.7 percent, 5.0 percentage points lower than the male participation rate. In contrast, the OECD average female labor participation rate is 65.2 percent and the average gender difference is 13.7 percentage points (OECD 2012). While the average unadjusted wage gap in Sweden between men and women was 14 percent in 2012, the adjusted figure after controlling for age, education, working hours, sector and occupation was 7 percent (Statistics Sweden 2014).

Nevertheless, the gender gaps in Sweden in terms of the female share of managers and board members are still large. Women occupied 29 percent of all managerial positions in the private sector (36 percent of all managerial positions including those in the public sector) in 2012 (Statistics Sweden 2014). When it comes to board members in listed companies, women made up 24 percent in 2013. Although this figure is high by international standards[2], these substantial gaps in managerial and higher positions may partially be attributed to a "glass ceiling" – a phenomenon in which women perform well in the labor market up to a certain point after which there is

---

[2]OECD (2012) makes an international comparison among the OECD countries using data from 2009 and 2010. According to their database, the female share of managers in Sweden in 2010 was 31.0 percent, which is close to the OECD average. On the other hand, the female share of listed companies' board members in Sweden was 19.3 percent in 2009, which is a much higher figure than the OECD average of 10.3 percent. The only country that had a larger ratio in the comparison was Norway, which introduced a legislated quota of 40 percent in 2006.

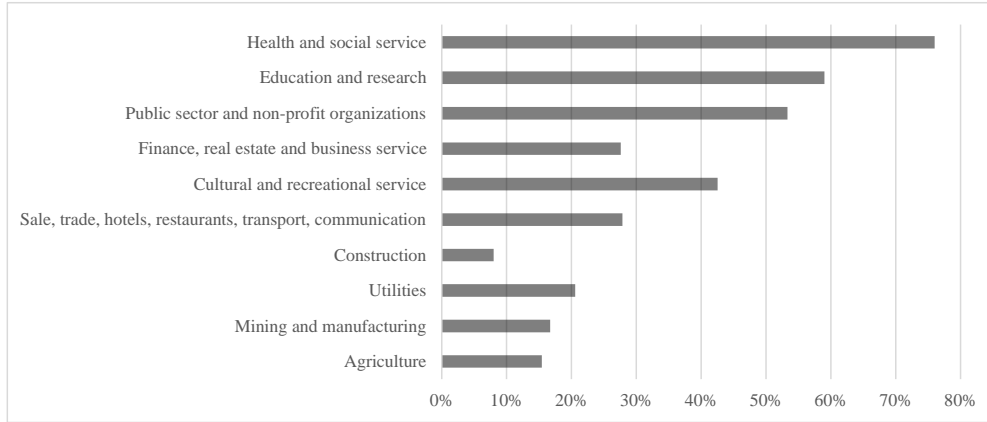Figure 2: Share of female managers by sector (2010)

Notes: The share of female managers in both the private and public sectors. These figures are based on the longitudinal integrated database for labor market research, provided by Statistics Sweden. For details, see Section 3.

an effective limit on their prospects (Albrecht et al. 2003; 2015; Wahlberg 2010).[3]

Figure 1 reports the share of female managers in Sweden in the last decade (For the source, see Section 3). Here we observe slow but steady progress in terms of the share of female managers. One of five managers was a woman in 2001. This share gradually increased and reached 28 percent in 2012. The penetration of women in managerial positions, however, varies largely across industries/sectors, a state of affairs that is presented in Table 2. Agriculture and traditionally male-dominated industries, such as mining, manufacturing and construction, have the lowest share of female managers. In contrast, more than half of managerial positions are occupied by women in service sectors with female-dominated workplaces, such as healthcare, social service and education organizations, as well as in the public sector.

---

[3]Albrecht et al. (2003), Albrecht et al. (2015) and Wahlberg (2010) find that the average gender wage gap in Sweden is mainly attributable to a wage gap in the upper tail of the wage distribution, and that gender differences in rewards based on workers' human capital characteristics (e.g. age, education, occupation, sector), not differences in these human capital characteristics themselves, is the primary factor responsible for it. They interpret their finding as strong evidence of a glass ceiling.

3 Data

Our study exploits two large register-based databases of annual observations provided by Statistics Sweden. The first is the longitudinal integrated database for labor market research that covers all individuals over 16 years old in Sweden. The variables we extract from the database include the birth year, gender and the highest education level attained for each individual, as well as the anonymized ID of the firm where the individual worked in November of each year and his/her occupational status at the firm. The occupational code tells us whether the individual is a manager or not.

The other database includes structural business statistics that cover all firms active in Sweden. We extract variables such as firm ID, three-digit industry classification code, sales, value added, intermediate input costs, and capital stocks. The firm IDs in this database are the same as those used in the first database on individuals, which allows us to construct a linked employer-employee dataset. We can thereby compute the numbers of employees and managers as well as their gender, age and level of education for each firm. We limit our analysis to service industries in the private sector.

The occupational code provided for individuals in the database is a core of this empirical study since it defines managers and other employees. This code is based on the Swedish Standard Classification of Occupations 1996 (SSYK-96), which is a national adaptation of the International Standard Classification of Occupations (ISCO-88) published by the International Labor Office. The occupations in SSYK-96 are organized in a hierarchical framework at four-digit level according to the skill level and skill specialization a position requires for an individual to conduct the task (Statistics Sweden 1998).

Based on the three-digit codes, we define employees classified as (121) directors and chief executives, (122) production and operations managers, (123) other specialist managers, and (131) managers of small enterprises as "managers" in our study.[4]

Information on the occupational status of each individual is first col-

---

[4]Groups 121-123 consist of middle managers. Group 121 also includes chief executives who are chairs of the boards as well, but excludes board members. The last group, 131, is made up of both executives and middle managers in small firms.

lected from the official statistics on wage and salary structures (Statistics Sweden 2011b), based on Statistics Sweden's own annual survey or databases held by employer organizations.[5] It covers all firms with more than 500 employees (around 11,000 firms), as well as a selection of smaller firms covered by employer organizations' databases.

Occupational information on employees in the other firms are collected separately through questionnaire surveys by Statistics Sweden. The sample size is around 47,000 annually using a rolling scheme. The number of firms from which occupational information is collected in this way is around 150,000, the majority of which are firms with 1 to 19 employees. According to the data documentation, these firms are surveyed every fourth or fifth year (Statistics Sweden 2011b). When a firm is not among the surveyed firms for a certain year, the occupational data from the last survey are filled in for the firm's employees for that year. This causes a time-lag before a change in the occupational status is reflected in the dataset. We will come back to this issue later.

4 Identification strategy

4.1 Identification using DID matching

A primary objective of our study is to identify the causal effect of the gender ratio of managers on firm productivity. If we discretize managerial gender composition, we can define the population average treatment on the treated (PATT) using the potential outcome framework or Rubin Causal Model (Holland 1986) as follows:

$$\tau^{PATT1} \equiv E[Y_i(1) - Y_i(0)|D_i = 1], \tag{1}$$

where $Y_i(D_i)$ is the potential firm productivity of the firm i and D_i is equal to one if a change of managerial gender composition exceeds some threshold and otherwise zero. The fundamental problem of identifying τ^{PATT1} is that we cannot observe both $Y_i(1)$ and $Y_i(0)$ at the same time, and therefore

---

[5]These organizations collect information on the salary and occupation of each employee of their member firms for use in collective bargaining with trade unions.

cannot directly calculate τ^{PATT1} from equation (1).

As is well known, the most powerful way to identify and estimate τ^{PATT1} is to implement a randomized control trial (RCT) in which D_i is randomly assigned to firms and to calculate the difference of averaged observed outcomes between the treated and control groups. If RCT cannot be implemented, researchers often try to find some quasi-experimental variation that can be regarded as an imperfect but nonetheless illuminating experiment.

In our case, however, to exploit experimental or even quasi-experimental variation to estimate τ^{PATT1} is quite difficult. To our knowledge, there is no experiment or quasi-experiment that effectively randomizes the share of female managers in actual firms in Sweden during the studied period. To mitigate this drawback, we try to construct a quasi-experimental setting to the greatest extent possible by utilizing the panel structure of our dataset, which we will explain below, and also by relying on some version of the "selection on observable" empirical strategy.

To accomplish this task, we first define our population average treatment effect on the treated conditional on covariates as follows:

$$\tau_t(x) \quad \equiv \quad E[Y_{i,t}(1) - Y_{i,t}(0)|D_{i,t} = 1, X_i = x]. \tag{2}$$

In this equation, $\tau_t(x)$ is the average treatment effect on the treated for $X_i = x$ at time t, where X_i is the vector of observed pre-determined covariates. $Y_{i,t}(D_{i,t})$ is the potential firm productivity of the firm i at time t. $D_{i,t}$ is defined to be equal to one if the gender composition of managers at t relative to some reference time exceeds a certain threshold, and otherwise zero. Our primary objective is then to estimate the following PATT:

$$\tau_t^{PATT2} \quad \equiv \quad E[\tau_t(X_i)]. \tag{3}$$

Our basic identifying assumption for the estimation of τ_t^{PATT2} is that conditional on observable covariates, changes in the share of female managers is independent of changes in firm productivity. This identifying assumption is essentially the same as the common trend assumption in a standard difference-in-differences (DID) method, but conditional on pre-

determined covariates. This assumption can be expressed as follows:

$$E[Y_{i,t}(0) - Y_{i,0}(0)|D_{i,t} = 1, X_i = x] = E[Y_{i,t}(0) - Y_{i,0}(0)|D_{i,t} = 0, X_i = x], \tag{4}$$

where we set $t = 0$ as the reference time, that is, a year just before the treatment is introduced. Using the assumption (4), $\tau_t(x)$ and τ_t^{PATT2} can be transformed into

$$\tau_t(x) = E[Y_{i,t}^{obs} - Y_{i,0}^{obs}|D_{i,t} = 1, X_i = x] - E[Y_{i,t}^{obs} - Y_{i,0}^{obs}|D_{i,t} = 0, X_i = x] \tag{5}$$

$$\equiv \tau_t^{DID},$$

and

$$\tau_t^{PATT2} = E[\tau_t^{DID}(X_i)], \tag{6}$$

where Y^{obs} indicates that Y is observed.[6] We then estimate average treatment effects on the treated (ATET) from the sample analog of equation (6). This estimator can be interpreted as a DID-matching estimator if we condition on X_i using a matching method.

By further exploiting the panel structure of our data, we also consider another definition of PATT based on a different matching strategy in which we add pre-determined lagged outcomes to the covariates X_i. This is akin to what Imbens and Wooldridge (2009) call the "uncounfoundedness-based" approach to DID methods, where the differences in outcomes between the treated and control groups are compared conditional on both pre-treatment lagged outcomes and other pre-determined covariates. Although the difference between these two approaches (standard DID approach and uncounfoundedness-based approach) appears rather subtle, they actually rely on different identifying assumptions (Imbens and Wooldridge 2009).

To judge which assumption is more plausible than the other requires substantive knowledge about the actual data generating process. Imbens and Wooldridge (2009) argue that the "uncounfoundedness-based" approach

---

[6]Equation (4) can be transformed into $E[Y_{i,t}(0)|D_i = 1, X_i = x] = E[Y_{i,t}(0) - Y_{i,0}(0)|D_i = 0, X_i = x] - E[Y_{i,0}(0)|D_i = 1, X_i = x]$ and then this equation is inserted into equation (2), resulting in equation (5).

is more attractive in a practical sense because it is unlikely that a treatment-control comparison conditional on lagged pre-treatment outcomes will make its causal interpretation less credible. On the other hand, Chabé-Ferret (2014) and Chabé-Ferret (2015) provide more critical views about the use of pre-treatment outcomes as additional covariates for matching and DID matching. In these studies, Chabé-Ferret investigates the performance of DID and matching when pre-treatment outcomes can be used as additional covariates for matching. Although his investigations with observational data and Monte Carlo simulations primarily focus on job training programs and relevant data generating processes, he concludes that matching estimation conditioning on pre-treatment outcomes sometimes produces biases.

In our estimation, we do not have any convincing a priori reasoning concerning whether conditioning on pre-treatment outcomes is plausible or bias-inducing. Based on several investigations we conducted using placebo tests, our preferred choice of covariates for matching is the set of pre-treatment outcomes and observed socio-economic characteristics of the firms in question[7]. Our empirical strategy is discussed more thoroughly in the following subsections.

4.2 Estimation procedures

4.2.1 Replacement effect and appointment effect

To discretize changes in the gender composition of managers, we define treatment as an increase in the number of female managers in a firm by one during a certain period. We assign firms that increased their number of female managers by one to the treatment group, while we assign those firms that kept the number constant to the control group. In principle we could also analyze other cases, such as, for example, cases in which we define treatment as an increase in the number of female managers by two or three. In this study, however, we limit our analysis due to the sample size.

In defining the control and treatment groups, we also need to take

---

[7]Pre-treatment values of our productivity measures, TFP and LP, are dropped from the covariates as they cause multicollinearity due to the measures' construction.

into account whether the total number of managers increased or not; it is not valid to compare a firm that created an additional manager post and assigned a woman to that post to a firm that made no change in the numbers of male and female managers at all and thereby retained as many managers as before. In the former case, effects on firm productivity can be attributed not only to the additional female manager but also to the fact that the firm has expanded the number of managerial posts. If the decision to employ an additional manager is a result of some organizational change within the firm, this may result in some productivity impact apart from the gender composition of the managers. Because we want to separate the effect of an increase in the number of female managers and that of an increase in the number of managers in general we must compare firms whose change in the total number of managers is the same. We therefore divide our sample into two cases and implement separate analyses as follows.

The first case deals with firms in which the total number of managers stays constant when treatment was introduced. In this case, we estimate the effect of "replacing" a male manager with a woman while keeping the total number of managers constant. We refer to this as a *replacement effect*. The second case deals with firms in which the total number of managers increased by one when treatment was introduced. In this case, we estimate the effect of "appointing" a female manager instead of a male manager to the new managerial position. We refer to this as an *appointment effect*.[8]

4.2.2 Selection of treated and control firms

As mentioned in Section 3, occupational data for employees working at the majority of small firms are not surveyed annually, so we cannot identify the exact timing of when a firm "replaces" a male manager with a female manager or "appoints" a woman to a new managerial position. According to page 14 in the data documentation (Statistics Sweden 2011b), occupational data for the small firms covered by Statistics Sweden's sample survey is collected in every fourth or fifth year using a rolling scheme. However, according to page 49 in the same documentation (Statistics Sweden

---

[8]We could, in principle, choose other cases too, such as one in which the total number of managers increased by two or more, or even decreased by one. However, we again limit our analyses to the two cases described above due to the sample size.

2011b) and also Statistics Sweden (2011a) and Statistics Sweden (2016), which show the distribution of the last survey year of the occupational data in 2008, 2009 and 2013, respectively, the majority of firms have their occupational data updated with at most a three-year lag.[9] We therefore reasonably assume that changes in occupational data, including the manager or non-manager status, observed in our dataset in $t = T$ have actually occurred within the period of $T - 4 < t < T$. This assumption is important in order to obtain a sufficient sample size.

Figure 3 illustrates our procedure for the estimation of both replacement and appointment effects. In this figure, Firms A and B are a matched pair of observations. We define Firm A as "treated" because the observed number of female managers in Firm A increases by one in standardized year $t = 4$. We in turn define Firm B as our "control" because it kept the observed number of female managers constant from $t = 3$ to $t = 4$. While these changes (or non-changes) in the numbers of managers are based on what we observe in the dataset, the actual changes (or non-changes) have occurred in some year between $t = 0$ and $t = 4$ due to the data limitation we have discussed.

In sum, we define $t \leq 0$, $0 < t < 4$, and $4 \leq t$ as pre-treatment, treatment-introduction, and post-treatment periods, respectively. We therefore use the covariate values at $t = 0$ as pre-treatment covariates when we run the DID-matching estimation discussed in Section 4.1. Estimates of the treatment effects in the post-treatment period can be interpreted as the causal effects of an increased number of female managers if the identifying assumptions are satisfied while the counterpart estimates in the pre-treatment period serve as measures of placebo effects.

In setting up the actual data, we define "treated" firms as follows. When the observed number of female managers in a firm changes by one in a specific year, we define this year as $t = 4$ (i.e. the first post-treatment year) as in Figure 3, and place this firm in the treatment group. When a

---

[9]For instance, the distribution of the last survey years for the occupational data for all employed 16-64 year-olds in 2009 is as follows: 4.2 percent (2010), 70.9 percent (2009), 8.1 percent (2008), 6.2 percent (2007), 3.1 percent (2006), 1.1 percent (2005), 2.9 percent (2004 or before), 3.5 percent (no data). This implies that occupational data for 92.5 percent of all employed 16-64 year-olds were updated at the earliest three years before 2009. The distributions of the last survey years for the occupational data in 2008 and 2013 indicate similar trends.

Figure 3: Illustration of the matching estimation
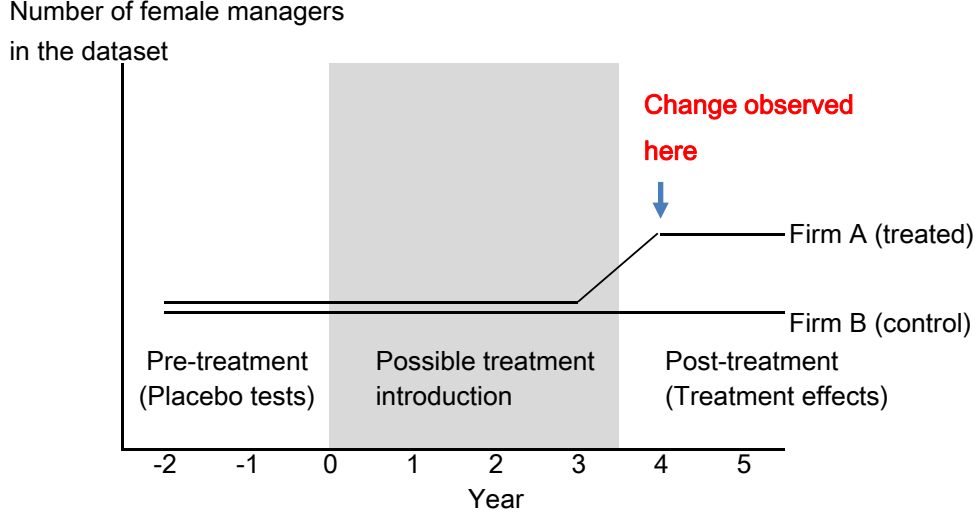
Notes: Firm A (treated): the observed number of female managers increases by one in standardized year $t = 4$. Firm B (control): the observed number of female managers constant from $t = 3$ to $t = 4$. While these changes (or non-changes) are based on what we observe in the dataset, the actual changes (or non-changes) have occurred in some year between $t = 0$ and $t = 4$ due to possible update lag.

firm's observed number of female managers stays constant for two years, we define these two years as $t = 3$ and $t = 4$ and place this firm in the control group. The critical assumption here is that the observed numbers of female managers in $t = 3$ reflect the actual manager numbers in some year t $(0 < t \leq 3)$, and that the actual change or non-change happened some time between $t = 0$ and $t = 4$.

One concern arises regarding the actual number of managers in $t = 4$. How can we be sure that the observed number of managers in $t = 4$ represents the actual value, not some previous non-updated value from $t = 1$, $t = 2$ or $t = 3$ due to the update lag? Regarding the treated group, we can be sure this is the case because we observe a change in the number in $t = 4$. When it comes to the control group, however, no such assurance is possible. To ensure that all firms in the control group do not experience any change in the number of managers in $t = 4$, we impose an additional restriction that the number of female managers was constant from $t = 3$ to $t = 7$. This control-firm criterion is based on the assumption that the

17

maximum lag in the update of occupational data is three years.[10]

Finally, we also need to control for the total number of managers. For the analysis of the *replacement effect*, both the control and the treatment groups kept the number constant between $t = 3$ and $t = 4$. For the analysis of the *replacement effect*, both groups increased it by one during the same period.

In sum, the treatment group ($D_i = 1$) and the control group ($D_i = 0$) in our analyses are defined as below. For the analysis of the *replacement effect*,

$$D_i = \begin{cases} 1, & \text{if } M_{i,3}^f + 1 = M_{i,4}^f \text{ and } M_{i,3} = M_{i,4} \\ 0, & \text{if } M_{i,3}^f = M_{i,t}^f \text{ and } M_{i,3} = M_{i,t} \text{ for } t = 4, 5, 6, 7, \end{cases} \tag{7}$$

where $M_{i,t}$ and $M_{i,t}^f$ stand for the observed numbers of total managers and female managers, respectively, at standardized year t that is set individually for each firm.[11] The additional conditions imposed on the control group are a consequence of the control-firm criterion discussed above. For the analysis of the *appointment effect*,

$$D_i = \begin{cases} 1, & \text{if } M_{i,3}^f + 1 = M_{i,4}^f \text{ and } M_{i,3} + 1 = M_{i,4} \\ 0, & \text{if } M_{i,3}^f = M_{i,4}^f \text{ and } M_{i,3} + 1 = M_{i,4}. \end{cases} \tag{8}$$

In this case we have no need to impose the additional conditions above on the control group.[12]

There is one issue we need to discuss before moving on. We have so

---

[10]Another concern regarding our definition of the treatment and control groups is that while we control for the number of female managers for $t = 3$ and $t = 4$ (and up to $t = 7$ for the control group), we do not precisely control for the number between $t = 0$ and $t = 2$. Some observations in the sample, especially those updated annually, may exhibit numbers in $t = 0$, $t = 1$ and $t = 2$ that are totally different from the number observed in $t = 3$. We are interested in the average treatment effect, however, and because this concern applies to both the treatment and control groups we assume it does not cause a severe bias in our estimation.

[11]That is, we pool series of observations from different points in time, instead of matching observations at the same point in time. We choose this approach to maximize the sample size in our analyses. Some firms may contribute to the sample with more than one firm-spell.

[12]Because we observe a change in the number of managers in $t = 4$, we can be sure that the occupational data has been updated at that point and that the observed and the actual number of managers are the same.

far only considered the gender composition of managers and changes in it. However, managers can be replaced even though the gender composition is unchanged. For example, the treatment group may include observations in which most of the existing managers have been replaced, whereas the control group may include observations where no managers have been replaced. Mobility of managers can be related to firm productivity if, for example, growing firms tend to reorganize and replace their managers, or existing managers tend to leave stagnating firms. For a valid comparison, we must perform matching among firms with a similar degree of manager mobility. At the same time, we are also aware of the possibility that any difference in manager mobility is partly a result of the treatment introduction. In this case, controlling for it may bring a bias in matching and lead to a misleading conclusion. We do not control for manager mobility in our baseline analysis, while we do control for it in a robustness check.

4.2.3 Propensity score matching

We employ propensity score (PS) matching as our matching method. This method is supposed to reduce selection bias by conditioning on covariates using a semi-parametric specification. An alternative matching method is, for instance, nearest-neighbor matching using some distance function of all covariates. Our preliminary application of the nearest-neighbor matching method with the Mahanobis distance metric, however, showed that the resulting balance of the covariates between the control and treatment groups is worse than when we apply PS matching (results not shown). We report below that the balance property of PS matching is acceptable for most of the covariates included in the analysis.

For covariate selection, we introduce a stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015). This procedure selects a subset of covariates and all their second-order interactions that are to be used for the estimation of the propensity score. For details on this procedure, see Appendix A. Throughout the study, we adopt a probit model in estimating propensity scores.

We then match each treated observation with h control observations whose propensity scores are closest to that of the treated observation. Note

that matching is performed with replacement, which implies that the same observation may enter the adjusted sample more than once.

The value of h is selected based on the relative size of the groups. For the analysis of the *replacement effect*, we set $h = 5$ because we have an abundance of observations in the control group relative to the treatment group. For the analysis of the *appointment effect*, we set $h = 2$ because we have two to three times more observations in the control group than in the treatment group.

After matching each treated observation to control observations, we move on to estimation of the average treatment effect on the treated (ATET) by computing the difference between the average changes in the outcome variables from $t = 0$ to an evaluation year $t \neq 0$ for the treatment and control groups. Our estimator is conceptually based on the sample analog of (5) and (6), but in practice we utilize propensity score $p(X_i)$ for matching instead of directly using X_i and use one-to-h matching with replacement for control-group construction. Estimation of the ATET when $h = H$ is expressed as:

$$\hat{\tau}_t^{ATET} = \frac{1}{N^{tr}} \sum_{i:D_i=1} \left\{ (Y_{i,t} - Y_{i,0}) - \frac{1}{H} \sum_{j \in \mathcal{H}_i} (Y_{j,t} - Y_{j,0}) \right\}, \quad (9)$$

where N^{tr} is the number of treated firms and $\mathcal{H}_i$ represents the set of control observations matched to unit i based on the propensity score. The number of control firms in each $\mathcal{H}_i$ is H.

We set $t = -2, -1, 1, 2, \ldots, 5$ so that we evaluate the ATET at every point from $t = -2$ to $t = 5$. We are primarily interested in the ATETs at $t = 4$ and $t = 5$, but to indirectly assess the plausibility of our identifying assumption of unconfoundedness, we perform placebo tests by evaluating the ATETs in the pre-treatment period, $t = -1$ and $t = -2$. In addition, we are also interested in the transition of ATETs during the period of possible treatment introduction, that is, $t = 1, 2,$ and 3.[13]

For each estimation of the ATET, we compute heteroskedasticity-consistent standard errors according to Abadie and Imbens (2006)'s approach, which

---

[13]The propensity score matching and the following estimation of the ATETs are performed using the user-written STATA program, psmatch2, coded by Leuven and Sianesi (2003).

calculates the conditional variance of within-group neighbors based on propensity scores. We set the number of neighbors to 3 throughout this study.

4.3 Outcome variables

4.3.1 Labor productivity and total factor productivity

Our aim in this study is to analyze the productive effects of gender equality at the managerial level. To measure the effects, we choose two measures of firm productivity: labor productivity (LP) and total factor productivity (TFP). In this subsection, we define these terms in the difference-in-differences framework and explain our choice of outcome variables.

We define labor productivity as the log of value added per unit of labor,

$$LP_{i,t} \equiv \ln\left(V_{i,t}/L_{i,t}\right) = \ln V_{i,t} - \ln L_{i,t}, \tag{10}$$

where $V_{i,t}$ and $L_{i,t}$ stand for value added and labor for firm i at time t, respectively. Since we have no information on hours worked in our dataset, we use the number of employees as the unit of labor $L_{i,t}$. In the difference-in-differences framework, a change in this measure from time 0 to t, i.e. $LP_{i,t} - LP_{i,0}$, enters as the outcome variable. This differenced term represents the growth rate of labor productivity, which is in turn equivalent to the difference in the growth rate between value added and labor:

$$LP_{i,t} - LP_{i,0} = (\ln V_{i,t} - \ln V_{i,0}) - (\ln L_{i,t} - \ln L_{i,0}) \tag{11}$$

A shortcoming of labor productivity measures is that they do not take into account differences in capital stocks across firms, an issue that can be overcome by a TFP measure. We define TFP as the log of value added minus the logs of labor and capital stocks with appropriate weights β_L and β_K, as

$$TFP_{i,t} \equiv \ln V_{i,t} - (\beta_L \ln L_{i,t} + \beta_K \ln K_{i,t}). \tag{12}$$

TFP is usually perceived as a residual in growth account after labor and capital are accounted for. However, it can also be interpreted as output per

a compound set of factor inputs, $\ln(V_{i,t}/L_{i,t}^{\beta_L} K_{i,t}^{\beta_K})$. It is then apparent that the differenced term $TFP_{i,t} - TFP_{i,0}$, which enters as an outcome variable in the diffenrence-in-differences framework, not only represents the growth rate of total factor productivity, but also shows the difference in growth rates between value added and weighted inputs:

$$
\begin{aligned}
TFP_{i,t} - TFP_{i,0} = {} & (\ln V_{i,t} - \ln V_{i,0}) \\
& - [\beta_L(\ln L_{i,t} - \ln L_{i,0}) + \beta_K(\ln K_{i,t} - \ln K_{i,0})]
\end{aligned}
\tag{13}
$$

We obtain consistent estimates of β_L and β_K from estimation of a production function based on the approach proposed by Ackerberg et al. (2015). Since we assume that the output elasticities β_L and β_K vary depending on firm size, we adopt a translog, rather than a Cobb-Douglas, function form for the production function. The online Appendix D explains the estimation procedure.

As is shown above, a productivity change is measured as the difference in growth between output and inputs. A positive change in productivity between two periods can therefore be caused by growth in the output and/or a reduction in the input. Whether it is the output side or the input side that drives the productivity change has different implications. Therefore, it is important to analyze how the output (i.e. value added) and the input (i.e. labor and capital) are affected by the treatment. Related to labor, we are also interested in shifts in the ratio of full-time to part-time workers. We will draw different implications if an increase in labor (i.e. the number of employees) is explained by a change in full-time or part-time workers. As we have no data on hours worked, we use labor costs per employee as a proxy for the share of part-time workers.

In summary, the outcome variables in our matching estimation are labor productivity $LP_{i,t}$, total factor productivity $TFP_{i,t}$, log of value added $\ln V_{i,t}$, log of labor $\ln L_{i,t}$, log of capital $\ln K_{i,t}$ and log of labor costs per employee $\ln(W_{i,t}/L_{i,t})$, where $W_{i,t}$ stands for labor costs.

4.3.2 Removing sector-specific trends

Our estimation using propensity score matching is based on firm-level observations. We pool series of observations of different sectors and years.

Consequently, our outcome variables are most likely to be affected by sector-specific time trends or shocks, which may contaminate estimates of the treatment effects.

We disentangle sector-specific time effects by running the following OLS regression for each outcome variable and each two-digit sector using all observations in our dataset:

$$Y_{i,t} = \alpha_0 + \sum_t \alpha_t D_t + e_{i,t} \tag{14}$$

where $Y_{i,t}$ is an outcome variable. We denote the detrended values of the original variable $Y_{i,t}$ as $\tilde{Y}_{i,t}$, which is defined as

$$\tilde{Y}_{i,t} = \hat{\alpha}_0 + \hat{e}_{i,t} \tag{15}$$

where $\hat{\alpha}_0$ and $\hat{e}_{i,t}$ indicate estimates of α_0 and $e_{i,t}$. The detrended values are interpreted as deviations from the sector-specific year average. In the estimation of treatment effects under the DID framework, the difference in these detrended values are entered as outcome variables.

4.4 Covariate candidates

Because we cannot exploit quasi-experimental exogenous variation, it is important to carefully choose pre-treatment covariates $X_{i,0}$ to make the common trend assumption (4) plausible. We use a rich set of covariates that are presumed to determine treatment selection and/or likely to be associated with outcome variables. Firstly, we control for firm size in terms of value added, the number of employees, and capital stocks because firm size is one of the most important determinants of productivity (e.g. Taymaz (2005) and Yang and Chen (2009)). Value added and capital stocks are deflated by respective deflators computed from the national accounts at the two-digit sectoral level.

We also control for the number of managers in total and that of female managers, and the share of female managers. We expect the total number of managers to positively affect treatment selection, because it may be easier for a firm with plenty of managerial positions to increase the number of female managers by replacement or expansion. We also expect that a

firm that already has female managers is more likely to appoint additional female managers than a firm that does not. However, this likelihood may actually decrease when the share of female managers is high.[14]

As we mentioned in Section 4.2.3, we adopt the stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015) (for details on this procedure, see Appendix A). In this procedure, we specify "basic" and "additional" covariates. The former refers to those covariates that are a priori regarded as important for explaining the treatment assignment or as likely related to the outcomes. We choose the above-mentioned six covariates as "basic".

As other potentially important covariates, referred to as "additional" covariates, we choose several variables that account for a firm's socio-economic characteristics related to the human capital of the employees. These covariates include the share of female employees, the share of university-educated employees and the shares of employees and current managers in different age groups. We also want to control for the share of part-time workers. Unfortunately, we have no data on hours worked by each employee. We use labor costs per employee at the firm level as a proxy for the share of part-time workers, though we are aware that variations in this proxy may also reflect differences in human capital.

Lastly, we add dummies for counties, the two-digit sector and years. These dummies only enter the probit model linearly (i.e. they have no interaction terms with other covariates). Our preliminary estimation of the probit models (results not shown) showed that some of the pseudo-outcomes in the pre-treatment period were significant, indicating that we were not successful in achieving common pre-treatment trends between the controls and the treated. This result suggested that we need to control for the dynamism of firm development. For example, a growing firm's expansion cannot be controlled for merely by the covariates at $t = 0$. We therefore added the one-period lagged values of the basic covariates as additional covariates to capture the dynamism from $t = -1$ to $t = 0$.[15]

---

[14]Our productivity measures, TFP and LP, do not enter the model as they cause multicollinearity.

[15]A similar approach, where treatment selection is conditioned on lagged values of some of the original covariates, is used by Acemoglu et al. (2014).

5 Estimation of the treatment effect

5.1 The balance of covariates

The descriptive statistics of the covariates and other major variables in our sample are presented in Table 1 (for the analysis of the replacement effect) and in Table 2 (for the analysis of the appointment effect). We have 11,216 control observations and 582 treated observations in the analysis of the replacement effect, whereas we have 3,801 control observations and 1,775 treated observations in the analysis of the appointment effect.[16]

The tables begin by reporting the potential covariates used in the estimation of propensity score. Note that we adopt the stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015). Because some of the covariates are dropped during the procedure, not all of those presented in this table are used in the actual estimation. The tables then go on to report other major variables. Here we present the estimated propensity score, TFP and LP. Although the log of the number of employees enters the probit estimation, we also report the number of employees for convenience in making comparisons.

To investigate the balance of the covariates between the control and treatment groups, we introduce normalized differences as suggested by Imbens and Wooldridge (2009) and Imbens and Rubin (2015). This is a scale-free measure of the differences between the control and treatment groups and is constructed as the difference in means, scaled by the square root of the average of the two within-group variances. For details of its computation, see Appendix B. This measure has an advantage over the t-value, which is usually used for a test of a difference in means, in that it is not affected by the sample size. Imbens and Wooldridge (2009) and Imbens and Rubin (2015) argue as a rule of thumb that a normalized difference over 0.25 in the absolute value is an indication of a large bias.

The normalized difference before and after the matching adjustment (labeled as *unadjusted* and *adjusted*, respectively) are presented in the right part of Table 1 and 2. The estimated probit model for treatment selection is reported in Table E.1 and Table E.2 in the online Appendix E for respec-

---

[16]Some firms in the dataset report having one manager though they only have one or two employees in total. We exclude such firms from our analysis.

Table 1: Normalized differences (analysis of the *replacement* effect)

Analysis of the *replacement effect*	Treated N₁ = 582		Control N₀ = 11,216		Normalized diff.	
Covariates	Mean	(s.d.)	Mean	(s.d.)	unadjusted	adjusted
Basic covariates						
ln (Value added)	16.844	(1.230)	15.260	(0.852)	1.496	0.097
ln (# of employees)	3.446	(1.101)	2.101	(0.684)	1.467	0.147
ln (Capital stocks)	14.581	(2.579)	13.075	(2.585)	0.583	0.244
# of all managers	5.675	(10.217)	1.428	(0.841)	0.586	0.162
# of female managers	1.701	(5.180)	0.292	(0.567)	0.382	0.105
Share of female managers	0.232	(0.278)	0.186	(0.349)	0.145	0.009
Additional covariates						
Share of female employees	0.448	(0.226)	0.403	(0.294)	0.172	0.060
Share of employees with university education	0.191	(0.208)	0.082	(0.157)	0.588	-0.111
ln (Labor costs per employee)	13.181	(0.370)	12.939	(0.337)	0.685	0.052
Share of employees in age -25 (benchmark)						
Share of employees in age 26-35	0.288	(0.157)	0.232	(0.180)	0.331	-0.063
Share of employees in age 36-45	0.267	(0.124)	0.256	(0.187)	0.068	0.084
Share of employees in age 46-55	0.176	(0.116)	0.217	(0.182)	-0.264	-0.135
Share of employees in age 56-65	0.122	(0.116)	0.126	(0.148)	-0.032	0.077
Share of employees in age 66-	0.011	(0.029)	0.013	(0.048)	-0.057	0.056
Share of managers in age -35 (benchmark)						
Share of employees in age 36-45	0.338	(0.318)	0.342	(0.440)	-0.010	0.006
Share of employees in age 46-55	0.298	(0.324)	0.389	(0.455)	-0.231	-0.002
Share of employees in age 56-65	0.188	(0.285)	0.160	(0.339)	0.091	0.012
Share of employees in age 66-	0.015	(0.079)	0.011	(0.089)	0.045	0.018
Dummy variables						
County dummies (only major countries presented)						
Stockholm county	0.402	(0.491)	0.234	(0.424)	0.366	0.082
Skåne county	0.119	(0.324)	0.135	(0.342)	-0.049	0.019
Västra götaland county	0.163	(0.370)	0.169	(0.375)	-0.015	-0.023
Sector dummies (aggregated into the 1-digit level)						
Wholesale. retail trade, hotels, restaurants	0.455	(0.498)	0.656	(0.475)	-0.413	0.015
Transport, communication, financial intermediation	0.081	(0.273)	0.084	(0.278)	-0.013	0.141
Real estate, renting, IT, research, business activities	0.297	(0.457)	0.167	(0.373)	0.313	-0.156
Education, health, social work	0.100	(0.300)	0.062	(0.241)	0.140	-0.019
Other community, social and personal service activities	0.067	(0.250)	0.031	(0.173)	0.167	0.145
Other major variables						
Propensity score	0.412	(0.349)	0.030	(0.072)	1.514	0.022
Total factor productivity (TFP)	12.873	(0.329)	12.671	(0.271)	0.672	-0.094
Labor productivity (LP)	13.398	(0.573)	13.160	(0.490)	0.447	-0.073
# of employees	65.893	(134.042)	10.871	(11.807)	0.578	0.253

Notes: The treatment group consists of the observations in which the firm *replaces* a male manager with a woman, whereas the control group consists of the observations in which the numbers of male and female managers are unchanged. Note that the stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015) chooses covariates and their interaction terms in the estimation of propensity score. Because some covariates are dropped during the procedure, not all of those presented in this table are used in the estimation. The probit model estimated to generate the propensity score for matching is reported in Table E.1 in the online Appendix E.

tive analysis type. We observe a considerable reduction in the normalized differences for many of the covariates after the matching adjustment. All potential covariates have normalized differences below 0.25 in the absolute value.

Table 2: Normalized differences (analysis of the *appointment* effect)

Analysis of the *appointment effect*	Treated N$_1$ = 1,775		Control N$_0$ = 3,801		Normalized diff.	
Covariates	Mean	(s.d.)	Mean	(s.d.)	unadjusted	adjusted
Basic covariates						
ln (Value added)	16.295	(1.225)	16.388	(1.163)	-0.078	0.013
ln (# of employees)	2.989	(1.072)	2.982	(1.010)	0.007	0.010
ln (Capital stocks)	13.880	(2.774)	14.186	(2.704)	-0.112	0.062
# of all managers	3.658	(7.157)	3.453	(5.296)	0.033	0.027
# of female managers	1.373	(4.390)	0.675	(1.593)	0.211	0.043
Share of female managers	0.344	(0.375)	0.174	(0.281)	0.513	0.006
Additional covariates						
Share of female employees	0.535	(0.261)	0.332	(0.229)	0.830	0.026
Share of employees with university education	0.188	(0.222)	0.156	(0.208)	0.146	-0.042
ln (Labor costs per employee)	13.088	(0.380)	13.152	(0.364)	-0.173	-0.009
Share of employees in age -25 (benchmark)						
Share of employees in age 26-35	0.277	(0.171)	0.268	(0.167)	0.052	-0.068
Share of employees in age 36-45	0.263	(0.146)	0.266	(0.144)	-0.021	0.041
Share of employees in age 46-55	0.173	(0.134)	0.192	(0.137)	-0.134	0.023
Share of employees in age 56-65	0.120	(0.129)	0.138	(0.135)	-0.140	0.016
Share of employees in age 66-	0.012	(0.036)	0.012	(0.035)	0.005	-0.006
Share of managers in age -35 (benchmark)						
Share of employees in age 36-45	0.345	(0.369)	0.327	(0.359)	0.049	0.053
Share of employees in age 46-55	0.300	(0.361)	0.310	(0.363)	-0.028	0.013
Share of employees in age 56-65	0.191	(0.314)	0.218	(0.329)	-0.084	-0.010
Share of employees in age 66-	0.020	(0.109)	0.019	(0.104)	0.006	-0.010
Dummy variables						
County dummies (only major countries presented)						
Stockholm county	0.372	(0.483)	0.304	(0.460)	0.144	-0.025
Skåne county	0.127	(0.333)	0.128	(0.334)	-0.004	0.033
Västra götaland county	0.180	(0.384)	0.194	(0.395)	-0.036	0.014
Sector dummies (aggregated into the 1-digit level)						
Wholesale. retail trade, hotels, restaurants	0.499	(0.500)	0.559	(0.497)	-0.121	0.041
Transport, communication, financial intermediation	0.066	(0.249)	0.094	(0.292)	-0.103	-0.007
Real estate, renting, IT, research, business activities	0.250	(0.433)	0.258	(0.438)	-0.020	-0.051
Education, health, social work	0.137	(0.344)	0.045	(0.207)	0.325	0.042
Other community, social and personal service activities	0.048	(0.214)	0.043	(0.204)	0.021	-0.050
Other major variables						
Propensity score	0.430	(0.199)	0.266	(0.158)	0.907	0.001
Total factor productivity (TFP)	12.813	(0.328)	12.846	(0.336)	-0.101	-0.011
Labor productivity (LP)	13.307	(0.536)	13.406	(0.564)	-0.182	0.011
# of employees	41.436	(102.779)	37.558	(91.961)	0.040	-0.010

Notes: The treatment group consists of the observations in which the firm *appoints* a female manager to a new managerial position, whereas the control group consists of the observations in which the firm appoints a male manager to a new managerial position. Note that not all of the covariates presented in this table are used in the final estimation of treatment effects because of the stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015). The probit model estimated to generate the propensity score for matching is reported in Table E.2 in the online Appendix E.

5.2 Estimation results

The estimates of the average treatment effect on the treated (ATET) are illustrated for each outcome in Figure 4. The dotted lines show the confidence intervals at the 95 percent level. For the original estimates and the heteroskedasticity-consistent standard errors, see Tables C.1 and C.2 in Appendix C.

When we look at the treatment effects after $t = 0$, we observe no significant effects on the two productivity measures. At the same time, the

Figure 4: DID-matching estimates of the treatment effects
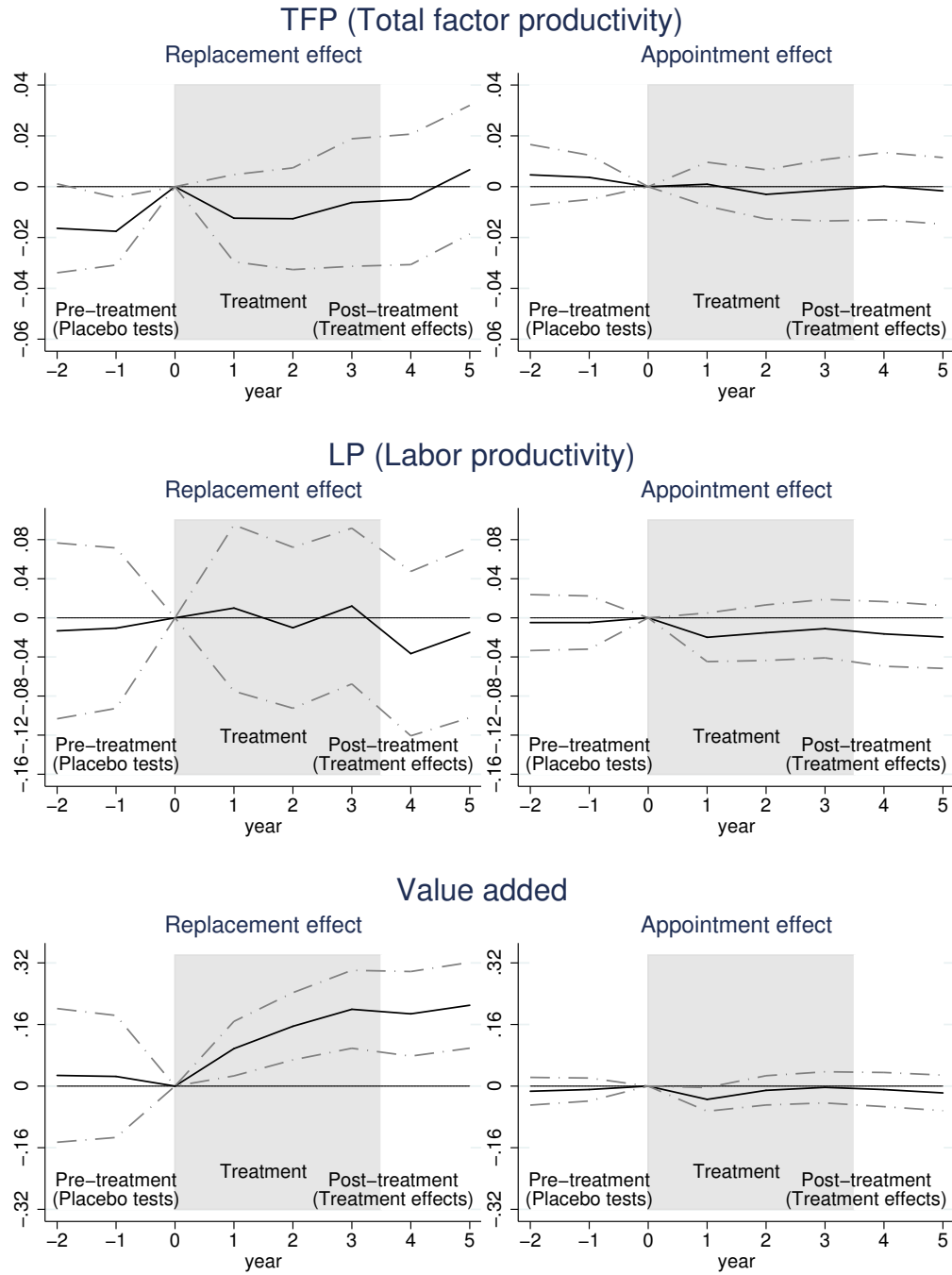
TFP (Total factor productivity)

LP (Labor productivity)

Value added

Notes: The solid and dotted lines show the point estimates of the average treatment effects on the treated (ATET) and their confidence intervals at the 95 percent level, respectively. The treatment is introduced during $0 < t < 4$. The estimates in $t \geq 4$ are our main focus. While the estimates in the period $t < 0$ serve as measures of placebo tests, the estimates in $0 < t < 4$ show the transition during the treatment introduction. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

Figure 4 (cont.): DID-matching estimates of the treatment effects
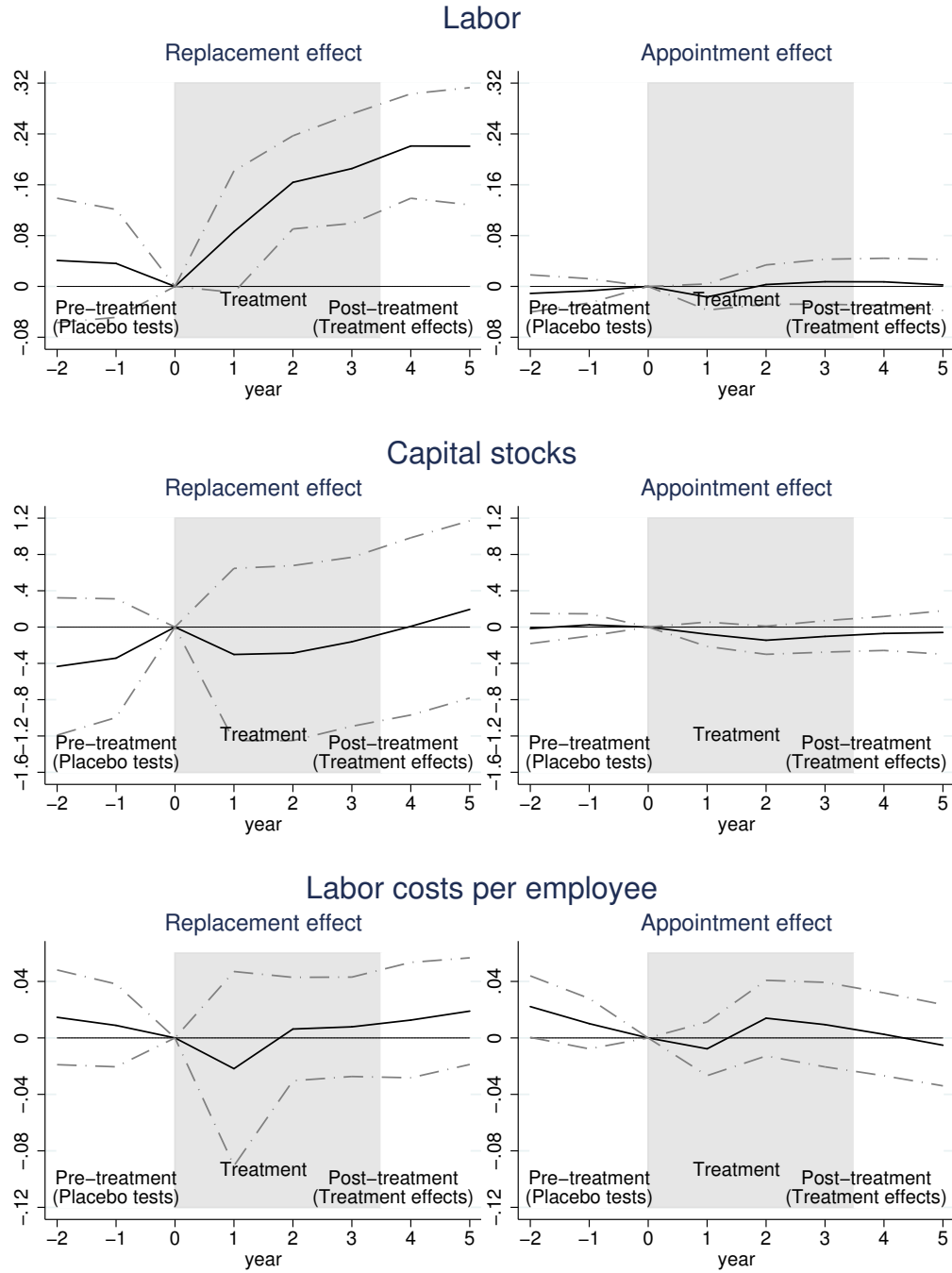
Labor

Capital stocks

Labor costs per employee

Notes: The solid and dotted lines show the point estimates of the average treatment effects on the treated (ATET) and their confidence intervals at the 95 percent level, respectively. The treatment is introduced during $0 < t < 4$. The estimates in $t \geq 4$ are our main focus. While the estimates in the period $t < 0$ serve as measures of placebo tests, the estimates in $0 < t < 4$ show the transition during the treatment introduction. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

results show significant positive effects on value added and labor input in the analysis of the replacement effect. The magnitude of the effects is around 0.19 and 0.22 respectively. The finding that significant effects on the two outcomes are already observed during the treatment introduction (i.e. before $t = 4$) suggests the update lag discussed above.

In the analysis of the appointment effect, we find no sustaining treatment effects, though value added and capital stocks occasionally show some significance at the 10 percent level (see Table C.2). Figures 4 and 4 illustrate that the estimates of the appointment effects are in general much smaller in magnitude than those of the replacement effects.

Our results are based on the assumption of unconfoundedness. Although we have included a rich set of covariates in the matching procedure, we cannot control for the effect of unobserved factors on changes in the outcome variables. To indirectly assess the validity of our identifying assumption, we perform placebo tests by evaluating the treatment effect in $t = -1$ and $t = -2$ (so-called *pseudo-outcomes*). If our assumption is valid, we do not expect any statistically significant estimates in the pre-treatment period. The results indicate significant placebo estimates in TFP (in the analysis of the replacement effect) and per-employee labor cost (in the analysis of the appointment effect), implying that we were not totally successful in balancing the two groups.

5.3 Robustness check

5.3.1 Control for manager mobility

We have found significant replacement effects on some of the outcome variables. However, one concern is a difference in manager mobility between the two groups in comparison. While treatment is defined as a change in the number of female managers, we have so far not controlled for other manager changes that do not affect gender composition (e.g. a replacement of male manager with another man).

Table 3 compares the number of newly appointed managers in $t = 4$ between the two groups following the matching adjustment in the respective baseline analysis. In the analysis of the replacement effect, manager mobility of the treated firms is clearly higher than that of the control firms. In a

way, this is to be expected. The control firms may or may not have replaced managers during the treatment introduction. They have, by definition, no manager change that affects the gender composition of the managers. In contrast, the treated firms replaced at least one manager. In the analysis of the appointment effect, manager mobility is similar between the two groups even though the difference is significant.[17]

Table 3: Manager mobility in the baseline analysis

# of new managers in $t = 4$	Treated (mean)	Control (mean)	Difference	(s.e.)
Analysis of the *replacement effect*	1.555	0.107	1.448 ***	(0.065)
Analysis of the *appointment effect*	1.426	1.279	0.147 ***	(0.047)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The average numbers of new managers in $t = 4$ are reported (after the matching adjustment). The difference and the standard errors are calculated by treating the new manager number as an outcome variable in the ATET computation. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

On one hand, the difference in manager mobility can be a result of the treatment introduction. If this is the case, we have a good reason not to control for this post-treatment variable. On the other hand, it is also plausible that a difference in manager mobility reflects the fact that the firms lie in different development phases. For example, growing firms may go through reorganization and replace their managers, or current managers may leave stagnating firms. In these cases, there is an argument to actively control for total replacement of managers in order to disentangle gender effects on firm productivity and related outcomes from other productivity-related factors.

To check robustness of the estimated replacement effect, we perform a new matching that accounts for manager mobility. In this matching, we limit our sample to the observations in which the firm only replaced one manager with a new one in $t = 4$ while keeping the total number of managers constant. This then means that the firms in the treatment group replaced a male manager with a female manager, whereas the firms in the control group switched managers of the same gender. In both groups, other

---

[17]Both the control and the treated firms assigned at least one new manager, which positively affects manager mobility. Therefore, it is expected that manager mobility is similar between the two groups.

managers are unchanged between the two periods.[18]

In addition, we also investigate whether the gender replacement effect is symmetric by performing yet another matching analysis. If a male-to-female manager change yields positive gender effects, we may find negative gender effects when a female manager is replaced by a male. In this analysis, we assign firms that replaced a female manager with a male manager in $t = 4$ to the treatment group. The control group consists of the firms that switched managers of the same gender. Again, in both groups, other managers are unchanged.

The descriptive statistics of the covariates and other major variables in our samples are presented in Table E.3 (male-to-female manager switch) and in Table E.4 (female-to-male manager switch) in the online Appendix E. We have 1,669 control observations and 428 treated observations in the former analysis, whereas we have 1,671 control observations and 342 treated observations in the latter analysis[19]. We match each treated observation with three observations from the control group based on the propensity scores estimated by probit models. The estimates of the probit models are presented in Tables E.5 and E.6 in the online Appendix E.

Figure 5 illustrates the estimates of the average treatment effects on the treated (ATETs) for both analyses by outcome. For the original estimates and the heteroskedasticity-consistent standard errors, see Tables C.3 and C.4 in Appendix C.

In the analysis of male-to-female manager changes, we still observe significant treatment effects on value added and labor even though the magnitude has decreased, from 0.19 to 0.08 for value added and from 0.22 to 0.07 for labor. This result suggests that manager mobility cannot entirely account for the positive replacement effects found in the baseline analysis.

In the analysis of female-to-male replacement, capital stocks respond positively during treatment introduction and LP reacted slightly negatively in $t = 5$. However, since these effects are either weak or not enduring,

---

[18]Although it is also possible to extend the sample by including observations with changes of two existing managers or more, we exclude them from the sample because there are few such observations.

[19]Although the control groups in both analyses are in principle the same, a difference can arise because some observations are dropped during the estimation of a probit model due to perfect prediction within some of the subgroups defined by the dummy variables.

Figure 5: DID-matching estimates of the treatment effects
(Controlling for manager mobility)
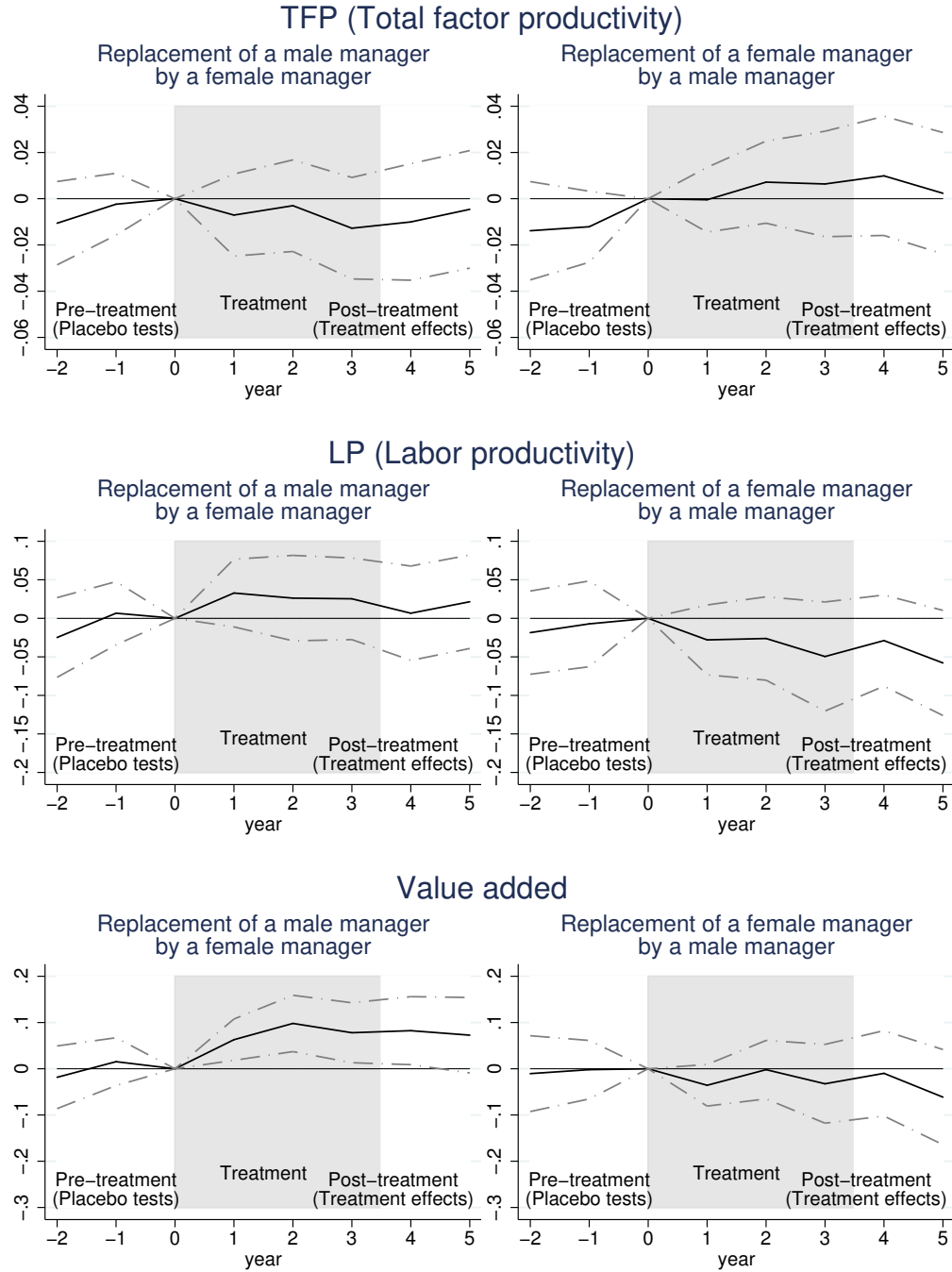
TFP (Total factor productivity)

Replacement of a male manager by a female manager

Replacement of a female manager by a male manager

LP (Labor productivity)

Replacement of a male manager by a female manager

Replacement of a female manager by a male manager

Value added

Replacement of a male manager by a female manager

Replacement of a female manager by a male manager

Notes: The solid and dotted lines show the point estimates of the average treatment effects on the treated (ATET) and their confidence intervals at the 95 percent level, respectively. The treatment is introduced during $0 < t < 4$. The estimates in $t \geq 4$ are our main focus. While the estimates in the period $t < 0$ serve as measures of placebo tests, the estimates in $0 < t < 4$ show the transition during the treatment introduction. The standard errors are heteroskedasticity-robust based on Abadie and Imbens (2006).

Figure 5 (cont.): DID-matching estimates of the treatment effects
(Controlling for manager mobility)
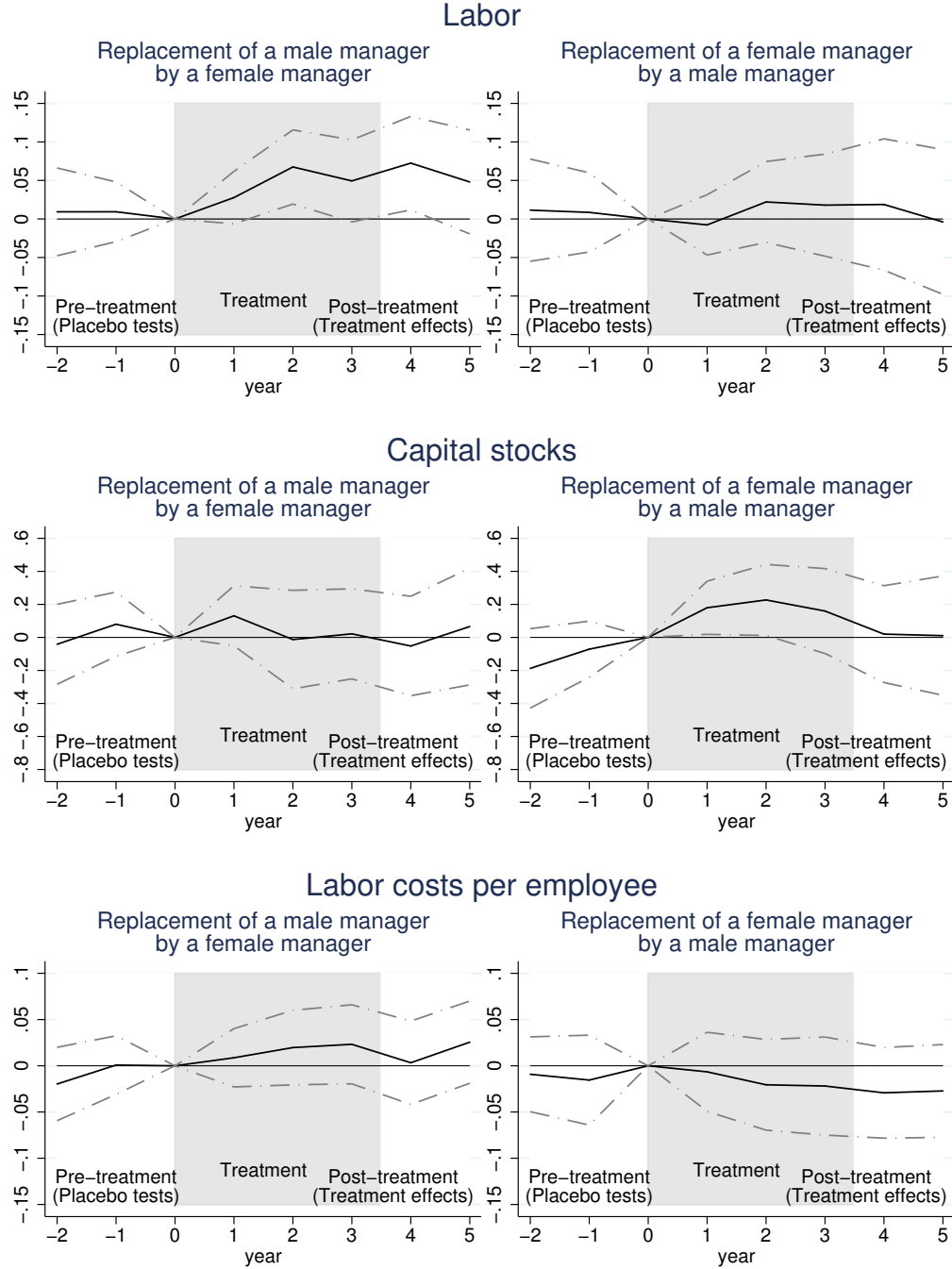
Labor

Capital stocks

Labor costs per employee

Notes: The solid and dotted lines show the point estimates of the average treatment effects on the treated (ATET) and their confidence intervals at the 95 percent level, respectively. The treatment is introduced during $0 < t < 4$. The estimates in $t \geq 4$ are our main focus. While the estimates in the period $t < 0$ serve as measures of placebo tests, the estimates in $0 < t < 4$ show the transition during the treatment introduction. The standard errors are heteroskedasticity-robust based on Abadie and Imbens (2006).

34

it is difficult to draw any conclusion from them. We find no significant treatment effects on the other variables.

5.3.2 Potential mechanisms

In Section 2.1, we listed potential mechanisms through which the gender composition of managers could affect firm productivity and related outcomes. They include (1) labor diversity, (2) gender-specific cognitive abilities, (3) a manager recruitment pool with greater human capital, and (4) signaling effects. While the first two mechanisms are not directly measurable, we can partly measure managers' human capital (3) and other employees' human capital (4) in relation to treatment introduction. In this subsection, we analyze the human capital composition of new managers and the whole workforce of the control and treatment groups to explore potential mechanisms that drive the positive male-to-female replacement effects on value added and labor.

First, Table 4 reports the average age of newly appointed managers at $t = 4$ and the share of those having university education among them in the male-to-female analysis. The average age of new managers in the treated firms is slightly younger at the 10 percent significance level, though the difference is only around one year. The share of new managers with university education is higher for the treated firms, but the difference is insignificant.

Table 4: Analysis of the new managers (male-to-female manager replacement)

Analysis of the *replacement effect* (a male-to-female change)	Treated (mean)	Control (mean)	Difference	(s.e.)
Age of new manager	43.084	44.146	-1.062 *	(0.614)
Share of new managers with univ. ed.	0.360	0.328	0.032	(0.032)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The table reports the average age of new managers and the share of the new managers with university education in $t = 4$ (after the matching adjustment). The difference and the standard errors are calculated by treating these variables as outcomes in the ATET computation. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

Next, Table 5 reports how the average age, the share of female employees, and the share of university-educated employees develop before and after treatment introduction. The shares of female employees and university-

Table 5: Analysis of all employees (male-to-female manager replacement)

Analysis of the *replacement effect* (a male-to-female manager change)					Control: N_0 = 1,669			Treated: N_1 = 428	
		pre-treatment			under treatment introduction			post-treatment	
Outcome variable (detrended)	time	-2	-1	0	1	2	3	4	5
Average age of employees	**ATET**	**-0.133**	**-0.160**	**0.000**	**-0.076**	**-0.150**	**-0.202**	**-0.237**	**-0.099**
	s.e.	(0.164)	(0.133)		(0.128)	(0.154)	(0.185)	(0.203)	(0.212)
Share of female employees	**ATET**	**-0.004**	**-0.001**	**0.000**	**0.010 ****	**0.012 ****	**0.012 ***	**0.043 *****	**0.040 *****
	s.e.	(0.006)	(0.005)		(0.005)	(0.005)	(0.006)	(0.007)	(0.007)
Share of employees with univ. ed.	**ATET**	**0.000**	**0.003**	**0.000**	**0.004**	**0.009 ****	**0.006**	**0.007**	**0.012 ***
	s.e.	(0.004)	(0.004)		(0.004)	(0.004)	(0.005)	(0.006)	(0.007)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The difference-in-differences treatment effects on these three variables regarding all employees are estimated in exactly the same way as the six outcome variables reported in Figure 5. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

educated employees in the treated firms changed positively by 4 percent and 1 percent, respectively, relative to the control firms, though the latter figure is only significant at the 10 percent level.

These results suggest that the treated firms are likely to attract younger managers, more female employees, and more educated employees after treatment. Although we cannot identify the exact mechanisms, the fact that managers became younger in the treated firms relative to the control firms may have contributed directly to firm growth in terms of value added and labor. Even if younger managers do not always imply higher human capital, this indicates that the recruitment pool of female managers is different from that of male managers in terms of age. In addition, the increased number of female managers may have triggered a signaling effect and increased the shares of female employees and university-educated employees, which in turn may have yielded impacts on firm growth.

Nevertheless, these differences in manager age and the share of female employees and university-educated employees are not striking. It does not seem plausible that these factors fully explain the positive manager-replacement effects on firm growth. To understand the strength of these factors' explanatory power, we perform yet another analysis of a male-to-female manager switch. This time we account for these three variables, though we are aware of the risk of controlling for the post-treatment variables. The result (not presented) shows that the positive replacement effects on value added and labor remain, indicating that we cannot rule out the effects of labor diversity and gender-specific cognitive abilities as causes of the observed effects on firm growth.

For comparison, we also investigate how the human capital of new managers and all employees develop subsequent to a female-to-male manager replacement. The results are presented in Tables 6 and 7. The average age and the education level of new managers do not differ significantly between the control and treated firms. We also observe that the average age of all employees became younger in the treated firms relative to the control firms. However, the difference is less than one year. We also observe a decrease in the share of female employees of around 4 percent. The fact that we did not observe any treatment effects on outcome variables may indicate that these differences are not important for outcomes, or that any effects are canceled out by each other.

Table 6: Analysis of the new managers (from-female-to-male manager replacement)

Analysis of the *replacement effect* (a female-to-male change)	Treated (mean)	Control (mean)	Difference	(s.e.)
Age of new manager	44.421	43.823	0.598	(0.824)
Share of new managers with univ. ed.	0.345	0.346	-0.001	(0.040)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The table reports the average age of new managers and the share of the new managers with university education in $t = 4$ (after the matching adjustment). The difference and the standard errors are calculated by treating these variables as outcomes in the ATET computation. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

Table 7: Analysis of all employees (female-to-male manager replacement)

Analysis of the *replacement effect* (a female-to-male manager change)		pre-treatment			under treatment introduction			post-treatment	
		Control: N₀ = 1,671					Treated: N₁ = 342		
Outcome variable (detrended)	time	-2	-1	0	1	2	3	4	5
Average age of employees	ATET	-0.378 **	-0.349 *	0.000	-0.283	-0.575 ***	-0.678 **	-0.911 ***	-0.509 *
	s.e.	(0.190)	(0.194)		(0.251)	(0.217)	(0.272)	(0.299)	(0.282)
Share of female employees	ATET	-0.024 ***	-0.023 **	0.000	-0.011	-0.012	0.004	-0.037 ***	-0.039 ***
	s.e.	(0.007)	(0.009)		(0.010)	(0.008)	(0.011)	(0.012)	(0.012)
Share of employees with univ. ed.	ATET	0.005	-0.001	0.000	0.004	-0.001	0.003	0.008	0.011
	s.e.	(0.006)	(0.005)		(0.006)	(0.006)	(0.009)	(0.009)	(0.010)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The difference-in-differences treatment effects on these three variables regarding all employees are estimated in exactly the same way as the six outcome reported in Figures 5. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

5.4 Discussion

Our baseline results suggest that gender composition does not have any impact on firm productivity. Nevertheless, we have found significant positive

effects on firms' rate of growth in terms of value added and the number of employees when the firms replace an existing male manager with a woman. In contrast, we have found no such effects when firms appoint a women, instead of a man, to a newly created managerial position.

In the following analysis, we further investigated these positive *replacement effects*. A drawback of our baseline analysis was that manager mobility differed substantially between the groups: while the treatment group consisted of observations in which one or more managers were replaced, the control group is dominated by observations in which there are no manager replacements, even within the same gender.

As a robustness check of our baseline result, we limited our sample to the observations that replaced one manager: treated firms implemented a male-to-female replacement whereas control firms carried out a manager replacement within the same gender. The other managers remained unchanged in both groups. The result showed that the positive effects on value added and labor became weaker but still remained significant. A further investigation revealed that the treated firms on average assign a younger manager than the control firms and that the share of female employees and university-educated employees increased in the treated firms after treatment introduction. These findings suggest that these shifts partially drive the positive replacement effects on firm growth in terms of value added and labor input.

At the same time, we cannot exclude potential effects of labor diversity and gender-specific cognitive abilities because we still observe a positive effect even after controlling for the differences in managers' and employees' human capital composition. Moreover, the finding that treated firms respond to the positive shock on output (value added) by increasing labor inputs rather than capital inputs suggests that male and female managers have different preferences when making decisions on internal resource allocation.[20]

---

[20]One may suspect that the input bias towards labor inputs is a result of the factor prices: i.e. firms with more female managers face lower labor costs due to a higher share of female employees with lower wages. However, this potential explanation is not plausible in this case because the gender wage gap in Sweden is fairly small. The fact that we have not found any negative effect on labor costs per employee supports its exclusion.

Finally, the absence of any significant appointment effect suggests that the gender of managers does not matter when firms expand.

6 Conclusion

In contrast to recent studies that examine the productivity effects of female board members or their interaction with female workers, we have analyzed whether gender composition at non-board managerial levels has any impact on firm productivity and other related outcomes in the Swedish service sector. To control for endogeneity between firms' outcomes and gender composition, we have applied a difference-in-differences matching method based on propensity score.

In the analysis of the *replacement effect* – the effect of an existing manager position being reassigned from a man to a woman, we found no treatment effects on total factor productivity and labor productivity. At the same time, we did observe significant impacts on firm size in terms of value added and the number of employees. Our robustness test confirmed this result.

In the analysis of the *appointment effect* – the effect of a new managerial position being given to a woman rather than a man, we found no impact on the productivity measures or other related outcomes. This result suggests that the gender of managers does not matter for firms in an expansion phase.

The implications of our empirical results are two-fold. First, assigning or promoting more women to managerial positions does not yield positive effects on firm productivity, nor does it cause negative productivity effects. As long as productivity is the most important measure for firms to consider in human resource management, female managers do not seem to be inferior to male colleagues, at least in the service sector in Sweden. Gender inequality at the managerial level in Sweden cannot be justified by gender difference in productivity. This result is consistent with recent studies that underline the fact that structural transformation towards a more service-oriented economy has raised the comparative advantage of female workers, and that it has partly driven the rise in labor participation of women as well as the decline in the wage gap between genders (e.g. Olivetti and

Petrongolo (2016)).

Second, our finding that a male-to-female replacement of managers influences firm growth in terms of value added and labor and internal resource allocation suggests that the gender composition of managers makes a difference to how firms are run. We have explored possible mechanisms based on previous studies, but determining the importance of these mechanisms requires further research.

References

Abadie, Alberto and Guido W. Imbens (2006) "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, Vol. 74, No. 1, pp. 235–267.

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson (2014) "Democracy Does Cause Growth," Working Paper 20004, National Bureau of Economic Research (NBER).

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer (2015) "Identification Properties of Recent Production Function Estimators," *Econometrica*, Vol. 83, No. 6, pp. 2411–2451.

Ahern, Kenneth R. and Amy K. Dittmar (2012) "The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation," *The Quarterly Journal of Economics*, Vol. 127, No. 1, pp. 137–197.

Albrecht, James, Anders Bjorklund, and Susan Vroman (2003) "Is There a Glass Ceiling in Sweden?" *Journal of Labor Economics*, Vol. 21, No. 1, pp. 145–177.

Albrecht, James, Peter Skogman Thoursie, and Susan Vroman (2015) "Parental Leave and the Glass Ceiling in Sweden," Working Paper Series 2015:4, Institute for Evaluation of Labour Market and Education Policy (IFAU).

Alesina, Alberto F. and Paola Giuliano (2009) "Preferences for Redistribution," Working Paper 14825, National Bureau of Economic Research (NBER).

Alesina, Alberto and Eliana La Ferrara (2005) "Ethnic Diversity and Eco-

nomic Performance," *Journal of Economic Literature*, Vol. 43, No. 3, pp. 762–800.

Ali, Muhammad (2016) "Impact of Gender-focused Human Resource Management on Performance: The Mediating Effects of Gender Diversity," *Australian Journal of Management*, Vol. 41, No. 2, pp. 376–397.

Altonji, Joseph G. and Rebecca M. Blank (1999) "Race and Gender in the Labor Market," in Orley Ashenfelter and David Card eds. *Handbook of Labor Economics*, Vol. 3: Elsevier, Chap. 48, pp. 3143–3259.

Azmat, Ghazala and Barbara Petrongolo (2014) "Gender and the Labor Market: What Have we Learned from Field and Lab Experiments?" *Labour Economics*, Vol. 30, pp. 32–40.

Bassett-Jones, Nigel (2005) "The Paradox of Diversity Management, Creativity and Innovation," *Creativity and innovation management*, Vol. 14, No. 2, pp. 169–175.

Becker, Gary S. (1971) *The Economics of Discrimination*: University of Chicago press.

Berliant, Marcus and Masahisa Fujita (2008) "Knowledge Creation as a Square Dance on the Hilbert Cube," *International Economic Review*, Vol. 49, No. 4, pp. 1251–1295.

Bertrand, Marianne (2011) "New Perspectives on Gender," in Orley Ashenfelter and David Card eds. *Handbook of Labor Economics*, Vol. 4: Elsevier, Chap. 17, pp. 1543–1590.

Borghans, Lex, Bas Ter Weel, and Bruce A. Weinberg (2006) "People People: Social Capital and the Labor-Market Outcomes of Underrepresented Groups," Working Paper 11985, National Bureau of Economic Research (NBER).

Carpenter, Mason A. (2002) "The Implications of Strategy and Social Context for the Relationship between Top Management Team Heterogeneity and Firm Performance," *Strategic Management Journal*, Vol. 23, No. 3, pp. 275–284.

Carter, David A., Frank D'Souza, Betty J. Simkins, and W. Gary Simpson (2010) "The Gender and Ethnic Diversity of US Boards and Board Committees and Firm Financial Performance," *Corporate Governance: An International Review*, Vol. 18, No. 5, pp. 396–414.

Chabé-Ferret, Sylvain (2014) "Bias of Causal Effect Estimators using Pre-

policy Outcomes," mimeo.

———— (2015) "Analysis of the Bias of Matching and Difference-in-Difference under Alternative Earnings and Selection Processes," *Journal of Econometrics*, Vol. 185, No. 1, pp. 110–123.

Croson, Rachel and Uri Gneezy (2009) "Gender Differences in Preferences," *Journal of Economic Literature*, Vol. 47, No. 2, pp. 448–474.

Datta Gupta, Nabanita, Anders Poulsen, and Marie Claire Villeval (2005) "Male and Female Competitive Behavior: Experimental Evidence," IZA Discussion Papers 1833, Institute for the Study of Labor (IZA).

Dohmen, Thomas and Armin Falk (2011) "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender," *American Economic Review*, Vol. 101, No. 2, pp. 556–590.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011) "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, Vol. 9, No. 3, pp. 522–550.

Eckbo, B. Espen, Knut Nygaard, and Karin S. Thorburn (2016) "Does Gender-Balancing the Board Reduce Firm Value?" CEPR Discussion Papers 11176, Center for Economic Policy Research (CEPR).

Eckel, Catherine C. and Philip J. Grossman (2008a) "Differences in the Economic Decisions of Men and Women: Experimental Evidence," in Charles R. Plott and Vernon L. Smith eds. *Handbook of Experimental Economics Results*, Vol. 1: Elsevier, Chap. 57, pp. 509–519.

———— (2008b) "Men, Women and Risk Aversion: Experimental Evidence," in Charles R. Plott and Vernon L. Smith eds. *Handbook of Experimental Economics Results*, Vol. 1: Elsevier, Chap. 113, pp. 1061–1073.

Farrell, Kathleen A. and Philip L. Hersch (2005) "Additions to Corporate Boards: the Effect of Gender," *Journal of Corporate Finance*, Vol. 11, No. 1, pp. 85–106.

Fortin, Nicole M. (2008) "The Gender Wage Gap among Young Adults in the United States: the Importance of Money versus People," *Journal of Human Resources*, Vol. 43, No. 4, pp. 884–918.

Funk, Patricia and Christina Gathmann (2015) "Gender Gaps in Policy Making: Evidence from Direct Democracy in Switzerland," *Economic Policy*, Vol. 30, No. 81, pp. 141–181.

42

Hambrick, Donald C., Theresa Seung Cho, and Ming-Jer Chen (1996) "The Influence of Top Management Team Heterogeneity on Firms' Competitive Moves," *Administrative Science Quarterly*, Vol. 41, No. 4, pp. 659–684.

Holland, Paul W. (1986) "Statistics and Causal Inference," *Journal of the American Statistical Association*, Vol. 81, No. 396, pp. 945–960.

Hong, Lu and Scott E. Page (2001) "Problem Solving by Heterogeneous Agents," *Journal of Economic Theory*, Vol. 97, No. 1, pp. 123–163.

——— (2004) "Groups of Diverse Problem Solvers can Outperform Groups of High-Ability Problem Solvers," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, No. 46, pp. 16385–16389.

Imbens, Guido W. (2015) "Matching Methods in Practice: Three Examples," *Journal of Human Resources*, Vol. 50, No. 2, pp. 373–419.

Imbens, Guido W. and Donald B. Rubin (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences*: Cambridge University Press.

Imbens, Guido W. and Jeffrey M. Wooldridge (2009) "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, Vol. 47, No. 1, pp. 5–86.

Jackson, Susan E., Karen E. May, and Kristina Whitney (1995) "Understanding the Dynamics of Diversity in Decision-Making Teams," in Richard A. Guzzo and Eduardo Salas eds. *Team Effectiveness and Decision Making in Organizations*: Jossey-Bass, Chap. 7, pp. 204–261.

Koenig, Anne M., Alice H. Eagly, Abigail A. Mitchell, and Tiina Ristikari (2011) "Are Leader Stereotypes Masculine? A Meta-Analysis of Three Research Paradigms," *Psychological Bulletin*, Vol. 137, No. 4, pp. 616–642.

Lazear, Edward P. (1999) "Globalisation and the Market for Team-Mates," *The Economic Journal*, Vol. 109, No. 454, pp. 15–40.

Leuven, Edwin and Barbara Sianesi (2003) "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing," Statistical software components, Boston College Department of Economics, https://ideas.repec.org/c/boc/bocode/s432001.html.

Niederle, Muriel and Lise Vesterlund (2007) "Do Women Shy Away from

Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics*, pp. 1067–1101.

Nkomo, Stella M. and T. Cox (1999) "Diverse Identities in Organizations," in Stewart R. Clegg, Cynthia Hardy, and Walter R. Nord eds. *Managing Organizations: Current Issues*: Sage, Chap. 5, pp. 88–106.

OECD (2012) *Closing the Gender Gap: Act Now*: OECD Publishing.

Olivetti, Claudia and Barbara Petrongolo (2016) "The Evolution of the Gender Gap in Industrialized Countries," *Annual Review of Economics*, Vol. 8, pp. 405–434.

Osborne, Evan (2000) "The Deceptively Simple Economics," *Journal of Labor Research*, Vol. 21, No. 3, pp. 463–475.

Pletzer, Jan Luca, Romina Nikolova, Karina Karolina Kedzior, and Sven Constantin Voelpel (2015) "Does Gender Matter? Female Representation on Corporate Boards and Firm Financial Performance - A Meta-Analysis," *PloS One*, Vol. 10, No. 6, p. e0130005.

Post, Corinne and Kris Byron (2015) "Women on Boards and Firm Financial Performance: A meta-analysis," *Academy of Management Journal*, Vol. 58, No. 5, pp. 1546–1571.

Rauch, James E. and Alessandra Casella (2003) "Overcoming Informational Barriers to International Resource Allocation: Prices and Ties," *The Economic Journal*, Vol. 113, No. 484, pp. 21–42.

Robinson, Gail and Kathleen Dechant (1997) "Building a Business Case for Diversity," *The Academy of Management Executive*, Vol. 11, No. 3, pp. 21–31.

Shen, Jie, Ashok Chanda, Brian D'netto, and Manjit Monga (2009) "Managing Diversity through Human Resource Management: An International Perspective and Conceptual Framework," *The International Journal of Human Resource Management*, Vol. 20, No. 2, pp. 235–251.

Spence, Michael (1973) "Job Market Signaling," *The Quarterly Journal of Economics*, Vol. 87, No. 3, pp. 355–374.

——— (2002) "Signaling in Retrospect and the Informational Structure of Markets," *American Economic Review*, Vol. 92, No. 3, pp. 434–459.

Statistics Sweden (1998) "SSYK 96 - Swedish Standard Classification of Occupations 1996 (In Swedish: Standard för svensk yrkesklassificering 1996)," Reports on Statistical Co-ordination for the Official Statistics of

Sweden 1998-3.

——— (2011a) "Integrated Database for Labour Market Research (In Swedish: Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier (LISA) 1990–2009)," Labour and Education Statistics 2011-4.

——— (2011b) "The Swedish Occupational Register (In Swedish: Yrkesregistret med yrkesstatistik - En beskrivning av innehåll och kvalitet)," Labour and Education Statistics 2011-5.

——— (2014) "Women and men in Sweden 2014: Facts and figures."

——— (2016) "Integrated Database for Labour Market Research (In Swedish: Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier (LISA) 1990–2013)," Labour and Education Statistics 2016-1.

Taymaz, Erol (2005) "Are Small Firms Really Less Productive?" *Small Business Economics*, Vol. 25, No. 5, pp. 429–445.

Wahlberg, Roger (2010) "The Gender Wage Gap across the Wage Distribution in the Private and Public Sectors in Sweden," *Applied Economics Letters*, Vol. 17, No. 15, pp. 1465–1468.

Yang, Chih-Hai and Ku-Hsieh Chen (2009) "Are Small Firms Less Efficient?" *Small Business Economics*, Vol. 32, No. 4, pp. 375–395.

Appendices

A   Imbens and Rubin's algorithm for covariate selection

We explain the stepwise procedure for covariate selection proposed by Imbens and Rubin (2015) and Imbens (2015). We firstly classify potentially important covariates into "basic" and "additional" covariates. The basic covariates include those variables that are a priori regarded as important for explaining treatment assignment or as likely related to the outcomes. We denote the number of the basic covariates and that of all potentially important covariates as K_B and K, respectively ($K_B \leq K$).

In the first stage, we select all of the basic covariates to be included in the probit model, regardless of their statistical association with the treatment indicator. We have K_B covatiates in the model in this stage.

In the second stage, we decide which of the remaining $K - K_B$ covariates will also enter the model linearly. We estimate a probit model with the covariates already included plus one of the remaining covariates, and run a likelihood ratio test to assess the null hypothesis that the coefficient of the additionally included covariate is zero. We repeat this for each of the remaining covariates. If at least one of the likelihood ratio test statistics is greater than the threshold $C_L = 1$, we add the covariate with the largest likelihood ratio statistic. Now we have $K + 1$ covariates included in the model, and we next decide which of the remaining $K - K_B - 1$ covariates will also enter the model linearly by running new $K - K_B - 1$ likelihood ratio tests. That is, we estimate a probit model with the already included covariates plus each of the remaining covariates. We again add the covariate with the largest likelihood ratio statistic, as long as the statistic is larger than the threshold $C_L = 1$. We repeat this process and add one covariate to the model at a time until the largest likelihood ratio statistic is below the threshold or there are no remaining covariates.

Let us assume that the number of the covariates in the model at this stage is K_L, including the initially included K_B basic covariates. In the third stage, we decide which of the interaction and quadratic terms of the

K_L covariates should be included in the model. The number of interaction and quadratic terms involving the K_L covariates is $K_L(K_L + 1)/2$, and we assess each of these one at a time. Exactly as in the second stage, we estimate a probit model with the covariates already included plus one of the $K_L(K_L + 1)/2$ second-order terms and run a likelihood ratio test to assess the null hypothesis that the coefficient of the additionally included term is zero. If at least one of the likelihood ratio test statistics is greater than the threshold $C_Q = 2.71$, we add the term with the largest likelihood ratio statistic. We repeat this procedure for the remaining $K_L(K_L + 1)/2 - 1$ second-order terms by running new $K_L(K_L + 1)/2 - 1$ likelihood ratio tests to decide which of them should be included in the model. In this way, we keep adding the term with the largest likelihood ratio statistic as long as this statistic is larger than the threshold $C_Q = 2.71$. When the largest likelihood ratio statistic is below the threshold or there are no remaining terms, we stop and estimate the propensity score using the covariates we have selected. The threshold values $C_L = 1$ and $C_Q = 2.71$ we adopted for the likelihood ratio tests are those recommended by Imbens and Rubin (2015) and Imbens (2015). They correspond to z-statistics of 1 and 1.645 respectively, implying that the threshold is generous for linear terms but relatively strict for the second-order terms.

We introduced a minor modification to this procedure. To account for the pre-treatment dynamism, we added the one-period lagged values of the basic covariates to additional covariates. But to avoid situations in which these lagged values crowd out other potentially important covariates, the validity of the lagged covariates was tested by likelihood ratio tests only after the other additional covariates were tested in the second stages. Similarly, the validity of the second-order terms involving the lagged covariates was tested at the end of the third stage,

B Normalized difference

The normalized difference is computed as

$$\Delta_x = \frac{\bar{x}^{tr} - \bar{x}^{cnt}}{\sqrt{[(s^{tr})^2 + (s^{cnt})^2]/2}}, \tag{B.1}$$

where $\bar{x}^{tr}$ and $(s^{tr})^2$ stand for the average and variance of covariate x for the treated observations respectively, while $\bar{x}^{cnt}$ and $(s^{cnt})^2$ stand for the counterparts for the control observations.

$\bar{x}^{tr}$ and $(s^{tr})^2$ are computed as follows:

$$\bar{x}^{tr} = \frac{1}{N^{tr}} \sum_{i:D_i=1} x_i, \qquad (s^{tr})^2 = \frac{1}{N^{tr}-1} \sum_{i:D_i=1} (x_i - \bar{x}^{tr})^2. \tag{B.2}$$

In contrast, the formula for $\bar{x}^{cnt}$ and $(s^{cnt})^2$ depend on whether the matching adjustment is introduced or not, because of our focus on the treatment effect on the treated. With an unadjusted sample, they are expressed as

$$\bar{x}^{cnt} = \frac{1}{N^{cnt}} \sum_{i:D_i=0} x_i, \qquad (s^{cnt})^2 = \frac{1}{N^{cnt}-1} \sum_{i:D_i=0} (x_i - \bar{x}^{cnt})^2, \tag{B.3}$$

where N^{tr} and N^{cnt} stands for the numbers of controls and treated respectively. With an adjusted sample, they are expressed as

$$\bar{x}^{cnt} = \frac{1}{N^{tr}} \sum_{i:D_i=1} \left(\frac{1}{H} \sum_{j\in\mathcal{H}_i} x_j \right), \quad (s^{cnt})^2 = \frac{1}{N^{tr}-1} \sum_{i:D_i=1} \left[\left(\frac{1}{H} \sum_{j\in\mathcal{H}_i} x_j \right) - \bar{x}^{cnt} \right]^2. \tag{B.4}$$

where $\mathcal{H}_i$ represents the set of the control observations matched to unit i based on the propensity score. The number of control firms in each $\mathcal{H}_i$ is H.

C   Estimation results

Table C.1: DID-matching estimates (analysis of the replacement effect)

Analysis of the *replacement effect*							Control: N_0 = 11,216	Treated: N_1 = 582
		pre-treatment			under treatment introduction			post-treatment	
Outcome variable (detrended)	time	-2	-1	0	1	2	3	4	5
Total factor productivity (TFP)	ATET	-0.016 *	-0.018 ***	0.000	-0.012	-0.013	-0.006	-0.005	0.007
	s.e.	(0.009)	(0.007)		(0.009)	(0.010)	(0.013)	(0.013)	(0.013)
Labor productivity (LP)	ATET	-0.013	-0.011	0.000	0.010	-0.010	0.012	-0.037	-0.015
	s.e.	(0.046)	(0.042)		(0.043)	(0.042)	(0.041)	(0.043)	(0.044)
ln (Value added)	ATET	0.028	0.025	0.000	0.097 ***	0.155 ***	0.200 ***	0.188 ***	0.210 ***
	s.e.	(0.089)	(0.081)		(0.036)	(0.045)	(0.052)	(0.056)	(0.057)
ln (# of employees)	ATET	0.041	0.036	0.000	0.086 *	0.164 ***	0.186 ***	0.221 ***	0.221 ***
	s.e.	(0.050)	(0.043)		(0.049)	(0.037)	(0.044)	(0.042)	(0.047)
ln (Capital stocks)	ATET	-0.434	-0.343	0.000	-0.302	-0.287	-0.162	0.008	0.195
	s.e.	(0.386)	(0.334)		(0.484)	(0.492)	(0.475)	(0.498)	(0.498)
ln (Labor costs per employee)	ATET	0.015	0.009	0.000	-0.022	0.006	0.008	0.013	0.019
	s.e.	(0.017)	(0.015)		(0.035)	(0.019)	(0.018)	(0.021)	(0.019)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The original estimates and the heteroskedasticity-consistent standard errors of the left part in Figure 4. The treatment group consists of the observations in which the firm *replaces* a male manager with a woman, whereas the control group consists of the observations in which the numbers of male and female managers are unchanged. The treatment is introduced during $0 < t < 4$. The estimates in $t \geq 4$ are our main focus. While the estimates in the period $t < 0$ serve as measures of placebo tests, the estimates in $0 < t < 4$ show the transition during the treatment introduction. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

Table C.2: DID-matching estimates (analysis of the appointment effect)

Analysis of the *appointment effect*							Control: N_0 = 3,801	Treated: N_1 = 1,775
		pre-treatment			under treatment introduction			post-treatment	
Outcome variable (detrended)	time	-2	-1	0	1	2	3	4	5
Total factor productivity (TFP)	ATET	0.005	0.004	0.000	0.001	-0.003	-0.001	0.000	-0.002
	s.e.	(0.006)	(0.004)		(0.004)	(0.005)	(0.006)	(0.007)	(0.007)
Labor productivity (LP)	ATET	-0.005	-0.005	0.000	-0.020	-0.015	-0.011	-0.016	-0.020
	s.e.	(0.015)	(0.014)		(0.013)	(0.015)	(0.015)	(0.017)	(0.016)
ln (Value added)	ATET	-0.013	-0.009	0.000	-0.034 **	-0.011	-0.003	-0.009	-0.018
	s.e.	(0.018)	(0.015)		(0.016)	(0.019)	(0.021)	(0.023)	(0.024)
ln (# of employees)	ATET	-0.011	-0.007	0.000	-0.017	0.003	0.007	0.007	0.002
	s.e.	(0.015)	(0.010)		(0.010)	(0.016)	(0.018)	(0.019)	(0.021)
ln (Capital stocks)	ATET	-0.016	0.024	0.000	-0.079	-0.145 *	-0.103	-0.070	-0.059
	s.e.	(0.085)	(0.062)		(0.069)	(0.079)	(0.088)	(0.096)	(0.122)
ln (Labor costs per employee)	ATET	0.022 **	0.010	0.000	-0.008	0.014	0.009	0.003	-0.005
	s.e.	(0.011)	(0.009)		(0.010)	(0.014)	(0.015)	(0.015)	(0.015)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The original estimates and the heteroskedasticity-consistent standard errors of the right part in Figure 4. The treatment group consists of the observations in which the firm *appoints* a female manager to a new managerial position, whereas the control group consists of the observations in which the firm appoints a male manager to a new managerial position. The treatment is introduced during $0 < t < 4$. The estimates in $t \geq 4$ are our main focus. While the estimates in the period $t < 0$ serve as measures of placebo tests, the estimates in $0 < t < 4$ show the transition during the treatment introduction. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

Table C.3: DID-matching estimates (a male manager replaced by a female manager)

Analysis of the *replacement effect* (a male-to-female manager change)		pre-treatment			under treatment introduction			post-treatment	
						Control: $N_0 = 1{,}669$		Treated: $N_1 = 428$	
Outcome variable (detrended)	time	-2	-1	0	1	2	3	4	5
Total factor productivity (TFP)	**ATET**	**-0.011**	**-0.002**	**0.000**	**-0.007**	**-0.003**	**-0.013**	**-0.010**	**-0.005**
	s.e.	(0.009)	(0.007)		(0.009)	(0.010)	(0.011)	(0.013)	(0.013)
Labor productivity (LP)	**ATET**	**-0.025**	**0.007**	**0.000**	**0.033**	**0.026**	**0.025**	**0.007**	**0.022**
	s.e.	(0.026)	(0.021)		(0.022)	(0.028)	(0.027)	(0.031)	(0.031)
ln (Value added)	**ATET**	**-0.018**	**0.015**	**0.000**	**0.063 *****	**0.098 *****	**0.078 ****	**0.082 ****	**0.073 ***
	s.e.	(0.035)	(0.026)		(0.023)	(0.031)	(0.033)	(0.038)	(0.042)
ln (# of employees)	**ATET**	**0.009**	**0.009**	**0.000**	**0.028**	**0.068 *****	**0.049 ***	**0.072 ****	**0.048**
	s.e.	(0.029)	(0.020)		(0.017)	(0.025)	(0.027)	(0.031)	(0.034)
ln (Capital stocks)	**ATET**	**-0.041**	**0.080**	**0.000**	**0.131**	**-0.013**	**0.022**	**-0.052**	**0.066**
	s.e.	(0.124)	(0.099)		(0.093)	(0.152)	(0.139)	(0.154)	(0.180)
ln (Labor costs per employee)	**ATET**	**-0.020**	**0.001**	**0.000**	**0.009**	**0.020**	**0.023**	**0.003**	**0.026**
	s.e.	(0.020)	(0.016)		(0.016)	(0.021)	(0.022)	(0.023)	(0.023)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The original estimates and the heteroskedasticity-consistent standard errors of the left part in Figure 5. The treatment group consists of the observations in which a male manager is replaced by a female manager, whereas the control group consists of the observations in which a manager is replaced by another person of the same gender. In both groups, the other managers are unchanged. The treatment is introduced during $0 < t < 4$. The estimates in $t \geq 4$ are our main focus. While the estimates in the period $t < 0$ serve as measures of placebo tests, the estimates in $0 < t < 4$ show the transition during the treatment introduction. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

Table C.4: DID-matching estimates (a female manager replaced by a male manager)

Analysis of the *replacement effect* (a female-to-male manager change)		pre-treatment			under treatment introduction			post-treatment	
						Control: $N_0 = 1{,}671$		Treated: $N_1 = 342$	
Outcome variable (detrended)	time	-2	-1	0	1	2	3	4	5
Total factor productivity (TFP)	**ATET**	**-0.014**	**-0.012**	**0.000**	**0.000**	**0.007**	**0.006**	**0.010**	**0.002**
	s.e.	(0.011)	(0.008)		(0.007)	(0.009)	(0.012)	(0.013)	(0.013)
Labor productivity (LP)	**ATET**	**-0.019**	**-0.007**	**0.000**	**-0.028**	**-0.026**	**-0.050**	**-0.029**	**-0.058 ***
	s.e.	(0.028)	(0.028)		(0.023)	(0.028)	(0.036)	(0.030)	(0.035)
ln (Value added)	**ATET**	**-0.011**	**-0.002**	**0.000**	**-0.036**	**-0.002**	**-0.032**	**-0.010**	**-0.061**
	s.e.	(0.042)	(0.032)		(0.023)	(0.032)	(0.043)	(0.047)	(0.053)
ln (# of employees)	**ATET**	**0.011**	**0.009**	**0.000**	**-0.008**	**0.022**	**0.018**	**0.019**	**-0.004**
	s.e.	(0.034)	(0.026)		(0.020)	(0.027)	(0.034)	(0.044)	(0.048)
ln (Capital stocks)	**ATET**	**-0.187**	**-0.071**	**0.000**	**0.180 ****	**0.227 ****	**0.160**	**0.020**	**0.011**
	s.e.	(0.123)	(0.087)		(0.082)	(0.110)	(0.131)	(0.150)	(0.184)
ln (Labor costs per employee)	**ATET**	**-0.009**	**-0.015**	**0.000**	**-0.007**	**-0.021**	**-0.022**	**-0.029**	**-0.027**
	s.e.	(0.021)	(0.025)		(0.022)	(0.025)	(0.027)	(0.025)	(0.026)

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The original estimates and the heteroskedasticity-consistent standard errors of the right part in Figure 5. The treatment group consists of the observations in which a female manager is replaced by a male manager, whereas the control group consists of the observations in which a manager is replaced by another person of the same gender. In both groups, the other managers are unchanged. The treatment is introduced during $0 < t < 4$. The estimates in $t \geq 4$ are our main focus. While the estimates in the period $t < 0$ serve as measures of placebo tests, the estimates in $0 < t < 4$ show the transition during the treatment introduction. The standard errors are robust for heteroskedasticity and computed according to Abadie and Imbens (2006).

Online appendices

D   Estimation of production function

We define a production function as

$$\ln V_{i,t} = \beta_l \ln L_{i,t} + \beta_k \ln K_{i,t} + \omega_{i,t} + \epsilon_{i,t}, \tag{D.1}$$

or,

$$v_{i,t} = \beta_l l_{i,t} + \beta_k k_{i,t} + \omega_{i,t} + \epsilon_{i,t}, \tag{D.2}$$

where the lower-case characters stand for the logs, $\omega_{i,t}$ is a productivity component and $\epsilon_{i,t}$ is an i.i.d. error term. Our challenge is to consistently estimate the production function (D.2) and to obtain the term $\omega_{i,t}$.

The coefficients β_l and β_k are most likely to vary across sectors. They are also likely to depend on firm size: for example, the degree of scale economy, which can be expressed as the sum $\beta_L + \beta_K$, is expected to decrease as firm size grows. We therefore adapt a more flexible functional form – a translog function

$$v_{i,t} = \beta_l l_{i,t} + \beta_{ll} l_{i,t}^2 + \beta_k k_{i,t} + \beta_{kk} k_{i,t}^2 + \beta_{lk} l_{i,t} k_{i,t} + \omega_{i,t} + \epsilon_{i,t}. \tag{D.3}$$

For estimation, we adopt the approach proposed by Ackerberg et al. (2015). We perform the following procedure for each two-digit service industry.

We start by assuming that capital k_t is determined in the previous period $t-1$ through investment chosen in $t-1$, while labor l_t can be chosen in t, $t-1$ or somewhere in between. We also assume that the demand function of intermediate input $m_{i,t}$ is a function of capital, labor and productivity and is given by $m_{i,t} = \tilde{f}_t(k_{i,t}, l_{i,t}, \omega_{i,t})$, where $\tilde{f}_t$ is strictly increasing in $\omega_{i,t}$. We then invert this function as $\omega_{i,t} = \tilde{f}_t^{-1}(k_{i,t}, l_{i,t}, m_{i,t})$ and replace $\omega_{i,t}$ in the production function. We further assume $\omega_{i,t}$ to follow

a first-order Markov process:

$$\omega_{i,t} = \mathbf{E}[\omega_{i,t}|\omega_{i,t-1}] + \xi_{i,t} \tag{D.4}$$

$$= \Psi(\omega_{i,t-1}) + \xi_{i,t} \tag{D.5}$$

In theory, we can estimate the function Ψ nonparametrically. However, for convenience of estimation, we choose a linear autoregressive function following Ackerberg et al. (2015):

$$\omega_{i,t} = \psi\omega_{i,t-1} + \xi_{i,t}. \tag{D.6}$$

In the first stage, we separate the i.i.d. error term $\epsilon_{i,t}$ by estimating

$$v_{i,t} = \tilde{\Theta}_t(k_{i,t}, l_{i,t}, m_{i,t}) + \epsilon_{i,t} \tag{D.7}$$

where $\tilde{\Theta}_t(k_{i,t}, l_{i,t}, m_{i,t}) = \beta_l l_{i,t} + \beta_{ll} l_{i,t}^2 + \beta_k k_{i,t} + \beta_{kk} k_{i,t}^2 + \beta_{lk} l_{i,t} k_{i,t} + \tilde{f}_t^{-1}(k_{i,t}, l_{i,t}, m_{i,t})$. For specification of the function $\tilde{\Theta}_t(k_{i,t}, l_{i,t}, m_{i,t})$, we adopt a third-order polynomial function of $k_{i,t}$, $l_{i,t}$ and $m_{i,t}$ combined with a linear function of dummies for three-digit industrial sectors and years.

In the second stage, we estimate the coefficients β_l, β_{ll}, β_k, β_{kk}, β_{lk} and ψ using the following moment conditions:

$$\mathbf{E}\left[\epsilon_{i,t} + \xi_{i,t} \begin{pmatrix} l_{i,t-1} \\ l_{i,t-1}^2 \\ k_{i,t} \\ k_{i,t}^2 \\ l_{i,t-1}k_{i,t} \\ \tilde{\Theta}_{t-1} \end{pmatrix}\right] = \mathbf{0} \tag{D.8}$$

where $\epsilon_{i,t} + \xi_{i,t}$ is replaced by

$$\epsilon_{i,t} + \xi_{i,t} = (v_t - \beta_l l_{i,t} - \beta_{ll} l_{i,t}^2 - \beta_k k_{i,t} - \beta_{kk} k_{i,t}^2 - \beta_{lk} l_{i,t} k_{i,t})$$
$$- \psi \left[\tilde{\Theta}_{t-1}(k_{i,t-1}, l_{i,t-1}, m_{i,t-1}) \right.$$
$$\left. - \beta_l l_{i,t-1} - \beta_{ll} l_{i,t-1}^2 - \beta_k k_{i,t-1} - \beta_{kk} k_{i,t-1}^2 - \beta_{lk} l_{i,t-1} k_{i,t-1} \right] \tag{D.9}$$

after rearranging (D.6) and (D.7).

Once we estimate β_l, β_{ll}, β_k, β_{kk}, β_{lk}, we can obtain $\omega_{i,t}$ from (D.3) and define this as the TFP measure.

E Supplementary tables

Table E.1: Propensity score estimation (analysis of the replacement effect)

Analysis of the *replacement effect*				
	Estimation of probit model		Marginal effect at means	
Covariates	coeff.	(s.e.)	coeff.	(s.e.)
Basic covariates				
ln (Value added)	-0.282	(0.185)	0.225 **	(0.110)
ln (# of employees)	-5.097 ***	(1.388)	0.449 ***	(0.099)
ln (Capital stocks)	1.607 ***	(0.537)	0.011	(0.018)
# of all managers	0.383 ***	(0.084)	0.132 ***	(0.042)
# of female managers	-0.153	(0.157)	0.035	(0.127)
Share of female managers	0.060	(0.180)	-0.429 **	(0.195)
Additional covariates				
Share of female employees	2.214 ***	(0.504)	0.488 ***	(0.158)
Share of employees with univ. education	1.365 ***	(0.445)	1.070 ***	(0.350)
ln (Labor costs per employee)	1.586 ***	(0.583)	0.463 ***	(0.143)
Share of employees in age -25 (benchmark)				
Share of employees in age 26-35	9.004	(6.083)	0.466 **	(0.230)
Share of employees in age 36-45	-3.043 ***	(1.174)	0.567 **	(0.227)
Share of employees in age 46-55	---	---	---	---
Share of employees in age 56-65	-4.612 ***	(1.515)	0.019	(0.308)
Share of employees in age 66-	-1.364	(6.773)	4.379 **	(1.863)
Share of managers in age -35 (benchmark)				
Share of managers in age 36-45	-0.227 *	(0.120)	-0.227 *	(0.120)
Share of managers in age 46-55	0.008	(0.161)	-0.198 *	(0.117)
Share of managers in age 56-65	0.463 **	(0.190)	-0.026	(0.144)
Share of managers in age 66-	---	---	---	---
Additional covariates (lagged values of basic covariates)				
ln (Value added) $_{-1}$	0.153	(0.160)	-0.134	(0.089)
ln (# of employees) $_{-1}$	---	---	---	---
ln (Capital stocks) $_{-1}$	---	---	---	---
# of all managers $_{-1}$	-0.147 **	(0.072)	0.065 *	(0.038)
# of female managers $_{-1}$	2.434 **	(1.117)	0.195 **	(0.089)
Share of female managers $_{-1}$	---	---	---	---
Interaction terms				
[ln (Value added)]*[Share of employees in age 26-35]	2.158 ***	(0.553)		
[ln (# of employees)]*[ln (Labor costs per employee)]	0.428 ***	(0.106)		
[ln (# of employees)]*[Share of employees in age 26-35]	0.246	(0.495)		
[ln (# of employees)]*[Share of employees in age 66-]	-4.352 ***	(1.570)		
[ln (Capital stocks)]*[ln (Capital stocks)]	-0.007 ***	(0.002)		
[ln (Capital stocks)]*[ln (Labor costs per employee)]	-0.120 ***	(0.040)		
[ln (Capital stocks)]*[Share of employees in age 36-45]	0.275 ***	(0.085)		
[ln (Capital stocks)]*[Share of employees in age 56-65]	0.352 ***	(0.107)		
[ln (Capital stocks)]*[Share of employees in age 66-]	1.132 ***	(0.436)		
[# of all managers]*[Share of female employees]	-0.129	(0.156)		
[# of all managers]*[Share of employees in age 26-35]	-0.288 **	(0.131)		
[# of all managers]*[Share of managers in age 46-55]	-0.339 ***	(0.108)		
[# of female managers]*[Share of female managers]	0.131	(0.227)		
[# of female managers]*[Share of employees in age 66-]	3.346 ***	(1.103)		
[# of female managers]*[Share of managers in age 56-65]	0.734 **	(0.292)		
[Share of female managers]*[Share of managers in age 56-65]	-2.365 ***	(0.768)		
[Share of female managers]*[# of female managers $_{-1}$]	-0.435 **	(0.195)		
[Share of female employees]*[Share of female employees]	-2.519 ***	(0.494)		
[Share of female employees]*[# of all managers $_{-1}$]	0.321 **	(0.150)		
[Share of employees with univ. ed.]*[Share of employees with univ. ed.]	-1.688 ***	(0.619)		
[ln (Labor costs per employee)]*[Share of employees in age 26-35]	-1.764 **	(0.685)		
[ln (Labor costs per employee)]*[Share of employees in age 26-35]	-0.167 **	(0.085)		
[Share of employees in age 26-35]*[ln (Value added) $_{-1}$]	-1.313 **	(0.553)		
[Share of employees in age 26-35]*[Share of managers in age 56-65]	-1.219 **	(0.466)		
[Share of employees in age 66-]*[Share of employees in age 66-]	-34.630 **	(13.627)		
[Share of managers in age 46-55]*[# of all managers $_{-1}$]	0.213 **	(0.101)		
Dummy variables				
County dummies	Yes			
2-digit sector dummies	Yes			
# of observations: Control: N_0 = 11,216				
Treated: N_1 = 582				

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The treatment group consists of the observations in which the firm *replaces* a male manager by a woman, whereas the control group consists of the observations in which the numbers of male and female managers are unchanged. The probit model is adopted. The stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015) chooses covariates and their interaction terms. The marginal effects are evaluated at the means of the covariates using the delta method.

Table E.2: Propensity score estimation (analysis of the appointment effect)

Analysis of the *appointment effect*				
	Estimation of probit model		**Marginal effect at means**	
Covariates	coeff.	(s.e.)	coeff.	(s.e.)
Basic covariates				
ln (Value added)	-0.011	(0.052)	-0.026	(0.041)
ln (# of employees)	0.024	(0.046)	0.024	(0.046)
ln (Capital stocks)	-0.022	(0.023)	-0.021	(0.016)
# of all managers	-0.020 *	(0.012)	-0.020 *	(0.012)
# of female managers	0.049 **	(0.019)	0.049 **	(0.019)
Share of female managers	-0.102	(0.163)	-0.306 **	(0.131)
Additional covariates				
Share of female employees	2.775 ***	(0.450)	1.613 ***	(0.105)
Share of employees with univ. education	-3.245 ***	(1.242)	0.199 *	(0.118)
ln (Labor costs per employee)				
Share of employees in age -25 (benchmark)				
Share of employees in age 26-35	-1.344 ***	(0.437)	-0.379 **	(0.192)
Share of employees in age 36-45	0.268	(0.191)	0.093	(0.169)
Share of employees in age 46-55	2.177	(2.043)	-0.409 **	(0.201)
Share of employees in age 56-65	-0.394 *	(0.217)	-0.394 *	(0.217)
Share of employees in age 66-				
Share of managers in age -35 (benchmark)				
Share of managers in age 36-45				
Share of managers in age 46-55				
Share of managers in age 56-65	-0.698 **	(0.307)	-0.061	(0.070)
Share of managers in age 66-				
Additional covariates (lagged values of basic covariates)				
ln (Value added) $_{-1}$				
ln (# of employees) $_{-1}$				
ln (Capital stocks) $_{-1}$	0.026	(0.016)	0.026	(0.016)
# of all managers $_{-1}$	0.010	(0.010)	0.010	(0.010)
# of female managers $_{-1}$				
Share of female managers $_{-1}$	0.331 ***	(0.123)	0.331 ***	(0.123)
Interaction terms				
[ln (Value added)]*[Share of employees with univ. edu.]	0.232 ***	(0.074)		
[ln (Value added)]*[Share of employees in age 46-55]	-0.289 **	(0.143)		
[ln (Capital stocks)]*[Share of female employees]	-0.091 ***	(0.030)		
[ln (Capital stocks)]*[Share of employees in age 46-55]	0.152 **	(0.059)		
[ln (Capital stocks)]*[Share of managers in age 56-65]	0.045 **	(0.021)		
[Share of female managers]*[Share of employees in age 36-45]	-0.766 **	(0.388)		
[Share of female employees]*[Share of employees with univ. ed.]	-0.867 **	(0.406)		
[Share of female employees]*[Share of employees in age 26-35]	0.996 **	(0.477)		
[Share of employees in age 26-35]*[Share of employees in age 26-35]	1.052 **	(0.493)		
Dummy variables				
County dummies	Yes			
2-digit sector dummies	Yes			
# of observations: Control: N_0 = 3,801				
Treated: N_1 = 1,775				

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The treatment group consists of the observations in which the firm *appoints* a female manager to a new managerial position, whereas the control group consists of the observations in which the firm appoints a male manager to a new managerial position. The probit model is adopted. The stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015) chooses covariates and their interaction terms. The marginal effects are evaluated at the means of the covariates using the delta method.

Table E.3: Normalized differences (a male manager replaced by a female manager)

Analysis of the *replacement effect* (a male-to-female manager change)	Treated N$_1$ = 428		Control N$_0$ = 1,669		Normalized diff.	
Covariates	Mean	(s.d.)	Mean	(s.d.)	unadjusted	adjusted
Basic covariates						
ln (Value added)	16.573	(1.128)	16.562	(1.162)	0.009	-0.023
ln (# of employees)	3.192	(0.972)	3.153	(1.013)	0.040	0.007
ln (Capital stocks)	14.413	(2.581)	14.229	(3.084)	0.065	0.082
# of all managers	3.846	(4.064)	3.808	(3.812)	0.010	-0.009
# of female managers	0.986	(1.621)	0.834	(1.393)	0.101	-0.066
Share of female managers	0.221	(0.290)	0.220	(0.314)	0.004	-0.085
Additional covariates						
Share of female employees	0.448	(0.229)	0.384	(0.260)	0.263	-0.027
Share of employees with university education	0.343	(0.264)	0.323	(0.258)	0.077	-0.030
ln (Labor costs per employee)	13.147	(0.383)	13.163	(0.380)	-0.042	-0.031
Share of employees in age -25 (benchmark)						
Share of employees in age 26-35	0.285	(0.160)	0.272	(0.161)	0.078	0.054
Share of employees in age 36-45	0.263	(0.135)	0.267	(0.130)	-0.033	0.003
Share of employees in age 46-55	0.182	(0.124)	0.187	(0.127)	-0.043	-0.044
Share of employees in age 56-65	0.120	(0.116)	0.139	(0.125)	-0.160	0.042
Share of employees in age 66-	0.011	(0.030)	0.012	(0.037)	-0.012	0.118
Share of managers in age -35 (benchmark)						
Share of employees in age 36-45	0.323	(0.337)	0.327	(0.339)	-0.012	0.033
Share of employees in age 46-55	0.315	(0.352)	0.284	(0.326)	0.092	-0.002
Share of employees in age 56-65	0.195	(0.305)	0.233	(0.324)	-0.119	-0.054
Share of employees in age 66-	0.016	(0.088)	0.024	(0.116)	-0.074	0.061
Dummy variables						
County dummies (only major countries presented)						
Stockholm county	0.371	(0.484)	0.330	(0.470)	0.087	-0.016
Skåne county	0.124	(0.330)	0.131	(0.338)	-0.022	-0.020
Västra götaland county	0.164	(0.370)	0.177	(0.382)	-0.035	-0.005
Sector dummies (aggregated into the 1-digit level)						
Wholesale. retail trade, hotels, restaurants	0.456	(0.499)	0.516	(0.500)	-0.121	0.023
Transport, communication, financial intermediation	0.082	(0.274)	0.090	(0.287)	-0.031	0.018
Real estate, renting, IT, research, business activities	0.299	(0.458)	0.269	(0.444)	0.067	-0.002
Education, health, social work	0.105	(0.307)	0.087	(0.282)	0.062	-0.040
Other community, social and personal service activities	0.058	(0.235)	0.038	(0.191)	0.097	-0.012
Other major variables						
Propensity score	0.261	(0.104)	0.189	(0.105)	0.679	0.001
Total factor productivity (TFP)	12.863	(0.363)	12.864	(0.322)	-0.005	-0.103
Labor productivity (LP)	13.381	(0.583)	13.409	(0.568)	-0.050	-0.056
# of employees	42.159	(66.080)	40.339	(54.345)	0.030	-0.019

Notes: The treatment group consists of the observationsin which a male manager is replaced by a female manager, whereas the control group consists of the observationsin which a manager is replaced by another person of the same gender. In both groups, the other managers are unchanged. Note that the stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015) chooses covariates and their interaction terms in the estimation of propensity score. Because some covariates are dropped during the procedure, not all of those presented in this table are used in the estimation. The probit model estimated to generate the propensity score for matching is reported in Table E.5.

Table E.4: Normalized differences (a female manager replaced by a male manager)

Analysis of the *replacement effect* (a female-to-male manager change)	Treated N_1 = 342		Control N_0 = 1,671		Normalized diff.	
Covariates	Mean	(s.d.)	Mean	(s.d.)	unadjusted	adjusted
Basic covariates						
ln (Value added)	16.583	(1.229)	16.557	(1.159)	0.022	0.000
ln (# of employees)	3.150	(1.016)	3.147	(1.008)	0.003	-0.057
ln (Capital stocks)	14.247	(2.789)	14.209	(3.062)	0.013	0.019
# of all managers	3.617	(3.701)	3.804	(3.802)	-0.050	-0.077
# of female managers	1.392	(1.658)	0.834	(1.394)	0.364	-0.037
Share of female managers	0.443	(0.359)	0.220	(0.314)	0.663	-0.013
Additional covariates						
Share of female employees	0.486	(0.242)	0.385	(0.259)	0.400	-0.041
Share of employees with university education	0.203	(0.220)	0.169	(0.204)	0.159	-0.016
ln (Labor costs per employee)	13.183	(0.401)	13.169	(0.382)	0.035	0.084
Share of employees in age -25 (benchmark)						
Share of employees in age 26-35	0.292	(0.186)	0.272	(0.161)	0.116	-0.020
Share of employees in age 36-45	0.256	(0.142)	0.267	(0.130)	-0.084	-0.081
Share of employees in age 46-55	0.172	(0.142)	0.187	(0.127)	-0.111	-0.001
Share of employees in age 56-65	0.130	(0.133)	0.140	(0.126)	-0.080	0.054
Share of employees in age 66-	0.015	(0.039)	0.012	(0.037)	0.082	0.108
Share of managers in age -35 (benchmark)						
Share of employees in age 36-45	0.358	(0.339)	0.326	(0.339)	0.094	-0.129
Share of employees in age 46-55	0.270	(0.330)	0.282	(0.326)	-0.037	0.048
Share of employees in age 56-65	0.211	(0.316)	0.235	(0.326)	-0.076	0.106
Share of employees in age 66-	0.029	(0.142)	0.023	(0.116)	0.041	0.075
Dummy variables						
County dummies (only major countries presented)						
Stockholm county	0.406	(0.492)	0.335	(0.472)	0.149	-0.109
Skåne county	0.073	(0.261)	0.131	(0.338)	-0.192	0.000
Västra götaland county	0.199	(0.400)	0.174	(0.379)	0.065	-0.009
Sector dummies (aggregated into the 1-digit level)						
Wholesale. retail trade, hotels, restaurants	0.459	(0.499)	0.515	(0.500)	-0.113	0.061
Transport, communication, financial intermediation	0.099	(0.300)	0.084	(0.278)	0.052	0.012
Real estate, renting, IT, research, business activities	0.287	(0.453)	0.269	(0.443)	0.040	0.003
Education, health, social work	0.099	(0.300)	0.087	(0.282)	0.043	-0.028
Other community, social and personal service activities	0.056	(0.229)	0.045	(0.207)	0.049	-0.112
Other major variables						
Propensity score	0.299	(0.187)	0.144	(0.124)	0.980	0.013
Total factor productivity (TFP)	12.877	(0.357)	12.870	(0.304)	0.021	-0.029
Labor productivity (LP)	13.433	(0.653)	13.410	(0.565)	0.038	0.091
# of employees	45.427	(107.309)	39.808	(53.121)	0.066	0.050

Notes: The treatment group consists of the observations in which a female manager is replaced by a male manager, whereas the control group consists of the observations in which a manager is replaced by another person of the same gender. In both groups, the other managers are unchanged. Note that the stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015) chooses covariates and their interaction terms in the estimation of propensity score. Because some covariates are dropped during the procedure, not all of those presented in this table are used in the estimation. The probit model estimated to generate the propensity score for matching is reported in Table E.6.

Table E.5: Propensity score estimation (a male manager replaced by a female manager)

Analysis of the *replacement effect* (a male-to-female manager change)				
	Estimation of probit model		**Marginal effect at means**	
Covariates	coeff.	(s.e.)	coeff.	(s.e.)
Basic covariates				
ln (Value added)	0.023	(0.068)	-0.016	(0.068)
ln (# of employees)	-0.010	(0.073)	-0.010	(0.073)
ln (Capital stocks)	0.013	(0.016)	0.028 *	(0.016)
# of all managers	-0.014	(0.015)	-0.014	(0.015)
# of female managers	0.044	(0.042)	0.069	(0.043)
Share of female managers	0.359	(0.372)	-0.201	(0.200)
Additional covariates				
Share of female employees	2.058 ***	(0.571)	0.679 ***	(0.196)
Share of employees with univ. education	---	---	---	---
ln (Labor costs per employee)	---	---	---	---
Share of employees in age -25 (benchmark)				
Share of employees in age 26-35	---	---	---	---
Share of employees in age 36-45	---	---	---	---
Share of employees in age 46-55	---	---	---	---
Share of employees in age 56-65	-0.513 *	(0.299)	-0.513 *	(0.299)
Share of employees in age 66-	0.175 *	(0.099)	0.175 *	(0.099)
Share of managers in age -35 (benchmark)				
Share of managers in age 36-45	---	---	---	---
Share of managers in age 46-55	---	---	---	---
Share of managers in age 56-65	---	---	---	---
Share of managers in age 66-	16.534 **	(7.076)	-1.086 *	(0.658)
Additional covariates (lagged values of basic covariates)				
ln (Value added) $_{-1}$	---	---	---	---
ln (# of employees) $_{-1}$	---	---	---	---
ln (Capital stocks) $_{-1}$	---	---	---	---
# of all managers $_{-1}$	---	---	---	---
# of female managers $_{-1}$	---	---	---	---
Share of female managers $_{-1}$	---	---	---	---
Interaction terms				
[ln (Value added)]*[Share of managers in age 66-]	-1.782 ***	(0.570)		
[ln (Capital stocks)]*[Share of managers in age 66-]	0.681 **	(0.287)		
[# of female managers]*[Share of managers in age 66-]	1.115 *	(0.619)		
[Share of female managers]*[Share of female employees]	-1.410 **	(0.572)		
[Share of female employees]*[Share of female employees]	-1.431 **	(0.672)		
[Share of female employees]*[Share of managers in age 66-]	3.062	(1.888)		
Dummy variables				
County dummies	Yes			
2-digit sector dummies	Yes			
# of observations: Control: N_0 = 1,669				
Treated: N_1 = 428				

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The treatment group consists of the observationsin which a male manager is replaced by a female manager, whereas the control group consists of the observations in which a manager is replaced by another person of the same gender. In both groups, the other managers are unchanged. The probit model is adopted. The stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015) chooses covariates and their interaction terms. The marginal effects are evaluated at the means of the covariates using the delta method.

Table E.6: Propensity score estimation (a female manager replaced by a male manager)

Analysis of the *replacement effect* (a female-to-male manager change)

Covariates	Estimation of probit model		Marginal effect at means	
	coeff.	(s.e.)	coeff.	(s.e.)
Basic covariates				
ln (Value added)	0.054	(0.145)	0.147	(0.140)
ln (# of employees)	-0.632	(0.791)	-0.194	(0.192)
ln (Capital stocks)	0.145 ***	(0.047)	0.024	(0.018)
# of all managers	-0.140 ***	(0.037)	-0.140 ***	(0.037)
# of female managers	-0.824 **	(0.348)	0.243 **	(0.097)
Share of female managers	2.314 ***	(0.368)	0.612 ***	(0.206)
Additional covariates				
Share of female employees	---	---	---	---
Share of employees with univ. education	6.540	(6.427)	0.427 *	(0.256)
ln (Labor costs per employee)	0.393 *	(0.202)	0.195	(0.164)
Share of employees in age -25 (benchmark)				
Share of employees in age 26-35	-1.571 **	(0.691)	-0.080	(0.271)
Share of employees in age 36-45	---	---	---	---
Share of employees in age 46-55	---	---	---	---
Share of employees in age 56-65	---	---	---	---
Share of employees in age 66-	-7.232	(5.888)	3.387 ***	(1.216)
Share of managers in age -35 (benchmark)			---	---
Share of managers in age 36-45	0.833 **	(0.402)	0.410 ***	(0.155)
Share of managers in age 46-55	---	---	---	---
Share of managers in age 56-65	---	---	---	---
Share of managers in age 66-	---	---	---	---
Additional covariates (lagged values of basic covariates)				
ln (Value added) $_{-1}$	-0.236 *	(0.121)	-0.236 *	(0.121)
ln (# of employees) $_{-1}$	1.622 **	(0.786)	0.384 **	(0.171)
ln (Capital stocks) $_{-1}$	---	---	---	---
# of all managers $_{-1}$	0.038	(0.031)	0.038	(0.031)
# of female managers $_{-1}$	0.778 **	(0.316)	0.020	(0.066)
Share of female managers $_{-1}$	---	---	---	---
Interaction terms				
[ln (Value added)]*[Share of employees with univ. ed.]	0.531 ***	(0.169)		
[ln (# of employees)]*[ln (Capital stocks)]	0.038	(0.053)		
[ln (# of employees)]*[Share of female managers]	-0.389 ***	(0.130)		
[ln (Capital stocks)]*[# of female managers]	0.082 ***	(0.021)		
[ln (Capital stocks)]*[Share of managers in age 66-]	0.576	(0.410)		
[ln (Capital stocks)]*[ln (# of employees) $_{-1}$]	-0.087	(0.054)		
[ln (Capital stocks)]*[# of female managers $_{-1}$]	-0.057 ***	(0.020)		
[# of female managers]*[Share of female managers]	-0.513 ***	(0.125)		
[# of female managers]*[Share of managers in age 66-]	2.613 **	(1.208)		
[Share of employees with univ. education]*[ln (Labor costs per employee)]	-1.131 **	(0.557)		
[Share of employees in age 26-35]*[Share of employees in age 26-35]	2.712 ***	(0.950)		
[Share of managers in age 36-45]*[Share of managers in age 36-45]	-0.879 **	(0.413)		
[Share of managers in age 36-45]*[# of female managers $_{-1}$]	0.174 *	(0.102)		
Dummy variables				
County dummies	Yes			
2-digit sector dummies	Yes			
# of observations: Control: N_0 = 1,671				
Treated: N_1 = 342				

Notes: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. The treatment group consists of the observations in which a female manager is replaced by a male manager, whereas the control group consists of the observations in which a manager is replaced by another person of the same gender. In both groups, the other managers are unchanged. The probit model is adopted. The stepwise procedure proposed by Imbens (2015) and Imbens and Rubin (2015) chooses covariates and their interaction terms. The marginal effects are evaluated at the means of the covariates using the delta method.