Managers and Productivity in the Public Sector

Alessandra Fenizia*

Abstract

This paper studies the productivity impacts of managers in the public sector using novel administrative data containing an output-based measure of productivity of public offices. Exploiting the rotation of managers across sites, I find that a one standard deviation better manager increases office productivity by 10%. These gains are driven primarily by the exit of older workers who retire when more productive managers take over. Empowering managers to directly change payrolls may generate large benefits to efficiency. Absent such civil service reforms, I use these estimates to evaluate the optimal allocation of managers to offices. I find that assigning better managers to the largest and most productive offices would increase output by at least 6.9%.

Key Words: Managers, Productivity, Public Sector, Effectiveness.

I Introduction

Public sector managers are the cornerstone of modern bureaucracies. They oversee day-to-day operations of complex public organizations and supervise policy implementation. The public sector represents a large share of modern economies and is not disciplined by economic forces of competition. As such, manager effectiveness may have important consequences for the performance of government agencies and ultimately citizens' welfare. For instance, delayed unemployment insurance benefit payments can aggravate hardship of the newly unemployed, and longer processing times for disability insurance claims can directly reduce employment and earnings for multiple years (Autor et al., 2015). However, we know little about the extent to which differences in manager quality ultimately impact public service provision.

*Corresponding author: Alessandra Fenizia, George Washington University, Department of Economic, G Street NW, Room #364, Washington, DC 20052, (e-mail: afenizia@gwu.edu). This paper is based on the main chapter of my dissertation that I completed at UC Berkeley. I am particularly indebted to my advisors David Card, Patrick Kline, and Hilary Hoynes for their invaluable support and guidance throughout the process. I am grateful to Eric Avis, Tito Boeri, Giulia Bovini, Luca Citino, Ernesto Dal Bó, Fred Finan, Michela Giorcelli, Zarek Brot-Goldberg, Daniel Haanwinckel, Erin Kelly, Nicholas Li, John Loeser, Juliana Londoño-Vélez, Isabela Manelici, Matteo Paradisi, Francisco Perez-Gonzalez, Daniel Rogger, Evan Rose, Raffaele Saggio, Yotam Shem-tov, David Silver, Bryan Stuart, Jose Vasquez, Mauricio Ulate, Christopher Walters, Guo Xu, and Danny Yagan for helpful suggestions. My special thanks goes to all the members of the administative and technical staff of the VisitINPS program who provided invaluable support and help. The findings and conclusions expressed are solely those of the author and do not represent the views of INPS.

On the one hand, managers may not be able to affect the performance of their organizations because they lack many of the tools available to private sector firms (e.g., firing, promotions, incentive-pay schemes). In most countries, public sector workers enjoy strong job security and often receive promotions and pay raises that depend on seniority rather than individual performance. On the other hand, public sector managers may play a particularly important role precisely because of the lack of other tools to motivate their workers.

One reason why little is known about the effectiveness of public sector managers is that is notoriously hard to objectively measure the performance of government agencies. Whereas profits are the principal objective in the private sector, government agencies have many obligations, and prioritizing them is subjective. A set of recent studies has made progress by measuring individual managerial practices, qualitative policies and procedures that are thought to be associated with well-run organizations, and has established that these measures positively correlate with public service delivery (Tsai et al., 2015; Bloom et al., 2015; Rasul and Rogger, 2018). Yet, it is unclear how to translate these correlations into quantitative measures of the causal impact of managers.

This paper studies the productivity impacts of managers in the public sector using novel data from the Italian Social Security Agency (*Istituto Nazionale di Previdenza Sociale* — INPS, hereafter). INPS administers applications for unemployment insurance, disability insurance, pensions, subsidies to the poor and other welfare and insurance programs. A key innovation is the use of an output-based measure of productivity of public offices constructed using detailed administrative quarterly data on both output—measured by a (complexity-weighted) standardized index of claims processed by the office— and on full-time equivalent workers assigned to the office. This is an ideal setting to isolate the contribution of managers to office performance because all sites are subject to the same rules, workers produce a homogeneous product, and there are virtually no differences in physical capital across offices. An additional advantage of my setting is that I do not need to rely on manager wages to infer their productivity (Eeckhout and Kircher, 2011).

I begin my analysis by documenting large variation in productivity across offices within INPS, dispersion that is not fully explained by regional differences that typically characterize the Italian economy. I then use a two-way fixed effects model to decompose log productivity into the components due to office characteristics, manager effects, and time effects. A simple model with additive office and manager components may raise two concerns. First, managers could be assigned to offices on the basis of unobserved factors that determine their comparative advantage. I test for match-driven sorting and find no evidence of comparative advantage-based mobility. Second, manager rotation might be correlated with office-specific trends. I find no evidence of sorting based on trends.

Using bias-corrected measures of the variance components, I find that manager fixed effects explain 9% of the total variation in productivity at the office level — about one third as much as the permanent component of productivity associated with different offices. Overall, a one standard deviation better manager increases office productivity by 10%. I also find that the (bias-corrected) covariance between manager and office fixed effects is negative, suggesting that INPS currently allocates the best managers to the least productive sites. This result is consistent with INPS trying to reduce inequality in productivity across sites.

In the second part of the paper, I exploit the rotation of managers as a natural experiment to study the mechanisms through which better managers achieve higher productivity. Previous research finds that effective private sector managers increase productivity by making better personnel (Hoffman and Tadelis, 2018) and investment decisions (Bennedsen et al., 2010, 2011). Public sector managers do not have the tools of their private sector counterparts, so increasing output may be particularly challenging. Instead, good managers may be those who are better at matching workers with tasks and use indirect strategies to elicit workers' effort and make unproductive workers quit or retire. My results are consistent with better task assignment from better managers. I find that the productivity gains are driven primarily by the exit of older workers who retire when a productive manager takes charge. Productive managers also maintain production without resorting to hiring or assigning overtime hours to compensate for the reduction in full-time equivalent employment. These findings are consistent with anecdotes suggesting that senior workers leave when better managers reassign more prestigious and better compensated tasks to other employees that are more junior but more productive. One might be worried that higher output per worker comes at the cost of quality of service provided. INPS also measures a quality index that captures both timeliness in processing claims and the rate of errors in subsequent random audits. I use this quality index to assess whether there is any a trade-off between productivity and quality of service, and I find that higher output per worker does not come at the cost of lower quality in this setting. These results imply that empowering managers to make payroll decisions would generate large efficiency gains for public sector offices.

As passing such drastic civil service reforms may not be feasible, in the final section of the paper, I discuss how governments could use these findings to improve public service provision by evaluating the efficiency gains from alternative managerial allocation schemes. The estimates from my productivity model imply that an optimal social allocation assigns the best managers to the largest and most productive offices. I find that if managers were reassigned on this basis, the agency output would increase by at least 6.9%.

This paper contributes to three strands of the literature. First, it contributes to the research that documents the impact of (1) managerial practices associated with better organizations and (2) managers themselves and their decisions. Both practices and managers matter for private sector firms (Bertrand and Schoar, 2003; Perez-Gonzales, 2006; Bloom and Van Reenen, 2007; Bloom et al., 2013; Lazear et al., 2015; Bloom et al., 2018; Bruhn et al., 2018; Hoffman and Tadelis, 2018; Giorcelli, 2019; Bandiera et al., 2020; Baltrunaite et al., 2020). In the public sector, studies have found that better organizational practices correlate with better outcomes in hospitals (Tsai et al., 2015), schools (Bloom et al., 2015), and civil service organizations (Rasul and Rogger, 2018; Rasul et al., 2019). But, it is not a given that managers' effectiveness in private sector firms, where all incentives are aligned to maximize profits and managers are imbued with broad authority, would necessarily carry over to the public sector. Most government operations in developed countries are very bureaucratic, fulfilling broad mandates with no semblance of competition or prices for individual services. But, the more bureaucratic the organization, the less it resembles a private sector counterpart's structure, and the fewer private sector tools are available to managers to increase productivity. My paper is the first to document managers' effectiveness in such an environment and study *how* managers improve office performance.[1]

Second, my work relates to the literature that studies the impact of civil servants on the performance of public sector institutions(Finan et al., 2017; Xu, 2018; Bertrand et al., 2019; Best et al., 2019; Choudhury et al., 2019; Khan et al., 2019; Janke et al., 2019). The primary challenge in this literature is credibly measuring output. For instance, two papers related to mine use different methods and arrive at different conclusions. Best et al. (2019) study public procurement. They conclude that bureaucrats vary in their ability to procure low cost goods, adjusting for observable quality differences using a machine learning classifier. Janke et al. (2019) study large public sector hospitals. They conclude that CEOs are not able to systematically improve a single performance index that combines multifaceted hospital output. The sophisticated approaches highlight the measurement challenges, especially since the economic agents multitask over the production of potentially many heterogeneous outputs. Implicitly, the credibility of the studies rely on civil servants uniformly agreeing on the researchers' measures as an objective. My setting provides several important advantages. First, I construct

[1] Choudhury et al. (2019) study public R&D labs in India and Janke et al. (2019) study hospitals in the UK. Neither studies a bureaucracy as constrained as mine, and both combine a number of outcomes to construct measures of productivity.

a comprehensive output-based measure of office productivity that is not subject to the concerns relative to multitasking that characterize previous studies. Second, I assess the productivity-quality trade-off using a measure of quality that is unavailable to the existing literature. Third, I use the rich INPS administrative data to recover the channels through which managers improve productivity and examine what makes for a productive manager.

Third, this paper also fits in the broad literature on productivity differentials between workplaces. Several papers have documented large and persistent differences in productivity across firms, even in narrowly defined industries (Syverson, 2004, 2011; Chandra et al., 2016). My paper contributes to this literature by providing compelling evidence that this phenomenon is not limited to the private sector and that it arises even *within* a large centralized public agency.

II Institutional Background

The *Istituto Nazionale di Previdenza Sociale* (INPS) employs 30,000 workers and administers applications for virtually all social welfare and insurance programs in the country including unemployment insurance, disability insurance, social security transfers, maternity leave, subsidies to the poor, and audits to firms and workers. Even though INPS is a large, centralized government agency, claim processing is decentralized. Every office has a catchment area and processes all claims that originate from it. The overall demand facing a given office largely reflects the demographic characteristics of residents and macroeconomic conditions. I study the offices that conduct the routine work associated with reviewing and processing claims. My sample includes 111 main satellite offices and 383 local branches.[2] Within each office, a single manager oversees production workers who assess whether to accept or reject claims (refer to Online Appendix C for additional details on within-office hierarchy).

At INPS, managers assign workloads and responsibilities, coordinate work inside the office, and ensure resources are used effectively. Their tasks include monitoring the production process and devising solutions whenever office performance falls short of production targets. However, managers are constrained from improving productivity through payroll decisions. Firing is uncommon in Italian government positions; a hiring freeze was instituted in 2008 (*blocco del turnover*) and covers the period of my

[2]These offices employ the vast majority of INPS workers. Refer to Online Appendix C for more details on the sample construction. Figure L.I in the Online Supplement shows the distribution INPS offices across the country.

analysis.[3] Thus, managers have to make the best out of their assigned set of workers. Anecdotally, more productive managers make their mark: they reassign workloads and responsibilities, change workplace practices, enforce break times, directly oversee employees' performance, and evaluate office operations using quantitative data.

Table I presents summary statistics of managers' characteristics. The first column includes all managers observed in my sample, while column 2 presents the characteristics of managers who are observed in at least two different offices (and therefore contribute to the estimation of the two-way fixed effects model discussed below). In the full sample, the average manager is 54 years old and has 27 years of civil service experience, commensurate with most managers having spent their entire career in civil service. Close to 60% were born in Southern Italy or the Islands, potentially reflecting the relative attractiveness of civil service jobs to people from those areas. About one-third of managers have a university degree in Law, and another 13% have a degree in Business, Administration, or Economics. Interestingly, over 20% have no university-level education. In comparison to the overall sample, managers who move across offices are younger, more likely to be male, and more likely to hold a university degree.

Next, I briefly describe manager rotation. INPS posts manager vacancies and their corresponding eligibility criteria on an internal website that is visible to all employees. As there are no official rules or unofficial guidelines on how to choose among qualified candidates, human resources officers select managers by making a case-by-case assessment (Online Appendix C). Managers stationed in main offices are forced to rotate every five years as part of anti-corruption law *190-2012*, which aims to prevent managers from becoming susceptible to corruption as a result of becoming too entrenched. Because of staggered tenures, relatively few vacancies are open in any given year, limiting the extent to which managers can sort.[4] However, this law does not apply to managers serving in local branches. As such, one may be concerned that managers may switch due to both plausibly exogenous reasons (e.g., retirement) and potentially endogenous choices (e.g., work closer to home). Nonetheless, the limited pool of candidates eligible to fill these positions, the lack of guidelines, and the many constraints related to the manager rotation limit the ability of managers to sort into offices. Specifically, if INPS decides to reallocate managers, the HR department faces a complicated problem. The same hiring and firing constraints that apply to production workers also apply to managers, so the HR department has to fill a given number of managerial positions by reshuffling a given set of managers. I further corroborate this argument by testing for

[3]The hiring freeze was introduced in 2008 and was aimed at progressively downsizing the public sector. This reform allows government agencies to hire one worker for four employees who leave.

[4]While ideally, I would like to limit my sample to managers stationed in main offices as their moves are plausibly more exogenous, I can not do so due to the limited sample size.

endogenous mobility in Section IV.

Workers and managers' salaries have a fixed component and a bonus. The former is tied to job title and the latter is a strictly increasing function of the levels of productivity and quality of service as well as the improvements of these two indicators relative to the previous year (refer to Online Appendix D for details on the bonus structure and the incentives that managers face in sorting into different office types). Managerial positions pay a fairly high salary as managers earn on average significantly more than the median Italian household income. While bonuses represent a small share of overall employee compensation, they amount to 15-30% of managers' salary.

III Data

This section details the quarterly office level data that form the basis of my analysis. These data are comprised of two main elements: data on office-level inputs and output and a personnel file that allows me to observe individual worker assignments to offices.

III.A Office-Level Productivity Measures

INPS has a centralized computerized quality control and internal monitoring system aimed at tracking every step of the production process. I use their internal monitoring data from Q1 2011 to Q2 2017. These data report inputs including the number of full-time equivalent workers devoted to production (FTE_{it}) at office i in quarter t as well as information on absences, overtime hours, and hours devoted to training by workers in each office. INPS also constructs a (complexity-weighted) standardized index of claims processed by each office as a measure for office output (Y_{it}). Specifically, the number of claims ($c_{v,it}$) of different types ($v = 1, \dots V$) processed by office i in quarter t are aggregated into a single output measure by weighting them by their complexity ($w_{v,t}$).

$$Y_{it} = \sum_{v=1}^{V} c_{v,it} \times w_{v,t}.$$

The weights represent the time employees *should* take to process each type of claim (refer to Online Supplement J for details). Importantly, appropriately weighting claims by their complexity controls for differences in tasks across offices (Autor et al., 2006; Autor, 2013; Stinebrickner et al., 2018). Although INPS employees' main task consists in processing paperwork, they also take turns working at the front-office where they assist beneficiaries. I provide more information on front office operations and show that this component of production is not driving my results in the Online Appendix A.

I combine the measures of office output and FTE employment to construct my measure of productivity (P_{it}) as output per worker:

$$P_{it} = \frac{Y_{it}}{FTE_{it} \times 3} = \frac{\sum_{v=1}^{V} c_{v,it} \times w_{v,t}}{FTE_{it} \times 3}.$$

The numerator is total quarterly output. I multiply average monthly FTE_{it} by 3 in the denominator to produce P_{it}, the number of complexity-adjusted claims per worker in an average month at office i in a given quarter t.[5] One advantage of my setting is that workers' tasks are extremely mechanical and INPS devotes a lot of effort into measuring each individual task. So, contrasting with other studies of productivity multitasking (Holmstrom and Milgrom, 1991) is much less of a concern in my setting.

A concern is that dispersion in productivity may be driven by demand volatility and that when demand is low workers are left idle. This is not the case due to two reasons: first, offices have a large backlog (refer to Table II and Online Supplement J). Second, managers who are in charge of offices facing low demand are instructed to contact a high-demand office and ask to transfer claims electronically to equalize workloads across sites (Online Appendix B). One may also worry that, if weights do not correctly reflect task complexity, managers could try to game the system by shifting production toward overvalued claims and neglect undervalued ones. I address these concerns in Section VI.

Throughout the analysis, I primarily focus on productivity as it relates to quantity of output. However, the data also allow me to test whether managers matter in affecting quality. In particular, INPS constructs an index of service quality, a weighted average of "timeliness" (the fraction of claims processed within the first thirty days) and the "error rate" (the fraction of claims that has to be processed more than once because of an error in initial processing).[6] My data does not contain these two sub-components, therefore I can not analyse them separately.

III.B Office-Employee Data

In addition to office-level productivity data, I have access to a personnel file that allows me to track employees over time within INPS (2005-2017). This dataset includes office

[5] As I use log productivity in my analysis, this normalization does not affect any of the estimated coefficients.

[6] INPS audits 5% of each office production twice per year, and most mistakes are detected during these audits. Some are also found when denied beneficiaries file an appeal (Online Supplement J). My data does not contain information on the audits, the number of mistakes, and the number of appeals filed so I can not analyze these components separately.

location, job title, hiring, firing, separations, and promotions.[7] Anecdotally, most employees are hired through a competitive examination (i.e., *concorso pubblico*) or from other government agencies. Workers rarely quit a public sector job, and the vast majority of them leave the INPS when they retire. Since I do not observe retirement directly, I use this anecdotal evidence to construct a proxy for it. I define retirements as voluntary separations of workers over age 60 (refer to Online Supplement J for details).

III.C Descriptive Statistics and Stylized Facts

In this subsection, I present an overview of Social Security Offices and I document two stylized facts related to productivity.

Table II reports the summary statistics for the full sample in column 1; columns 2 and 3 display the statistics for main offices and local branches respectively. Main offices are substantially larger than local branches. A typical main office employs on average 115 workers, while a local branch has on average only 16 employees. As labor is the main input of the production process, larger office size translates to higher output. Offices have a large backlog which amounts to 80% of the average quarterly inflow of new claims. While all offices have large backlogs, this phenomenon is more pronounced in main offices. Interestingly, main offices are 12% more productive than local branches on average. Despite these stark differences between main offices and local branches, the quality index and absenteeism rates do not seem to differ substantially across these two types of production sites. Overall, employees devote a very small fraction of their time to training and overtime work (column 1). Hiring is extremely limited in this context, 0.5 workers per office separate from INPS on average every quarter (48% of which are due to retirement), and 0.3 workers transfer to another office within the Social Security Agency (Online Appendix J).

In the remaining part of this Section, I document two stylized facts. First, there is a surprising amount of variation in productivity across offices over time, even *within* a large centralized agency. Offices located at the 90-th percentile of the productivity distribution are 2.6 times more productive than those at the 10-th percentile (Figure H.I). Although comparing productivity differentials across industries is notoriously hard, I benchmark my estimates with previous studies. I compare the distribution of log productivity in my sample (Panel A of Table H.I) with the within-industry plant-level distribution moments in Syverson (2004) (Panel B). There might be reasons for believing the dispersion in productivity across offices that belong to the same centralized agency is substantially smaller than the one across plants within the same industry; yet my es-

[7] The personnel data does not include information on wages and earnings.

timates are somewhat smaller, but comparable, to those in Syverson (2004). Second, there are large productivity differentials not only across but also within regions. Figure H.II plots the average productivity in each province over my sample period. This figure shows that while the North is more productive on average, there is a substantial variation within each geographical region.

IV Do managers matter in the public sector?

I develop a framework which exploits manager rotation across sites to decompose productivity into a manager and an office component. I discuss the identification challenges that arise in this context and estimate the model. I then perform a series of diagnostic checks which evaluate the model specification. I conclude this section by summarizing the implications of this model in a variance decomposition exercise and comparing my estimates to the literature.

IV.A Model

I begin by assuming that the aggregated output index (Y_{it}) is produced according to a Cobb-Douglas production function:

$$Y_{it} = \exp\left(A_{it} + M_{it}\right) K_{it}^{a} H\left(\boldsymbol{L}_{it}\right)^{1-a}. \tag{1}$$

A_{it} represents total factor productivity and it is the sum of an office-specific component (α_i), an aggregate shock (ν_t), an office-specific trend (ζ_{it}), and a transitory shock (ε_{it}), i.e., $A_{it} = \alpha_i + \nu_t + \zeta_{it} + \varepsilon_{it}$. $m(i,t)$ maps office i and quarter t to the manager's identity m. M_{it} represents managerial talent, and it is the sum of the innate component of managerial talent ($\lambda_{m(i,t)}$) and an office-manager match component ($\eta_{im(i,t)}$), i.e., $M_{it} = \lambda_{m(i,t)} + \eta_{im(i,t)}$. I assume that there are ℓ worker types who differ in their innate productivity. Denote bold $\boldsymbol{L}_{it} = (L_{it}^1, L_{it}^2, \ldots, L_{it}^L)$ to be the vector of workers of different types, $H\left(\boldsymbol{L}_{it}\right)$ to be a labor aggregate that enters the production function, and abusing notation slightly L_{it} to be the number of workers at the office $L_{it} = \sum_\ell L_{it}^\ell$.

The Italian context suggests some simplifying assumptions. All production employees work on the same software and labor is the main input of production. There are virtually no differences in per-worker physical capital across sites and little scope for manager input. I specify total physical capital as $K_{it} = k_t \times L_{it}$, the product between a per-worker capital component k_t, which does not vary across offices, and office size.

Managers are likely to matter the most with respect to workers. First, despite institutional constraints, a good manager may be able to affect worker composition and hence workers' average productivity $\mu_{m(i,t)} = \ln \frac{H(\boldsymbol{L}_{it})}{L_{it}}$ (Hoffman and Tadelis, 2018) and office size $L_{it} = \sum_\ell L_{it}^\ell$ (Lucas, 1978). Second, a productive manager may be better at eliciting workers' effort and assigning workers to tasks to suit their comparative advantage, reflected in higher $\lambda_{m(i,t)}$. Third, a manager may have a comparative advantage at overseeing a specific office, captured by a higher $\eta_{im(i,t)}$.

There are two key innovation of this paper. First, by defining a comprehensive measure of output, INPS explicitly defines the managers' objective function. INPS rewards managers based on productivity, defined as output per worker $P_{it} = \frac{Y_{it}}{L_{it}}$. Managers select $L_{it}^1, L_{it}^2, \ldots, L_{it}^L$ and assign workers to tasks to maximize their bonus $\max_{L_{it}^1, L_{it}^2, \ldots, L_{it}^L} = \text{bonus}\left(P_{it}, \text{quality}_{it}\right)$.[8] Managers maximize productivity and quality like private sector CEOs maximize profits. This is typically not the case in public sector institutions where, with their own diverse beliefs of the institution mission, managers often have subjective perceptions of how to weigh the outputs, making it difficult for researchers to not only measure each output that managers are trying to optimize, but also separate differences in ability from differences in beliefs. Second, the vast majority of papers that study the public sector directly focus on bureaucrats and front-line providers, L_{it}^ℓ in my framework. In this paper, I study the direct impact of the innate component of managerial talent on office performance as well as the effects mediated through personnel decisions. Because technology is constant across offices, the same rules apply to all sites, and offices produce a homogeneous product, this is an ideal setting to study the impact of managers on office-level outcomes. In other words, sites are not subject to many of the factors that confound interpretation of managers' effects in other studies.

Substituting the production function in equation (1) into the definition of productivity and taking logs yields:

$$\ln P_{it} = \alpha_i + \underbrace{a \ln k_t + \nu_t}_{\tau_t} + \underbrace{\lambda_{m(i,t)} + \mu_{m(i,t)}}_{\theta_{m(i,t)}} + \underbrace{\eta_{im(i,t)} + \zeta_{it} + \varepsilon_{it}}_{u_{it}}. \tag{2}$$

I summarize (2) with a combination of office, time, and manager effects in my main estimating equation:

$$\ln P_{it} = \alpha_i + \tau_t + \theta_{m(i,t)} + u_{it}. \tag{3}$$

My main parameters of interest are the portable components of managers' ability $\theta_{m(i,t)}$.[9]

[8] The bonus formula is monotonically increasing in the arguments. It is qualitatively summarized in Online Appendix D and formalized in the Online Supplement I.

[9] Given my relatively short panel, I can not estimate time-varying manager effects to allow for career dynamics and manager learning. While these phenomena are potentially very interesting, they are unlikely to represent a major concern in this setting as most managers have been working in the public

I refer to them is as manager quality or managerial talent interchangeably.

IV.B Identification

The rotation of managers across sites allows me to separately identify the impact of managers and office heterogeneity on productivity. The office effects (α_i) proxy for the time-invariant characteristics of the office (e.g., geographical location, average quality of the workers at office i, main office vs local branch) and for size/composition of the workforce to the extent that these variables do not change over time. I include time fixed effects τ_t to absorb seasonality and macroeconomic shocks.

Additionally, model (3) postulates that productivity changes discretely as a new manager takes over. However, in reality, managers may take some time to change work practices. As I do not want the estimated manager effects to be confounded by switching costs or measurement error in manager identity[10], I estimate (3) excluding the first quarter in which the new manager is in charge. One may be concerned that the adjustment may take longer than a quarter and that there may be some persistent impact of past managers on office productivity. If this was the case the estimated manager fixed effects would represent a lower bound of the true managerial talent.

I can re-write (3) in matrix notation as:

$$\ln(P) = D\alpha + G\theta + T\tau + u, \tag{4}$$

where D, G, and T collect all the office, manager, and time dummies respectively. OLS identifies the parameters of interest under the following identifying assumptions:

$$E[d_i' u] = 0 \ \forall i, \tag{5}$$

$$E[g_m' u] = 0 \ \forall m, \tag{6}$$

where d_i is the i-th row of the matrix D and g_m is the m-th row of the matrix G.

It is well known that the office and manager effects in (3) are identified by movers. As I can separately identify manager from office effects only within a connected set (Abowd et al., 1999), I can meaningfully compare the estimated fixed effects only within and not across connected sets (refer to Online Supplement K for a discussion of the normalization). The identifying assumptions (5) and (6) impose that manager mobility is as-good-as random, conditional on office and time fixed effects.

sector for almost 30 years and are in the very last part of their career (Table I).

[10] If the switch does not occur on the first day of the quarter, I assign the quarter of the switch to the manager with the longest spell in that quarter.

Loosely speaking, these orthogonality conditions are satisfied if the assignment of managers to offices depends only on the permanent component of office productivity (α_i) and/or the permanent component of managerial ability ($\theta_{m(i,t)}$). For example, better managers sorting into more productive offices would not violate the identifying assumptions. By the same token, if productive managers were systematically sent to local branches or to a specific geographical area, this would not represent a threat to the identification strategy. (5) and (6) allow for rich patterns in the sorting of managers to offices. Violations of the exogenous mobility assumption occur when managers sort on the error term.

I follow Card et al. (2013) and consider three forms of endogenous mobility that depend on any office-manager match component of productivity ($\eta_{im(i,t)}$), on any office-specific trend in productivity (ζ_{it}), or on any transitory component of office productivity (ε_{it}). In particular, I specify the following composite structure of the error term:

$$u_{it} = \eta_{im(i,t)} + \zeta_{it} + \varepsilon_{it}, \tag{7}$$

I assume that $\eta_{im(i)}$ has mean zero for all offices and all managers in the sample. ζ_{it} is a drift component, which captures offices improving or deteriorating over time; I assume this component has a mean zero for each office but contains a unit root. ε_{it} is an idiosyncratic error term and represents transitory shocks; I assume that ε_{it} has mean zero for each office.

Given the error structure posited in equation (7), the assumptions in equations (5) and (6) rule out three types of sorting. First, managers can not sort into offices on the basis of their comparative advantage. Second, better managers cannot be systematically sent to offices whose performance is worsening over time (i.e., matching based on underlying trends in productivity). Third, a better manager cannot join an office in response to a negative transitory productivity shock. I use the estimated manager effects to conduct a series of tests for the presence of endogenous mobility in Subsection IV.D.

IV.C Results

Table III describes the structure of my sample of quarterly level observations on office-level productivity. The first column reports the statistics for the full sample, while the second column restricts attention to the balanced-analysis sample. The latter includes the subset of offices for which I observe the outgoing manager being in charge for at least four quarters before the change in leadership and the incoming manager being assigned to the office for at least six quarters after that. The full sample contains 851 managers, 494 offices, and 276 connected sets (Table III, column 1). Roughly one-fourth of these managers move across sites and almost 80% of offices experience a change in management between 2011 and 2017 (column 1). The remaining 20% of the offices do not contribute to the estimation of the manager effects. All offices experience a change in leadership in the balanced-analysis sample by construction, and 30% of managers move across sites (column 2).

In order to assess the amount of dispersion in public sector productivity attributed to managers, I follow Bertrand and Schoar (2003). I compare the adjusted R^2 estimated from a regression of the logarithm of productivity on office and time fixed effects, model (8), to the one from model (3) which also includes manager fixed effects.

$$\ln P_{it} = \alpha_i + \tau_t + \tilde{u}_{it}. \tag{8}$$

Model (3) nests (8) under the assumption that managers have no impact on office productivity. Table IV reports the estimates from (8) and (3) in columns 2 and 3, respectively. The adjusted R^2 increases from 0.69 in column 2 to 0.76 in column 3, suggesting that managers explain a non-trivial amount of the variation in productivity across sites. Although the increase in the adjusted R^2 might seem small, its magnitude is very similar to that reported in Bertrand and Schoar (2003).

To test more formally whether managers affect productivity, I perform an F-test for the null hypothesis that the manager effects are jointly zero. I reject the null at any standard significance level (p-value=0.000). Notice that the adjusted R^2 of columns 2 and 4 are high relative to the one reported in column 3, which suggests that manager and office fixed effects are highly correlated in this setting. These R^2's are lower than those of two-way fixed-effect models that decompose wages (Card et al., 2013). The reason is that productivity is intrinsically more volatile than wages.[11]

IV.D Diagnostic Checks

In Section II I argued that the institutional framework severely limits the ability of managers to sort into offices. I now test for detectable evidence of this phenomenon focusing on patterns that might be related to the components of the error specified in equation (7). First, I discuss sorting on the drift component (ζ_{it}). Second, I address concerns related to managers being assigned to offices on the base of unobservable factors determining their comparative advantage ($\eta_{im(i)}$). Third, I consider sorting on the transitory component of the error term (ε_{it}).

[11] For completeness, I report the same exercise using quarterly data in Online Appendix H (Table H.II). Although quarterly productivity is a somewhat noisier outcome, the results are largely unchanged.

One might be wary of endogenous mobility related to office-specific trends in productivity. As a concrete example, if good managers were able to systematically move to offices which are improving over time, my model would overestimate their managerial quality. I investigate this concern in Table V by evaluating the correlation between baseline office characteristics and estimated fixed effects of future managers. Intuitively, managers cannot impact office performance before they take charge, hence any correlation between future manager ability and baseline office characteristics is indicative of sorting. As a benchmark, if managers were randomly assigned one would expect manager productivity to be uncorrelated with observable pre-determined characteristics of the office.

The results in column 1 of Table V show that, more productive managers are less likely to serve in main offices and more likely to be assigned to Northern or Central Italy. Importantly, future manager effects do not appear to be correlated with office growth rates. I also test whether the explanatory variables are jointly statistically significant and whether growth rates can jointly predict future manager fixed effects. I can reject the latter but not the former. Overall, there is some evidence of managerial sorting on geography and office type, but not on growth rates. As I discussed earlier, manager assignment being correlated with time-invariant characteristics of the office does not pose a threat to my empirical strategy. I repeat the same exercise using the change in the estimated fixed effects as the dependent variable (column 2 of Table V). The overall pattern of results is largely unchanged. These findings show that there is no evidence of managers sorting on the drift component. They also further motivate the use of office fixed effects in my main specification to control for sorting based on time-invariant characteristics of the office.[12]

Another way to test whether the sorting of managers to offices is driven by serially correlated error components in office or manager productivity is to examine the residuals from (3) associated with specific forms of manager changes. When an office goes through a change in management, it can experience three types of transitions: an overall increase in manager ability ($\widehat{\Delta M_i} > 0$, where $\widehat{\Delta M_i}$ represents the change in the estimated manager fixed effects), a decrease in management quality ($\widehat{\Delta M_i} < 0$), or no significant change ($\widehat{\Delta M_i} \approx 0$). Figure I reports the event study for (trend-adjusted) office productivity for these types of transitions (i.e., tertiles of $\widehat{\Delta M_i}$). Figure I shows that average log productivity remains relatively flat in the four quarters before the change in management and jumps discontinuously at the time of the event. The lack of pre-trends corroborates the claim that officers do not sort into sites based on the drift component. I test more formally for the presence of pre-trends in Section V, and I do not find evidence

[12]These results are robust to the exclusion of the lagged variables (Table L.I in the Online Supplement).

of this phenomenon. Importantly, the fact that productivity appears to be slightly lower in the quarter of the switch than in the following three quarters, motivates my choice of excluding the quarter in which the takeover takes place from my two-way fixed-effect model.

Next, I test for sorting on the match component of productivity by examining gains and losses as managers move from office to office. As already noted, Figure I displays a remarkably symmetric pattern, which is not consistent with managers sorting on their idiosyncratic match component (i.e., sorting based on comparative advantage of specific managers at specific offices). I also compare the fit of model (3) with a fully saturated model that includes manager-by-office dummies. In the presence of match components, the latter should fit substantially better than the former. The adjusted R^2 of the fully saturated model is only marginally higher than the two-way fixed effect specification (0.764 in column 5 vs. 0.762 in 3 of Table IV), suggesting that match components are not quantitatively relevant in this context.

Finally, manager rotation could be correlated with the transitory component of the error term. This could be the case if managers were to relocate to a less productive office after a particularly bad ε_{it} draw. Once again, this is not consistent with the lack of pre-trends reported in Figure I and in Section V.[13]

A final set of concerns about model (3) regards the assumption of additive separability between the permanent office component and managerial ability. A violation of the additive separability assumption would result in abnormally large/small mean residuals for some office-manager pairs. To assess whether this is the case, I divide the estimated manager and office effects into quartiles. I compute the mean residual for each cell. Figure II reports these statistics. All values are rather low, and the highest mean residual is equal to 0.01.[14] Overall, this finding suggests that match effects, if present, are not quantitatively relevant in this context. The analysis presented in this subsection supports the claim that the two-way fixed effect model approximates the data fairly well in this setting.

IV.E Variance-Covariance Decomposition

We might expect social norms and workforce composition (proxied by office fixed effects) to be important drivers of productivity. However, it is less obvious whether public

[13] Refer to Online Appendix D for a discussion of how the bonus structure can induce managers to sort into different office types and whether this is a threat to the empirical strategy.

[14] The mean residuals have been computed using all the connected sets in which there are at least four offices and four managers. Managers and offices are ranked within a connected set. Figure H.III reports the same exercise on the largest connected set.

sector managers, who operate in a severely constrained environment, can have an impact on productivity. I propose a variance decomposition exercise that allows me to assess whether these two dimensions matter and their relative importance. If manager ability and office characteristics are important determinants of productivity, then $Var(\alpha_i)$ and $Var(\theta_{m(i,t)})$ should explain a large share of the variation in observed productivity. I use (3) to decompose the variance of productivity into (Abowd et al., 1999):

$$\begin{aligned} Var(\ln P_{it}) = & Var(\alpha_i) + Var(\theta_{m(i,t)}) + Var(\tau_t) + Var(u_{it}) \\ & + 2Cov(\theta_{m(i,t)}, \alpha_i) + 2Cov(\theta_{m(i,t)}, \tau_t) + 2Cov(\alpha_i, \tau_t). \end{aligned} \tag{9}$$

Table VI reports the bias-corrected variances and covariances estimated on the largest connected set (Andrews et al., 2008; Gaure, 2014). This procedure allows me to obtain consistent estimators of the variances and covariances of interest in the presence of limited mobility bias. Manager fixed effects explain roughly 9% of the variance in log productivity, about one third as much as the permanent component of productivity associated with different offices. Time fixed effects explain a non-trivial share of the variation in productivity, which is mainly driven by seasonality in productivity and the overall improvement in the Social Security Agency performance over time.

Interestingly, the bias-corrected covariance between manager and office effects is negative, namely more productive managers currently work at less productive offices (i.e., negative assortative matching). This finding is crucial for the interpretation of the counterfactual exercises I develop in Section VII. It is worth emphasizing that this result is somewhat unusual. Most economic models predict positive assortative matching and the recent literature on wage determination suggests that higher-wage workers tend to sort to firms that offer higher wage premiums (Card et al., 2013; Song et al., 2015). This result is consistent with INPS trying to reduce inequality in productivity across sites.[15]

IV.F Manager Effects and Observable Characteristics

I conclude this Section by discussing the magnitude of the estimated manager effects and showing how they correlate with observable characteristics of top-level officers.

Managers have a large impact on office performance. A two standard deviation in-

[15] Several papers have found evidence of negative assortative matching using the two-way fixed effects framework. However, these findings may be tainted by limited mobility bias. I refer to Andrews et al. (2012) for a complete treatment of this issue. A recent study by Adhvaryu et al. (2020) documents negative assortative matching between managers and production lines in an Indian garment factory and argue that it is driven by the strong incentives in reducing delays on any particular order. The work of Limodio (2019) also suggests that high performing World Bank managers often work in poorly performing countries and that the negative assortative matching strengthens in the aftermath of a natural disaster.

crease in managerial talent leads to a 20% increase in office productivity (Table VI). Although it is challenging to directly compare point estimates across industries and countries, I benchmark the magnitude of this effect with the work of Bloom et al. (2013). The authors find that the adoption of management practices induces a 17% increase in productivity in textile firms in India. My results suggest that a very good manager has a comparable effect on productivity in the Italian government.

To study how estimated managerial talent correlates with observable characteristics, I regress the estimated manager fixed effects from (3) on gender, experience, experience squared, a set of dummies for region of birth, and a set of indicators for the highest educational attainment, as well as connected set fixed effects (Table VII). Female managers appear to be on average more productive than their male counterparts. Not surprisingly, managerial talent is strongly correlated with experience although it exhibits decreasing marginal returns. There is some suggestive evidence that managers born in Southern Italy or the Islands are more productive than those from the North and that those who never attended college are better than those who studied law or STEM. Importantly, these coefficients should not be interpreted causally; these correlations can be explained by differential selection patterns into public sector jobs and managerial career. In particular, these findings are consistent with negative selection into government jobs for men, those born in the North, and those who have a STEM major.

In this section, I have shown that there is substantial variation in managerial talent within this large centralized agency and managers have a quantitatively meaningful impact on office productivity. Next, I open the black box of manager fixed effects and try to analyze the specific mechanisms that drive the effects of more and less productive managers.

V What makes for a productive manager?

Better managers could affect office productivity through a variety of mechanisms that include better personnel decisions, more competent management of office operations, and eliciting effort from workers. Given the institutional constraints discussed in Section II, managers are unlikely to have an impact on hiring and firing. However, they can in principle affect office operations by changing workers' time allocation and the assignment of tasks to workers. As employees enjoy strong job security, soft skills may be particularly important when eliciting effort from workers. In this section, I utilize manager rotations as a quasi-experimental analog of random assignment of managers to offices to characterize *how* managers matter. I first decompose the productivity gains induced by a change in leadership into its effects on output and full-time equivalent employment. Second, I explore how changes in managerial talent impact workers through personnel decisions and changes in their time allocation. Third, I evaluate the productivity-quality trade-off. Finally, I test for heterogeneous treatment effects.

V.A Event Study Strategy

I begin by specifying a basic event study regression that relates outcome y (e.g., output, FTE, new hires etc.) to changes in managerial talent ΔM_i:

$$y_{it} = \alpha_i + \sum_{k \neq -1} \left[\pi_0^k D_{it}^k + \pi_1^k D_{it}^k \Delta M_i \right] + g_t\left(X_{it}\right) + \varepsilon_{it} \tag{10}$$

where k indexes quarters relative to a change in management (positive values of k refer to quarters after the event and negative values to those prior) and the π_0^k coefficients capture dynamics related to a change in leadership that are common across all offices. The main objects of interest are π_1^k's, which capture the extent to which effects differ depending on the change in quality of incoming managers relative to the managers they replace (i.e., ΔM). This coefficient represents a lower bound on the true effect as ΔM is estimated with error. Permanent differences in office productivity are captured by α_i and $g_t\left(X_{it}\right)$ controls flexibly for time trends. The identifying assumption is that changes in management quality are not coincident with the evolution of other unobservable factors. Event study frameworks of the type described in equation (10) are typically estimated via OLS, and the identifying assumption is tested by evaluating the pre-trends.

If manager effectiveness were observable to the econometrician, equation (10) could be estimated directly. Managerial talent is fundamentally unobservable but the two-way fixed effects model enables me to estimate it by exploiting the rotation of managers across sites. However, using the first-step estimates as covariates in (10) could bias π_1^k. Idiosyncratic productivity shocks could affect both my estimates of manager effectiveness and the outcome of interest, creating a spurious correlation even in the absence of a causal relationship. A natural solution is to purge idiosyncratic shocks by estimating the first step leaving out data where correlations may arise. I overcome this challenge by modifying the standard event study specification in two ways. First, I subtract from equation (10) in each event time the corresponding values in event time $k = -1$. Second, to purge the regressor of potential mechanical correlations, I generate the change in manager productivity by using estimates from separate two-way fixed effect models, *excluding* data from event times -1, 0, and k.[16] Formally,

$$\Delta y_i^k = \pi_0^k + \pi_1^k \widehat{\Delta M}_i^{L,k} + \Gamma^k X_i + \Delta \varepsilon_i^k \tag{11}$$

where

$$\widehat{\Delta M}_i^{L,k} = \hat{\theta}_{i,incoming}^{L,k} - \hat{\theta}_{i,outgoing}^{L,k}$$

and the $\hat{\theta}_{i,\cdot}^{L}$'s are the leave-out estimated manager effect of the incoming and outgoing managers, respectively (where L superscript stands for "leave-out").[17] X_i includes indicators for being in the Center-North of Italy, for being main offices, for quartiles of baseline productivity, and two-way interactions between each of these. I also flexibly control for trends by including time dummies and time dummies interacted with the dummy for being in the Center-North of Italy. The specification in (11) suggests estimating separate regression models for each event time. I focus on the $[-4,6]$ window, and I limit my sample to the subset of events which are balanced over this time horizon (i.e., balanced-analysis sample).[18] I bootstrap the standard errors to account for the presence of a generated regressor.

The separate regression models and leave-out procedure ensure that $\widehat{\pi_1^k}$'s are not spuriously driven by contemporaneous, idiosyncratic shocks that affect both the estimated manager effects and the outcomes of interest. The underlying parameters of the differenced model and the standard model in equation (10) are nonetheless the same, allowing me to directly test the over-identifying restrictions of parallel trends. I interpret violations of the parallel pre-trend assumption as evidence of time-varying unobservable confounding factors.

This procedure ensures that the outcomes of interest are not directly related to my measures of manager ability. However, if unobserved productivity shocks u_{it} are serially correlated, then my leave-out measure may still be spuriously, yet indirectly correlated with outcomes. While leaving out more data in the first-step mitigates concerns relative to mechanical correlation, it also increases measurement error in my estimates of manager effectiveness. Given the limited number of years in my sample, I do not have enough data to pursue a leave-office-out estimation strategy. Reassuringly, serially correlated productivity shocks do not appear to be a concern in my setting. First, the autocorrelation coefficient from fitting an AR(1) model to the residuals from equa-

[16] As described in Section IV, managers may take some time to change work practices. Hence, I never include the quarter of the switch (i.e., $k = 0$) when estimating manager effects via (3).

[17] $\widehat{\Delta M}_i^{L}$ ranges from -0.4 to 0.55 and looks approximately normally distributed (Figure H.IV). Crucially, $\widehat{\Delta M}_i^{L}$ does not depend on the normalization I choose (Online Supplement K).

[18] I can not expand my window further as the balanced sample becomes excessively thin.

tion (3) is extremely small ($\hat{\rho} = 0.04$). Second, in simulation analyses, I hold constant the mobility structure from my sample and generate simulated outcomes suffering from different degrees of autocorrelation. I show that a substantial degree of autocorrelation is needed to represent a serious threat to my empirical strategy (Online Appendix E).

V.B Decomposition of Productivity Impacts

The two-way fixed effect model documents that managers matter and parsimoniously summarizes their contribution in a single measure. However, this measure does not give any insight into the mechanisms and their timing. I begin the analysis of the mechanisms by examining the timing of the productivity gains; next, I decompose them into their effect on output and full-time equivalent employment.

Figure IIIa reports the estimated impact of an increase in managerial quality on office-level productivity.[19] This figure collects the estimated coefficients and their 95% bootstrapped confidence intervals obtained by running a separate regression for each time horizon (k).[20] Reassuringly, changes in managerial talent do not predict changes in productivity before the event takes place (placebo tests), which alleviates concerns regarding endogenous mobility. Productivity starts increasing when the better manager takes office and it stabilizes to its new level one quarter after the change. These results show that improving management quality increases productivity, but that the effect does not fully materialize in the first quarter. This is consistent with the presence of adjustment costs associated with manager rotation.

Next, I discuss how I can decompose the impact of managerial talent on office productivity into its effect on output (numerator) and FTE (denominator). Holding office composition constant, managers may increase output by eliciting more effort from workers or by better matching employees to tasks. Because public sector managers have limited discretion over workers' promotions and compensation, increasing output may be particularly challenging. Managers also have limited ability to make de jure personnel decisions (e.g., hiring and firing). In this setting, good managers may be those who are better at matching workers with tasks and use indirect strategies to elicit workers' effort and make unproductive workers quit or retire.

When a better manager takes over there is a modest (although not statistically significant) increase in output (Figure IIIb). This pattern is consistent with more effective managers better matching workers to tasks and eliciting more effort. This finding is striking in light of the sharp decrease in the number of employees assigned to the of-

[19] Table H.III reports the results of Figures IIIa, IIIb, and IIIc in a table format.

[20] I bootstrap standard errors and confidence intervals over 1,000 replications.

fice after a better manager takes charge (Figure IIIc). Most of the productivity gains are driven by the reduction in the number of workers assigned to the office. A 10% increase in managerial talent increases office output by 1.7% and generates a 4.9% reduction in full-time equivalent employment six quarters after the event (Table H.III , column 2 and 3, respectively). This finding speaks directly to the consequences of downsizing the public sector and suggests that reducing the number of public sector employees does not necessarily lower the volume of services provided.

V.C Reduced Form Impacts of Managerial Talent on Workers

How do managers reduce office size in such a constrained environment? Table VIII explores the channels through which managers impact the composition of workers they supervise. The dependent variables represent *cumulative* flows.

Better managers drive older workers to retire (column 1). The bulk of the manager-induced retirements occur in the first two quarters after the change in leadership. Managers and higher level officials cannot force older workers to retire, and they cannot negotiate severance packages to persuade them to leave.[21] So, why do retirement-age workers only leave when better managers arrive? One plausible explanation comes from the norm that more senior workers are usually given more prestigious tasks that come both with additional responsibilities and importantly, additional compensation. Anecdotal evidence suggests that more productive managers reallocate these tasks when they take charge. Either because they are slighted or because they lose the extra compensation, senior employees retire. Unsurprisingly, an increase in management quality does not have any statistically significant impact on hiring and firing (columns 2 and 3) because managers have limited hiring and firing authority. The arrival of a more productive manager is associated with fewer inbound (column 4) and outbound transfers (column 5) three quarters after the change in leadership. One may be concerned that if transfers were to occur coincidentally with the change in leadership, this could affect the interpretation of my results. However, as office productivity increases in the quarter in which a better manager takes charge and the effects on transfers materialize a few quarters later, inbounds and outbounds transfers are unlikely to be driving my results.

I also investigate how time allocation changes with the takeover of a more productive manager. Changes in managerial quality do not appear to produce strong and persistent effects on training, overtime work, and total hours (Online Appendix F), while there is some suggestive evidence that better managers may be able to reduce absenteeism rate. Remarkably, more productive managers succeed in keeping up production

[21] See Online Supplement J for a description of the retirement system.

without resorting to more overtime hours to compensate for the reduction in FTE (column 3 of Table F.I). This is consistent with more productive managers being able to better match workers with tasks.

V.D Reduced Form Impacts of Managerial Talent on Quality and Backlog

One might fear that there is some trade-off between productivity and quality of service provided. Column 1 of Table IX shows that the arrival of a more productive manager does not negatively impact quality. This result is likely to be driven by the fact that the incentive-pay scheme provides a direct incentive for managers to increase productivity without letting quality deteriorate. There is also some suggestive evidence that effective managers lower backlog (column 2 of Table IX). In light of the work by Autor et al. (2015), this result points toward potentially large benefits to the claim beneficiaries driven by a reduction in processing time.

V.E Heterogeneity

A productive manager might be able to have a larger impact on an unproductive rather than on a very productive office. Likewise, she could be more effective in a smaller than in a larger site, or in some specific geographical areas. I test for heterogeneous treatment effects in Online Appendix G and show that productivity gains do not appear to differ by geographical location, office size, main offices vs local branches, baseline productivity and social capital, although I have admittedly limited power to detect heterogeneous effects.[22]

VI Robustness Checks

In this section, I address some concerns regarding my empirical strategy. I first show that manager effects are not confounded by demand shocks. Second, I test whether managers appear to game the system. Third, I relax the assumption that manager effects translate linearly in office level outcomes.

[22]These findings are in line with the absence of heterogeneous treatment effects postulated by my two-way fixed effects model (see 3).

VI.A Demand

A concern might be that high-fixed effect managers are classified as "productive" because they happen to be working at offices that received a site-specific positive demand shock. Demand for services is measured as the number of claims that originate from the office catchment area and it is exogenous to the office, as it depends on the demographic characteristics of those living in the catchment area as well as its economic condition. Unlike managers in the private sector, top-level bureaucrats cannot advertise their products or take actions to affect the local demand for public services. I use model (11) to test whether demand is higher on average when a more productive manager is in charge. Column 3 of Table IX shows that more productive managers are not in charge during periods of high demand. Demand appears to be extremely volatile even when controlling for time fixed effects, which is consistent with the fact that it is mostly driven by local idiosyncratic shocks. As an additional robustness check, I correlate my estimated manager fixed effects with those estimated by controlling for the logarithm of the number of claims originated in each quarter. The correlation is extremely high (98.8%). I conclude that the estimated manager effects do not appear to be confounded by demand shocks.

VI.B Do Managers Game the System?

As bonuses are a strictly increasing function of productivity, one could worry that the managers I classified as productive are, in fact, those who are able to game the system. In particular, if the weights are mismeasured, managers might try to shift production toward the overvalued products and shift away from the undervalued ones.[23]

First, manipulation should be mitigated by the fact that if managers decided not to process undervalued products in a timely fashion, this would reflect negatively on the quality index and, in turn, on their bonuses. Quality is not negatively impacted when a more productive manager takes over (column 2 of Table IX), which attenuates these concerns. Second, 5% of all claims processed by each office are audited twice per year; the purpose of this cross-check is to monitor the production process and detect anomalies or illicit behavior. Managers are responsible for the claims processed under their watch and are held personally accountable. Third, if the backlog of a specific type of claim increases across multiple offices, INPS reassesses the weight associated with that product.

In order to test more formally whether managers game the system, I divide all the products into nine categories, and I estimate whether the number of claims processed in

[23]Notice that if the weights are measured correctly, there is no scope for gaming.

each of these categories changes differentially when a better manager takes office.[24] I interpret shifts toward a product category or away from it as evidence of gaming. More specifically, I test the presence of such behavior estimating (10) where I estimate ΔM_i with $\widehat{\Delta M}_i = \hat{\theta}_{incoming} - \hat{\theta}_{outgoing}$ and I control for demand for the nine broad product categories. Controlling for demand is important in this context because it controls for the shifts toward some products dictated by external factors that are not under the control of the manager. Figures IVa and IVb show that there is no evidence of productive managers shifting the production mix. Overall, I find no evidence consistent with managers gaming the system.

VI.C Quartile Specification

Model (11) assumes that productivity gains are a linear function of changes in managerial talent. In this subsection, I relax this assumption and propose an alternative exercise where I divide my measure of changes in manager ability (i.e., $\widehat{\Delta M}_i^L$) into four quartiles. I estimate the impact of the change in management for each of these groups:

$$\Delta y_{it}^k = \beta_0^k + \sum_{v=2}^{4} \beta_v^k \times Q_{iv} + \Delta\tau + \psi^k \Delta X_{it} + \Delta\varepsilon_{it}^k. \tag{12}$$

Let Q_{iv} be a dummy that takes a value of one if office i belongs to the v-th quartile of the $\widehat{\Delta M}_i^L$ distribution. All the other variables are defined as above. Since I omit the first quartile, β_v^k identifies the difference between offices belonging to the v-th and the first quartile at event time k. I iterate over the values of k as described in Section V.

Figure Va shows that the higher the treatment intensity (i.e. $\widehat{\Delta M}_i^{L,k}$), the larger the treatment effect. Figures Vb and Vc display the same pattern. These findings suggest that the linear specification is not a poor approximation of the data.

VII Counterfactual Exercises

In this section, I discuss how governments could use these findings to improve public service provision. The results presented in the previous sections imply that empowering managers to make payroll decisions would generate large efficiency gains for public sector offices. As passing such drastic civil service reforms may not be feasible, I use my estimates to construct counterfactual exercises that evaluate the efficiency gains

[24] The categories are defined as follows: 1. Insurance and pensions; 2. Subsidies to the poor; 3. Services to contributors; 4. Social and medical services; 5. Specialized products; 6. Archives and data management; 7. Administrative cross-checks; 8. Checks on benefits; 9. Appeals.

from alternative managerial allocation schemes.

I assume that the social planner maximizes the aggregate agency output and that she cannot directly influence the permanent office component of productivity (α_i) and the number of workers assigned to the office (w_i).[25] She can, however, hire and fire managers and freely assign them to offices. Let $P_i\left(\alpha_i, \theta_{m(i)}\right)$ be the average productivity at office i, which depends on some time-invariant office characteristics (α_i) and on the ability of the manager ($\theta_{m(i)}$). Let w_i represent the full-time equivalent number of workers assigned to the office. Assuming $\ln P_i\left(\alpha_i, \theta_{m(i)}\right) = \alpha_i + \theta_{m(i)}$, the planner's objective function is:

$$\max_{\underline{\theta}} \sum_i \gamma_i e^{\theta_{m(i)}}, \tag{13}$$

where $\gamma_i = e^{\alpha_i} w_i$. I choose an additively separable model as I have shown that it approximates the data fairly well and it fits naturally in the framework I have developed.

I consider four counterfactual policies that the social planner can implement: 1) She can maximize (13) by reassigning existing managers to offices; 2) She can fire the bottom 20% of managers and substitute them with the median manager (but allocate them as in the current environment); 3) She can implement both policies at once; and 4) She can randomly assign managers to offices. Being able to reassign managers within but not across connected sets puts additional constraints on (13).[26] As the optimal solution to the constrained problem can never exceed that of the unconstrained one, the estimated impact of the first and third intervention represent a lower bound on the true effect.

$P_i w_i$ is twice differentiable and supermodular, therefore the optimal allocation is an assortative matching equilibrium where the best managers are sent to offices which are both productive and large (Becker, 1974), where γ_i implicitly weights these two dimensions. Such an allocation exacerbates productivity inequality across sites. Traditionally, the argument about equalizing quality of services across offices relies on the idea that beneficiaries should not receive a different treatment depending on where they live. As claims can be electronically redistributed across sites at virtually no cost and processed anywhere, it is unclear why productivity should be equalized across offices in this setting.

[25] Unlike managers, production line workers enjoy strong job protection and can not be forced to move from one site to another. Moreover, financial incentives play a limited role as they are a function of group performance and promotions depend on seniority.

[26] As my sample contains multiple connected sets, I implement each policy within each connected set. Connected sets reflect the mobility patterns in my sample and often overlap with broad geographical regions (Appendix C).

Table X reports the efficiency gains from alternative managerial allocation schemes. If the social planner reassigns managers using the optimal allocation rule, aggregate productivity increases by at least 6.9%. If instead, she fires the bottom 20%, aggregate productivity raises only by 2.9%. In this setting, the first policy is more effective than the second because there are strong complementarities between managerial talent and the permanent component of office productivity and, as the most productive managers are currently allocated to the least productive offices (Table VI), there is quite some scope for reallocation. Implementing both these policies simultaneously increases aggregate productivity by at least 7.4%, corroborating the finding that managerial allocation is key in this context. Last but not least, I randomly reassign managers to offices and I take the average of the implied productivity gains and losses over 1,000 iterations. If managers were randomly reassigned, aggregate productivity would increase by 2%. This is because random assignment moves the allocation closer to the socially optimal one by undoing the negative assortative matching equilibrium.[27] In practice, reallocating managers across sites is feasible and INPS has experimented with it in 2019 when Tridico was appointed to INPS President. However, hiring and firing top-level officials is extremely challenging in the Italian legal framework and it is unlikely to be a viable policy option.

VIII Conclusion

Following a longer tradition of productivity studies in the private sector, this paper is the first to estimate the productivity impacts of managers in the public sector bureaucracy. Bureaucracy represents a large part of modern economies. But because prices, profit-motive, and competition are not driving forces in the public sector, simply measuring productivity when government has many objectives has limited previous research to simpler settings where agents' objectives are less subjective (e.g. schools).

I overcome this challenge by using novel administrative data containing an output-based measure of productivity of public offices. Using the rotation of managers as a source of quasi-experimental variation, I find that public sector managers have a quantitatively meaningful impact on the productivity of the offices they oversee, even with only limited ability to make personnel decisions. In this context, it is all the more surprising that manager-driven productivity increases are mainly driven by the exit of older workers. A good manager also sustains production levels without resorting to hiring or overtime to compensate for the decrease in full-time equivalent employment. These

[27] Table H.IV report the estimates computed on the largest connected set (Online Appendix H). Reassuringly, the pattern of results is unchanged.

findings are consistent with anecdotes suggesting that senior workers leave when better managers reassign more prestigious and better compensated tasks to other employees that are more junior but more productive.

These results speak directly to the debate on downsizing the public sector and they suggest that managerial talent can go a long way in sustaining adequate public service provision in a context where the workforce is shrinking. These results also imply that broadly empowering managers to make payroll decisions would generate large efficiency gains for public sector offices. As passing such drastic civil service reforms may not be feasible, in the final part of the paper, I discuss how governments could use these findings to improve public service provision by evaluating alternative managerial allocation schemes. I estimate that the agency can substantially increase output by reallocating managers across sites.

These findings are broadly relevant for agencies where officers primarily engage in back-office duties and process paperwork, the vast majority of the modern bureaucracy. These include the Social Security Administration, taxation authorities (e.g., IRS and state revenue agencies), immigration agencies (e.g., USCIS), national and local agencies dispensing welfare transfers such as TANF or food stamps, and offices granting licenses and permits (e.g., DMV, State Corporation Commission, Board of Accountancy etc).

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Adhvaryu, Achyuta, Vittorio Bassi, Anant Nyshadham, and Jorge A Tamayo, "No Line Left Behind: Assortative Matching Inside the Firm," Working Paper 27006, National Bureau of Economic Research April 2020.

Andrews, M. J., L. Gill, T. Schank, and R. Upward, "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias," *Economics Letters*, 2012, *117* (3), 824 – 827.

_ , _ , Thorsten Schank, and Richard Upward, "High wage workers and low wage firms: negative assortative matching or limited mobility bias?," *Journal of the Royal Statistical Society Series A*, 2008, *171* (3), 673–697.

Autor, David H., "The "task approach" to labor markets: an overview," *Journal for Labour Market Research*, Sep 2013, *46* (3), 185–199.

— , **Lawrence F. Katz, and Melissa S. Kearney**, "The Polarization of the U.S. Labor Market," *American Economic Review*, May 2006, *96* (2), 189–194.

— , **Nicole Maestas, Kathleen J Mullen, and Alexander Strand**, "Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants," Working Paper 20840, National Bureau of Economic Research January 2015.

Baltrunaite, Audinga, Giulia Bovini, and Sauro Mocetti, "Opening the Black-Box of Managerial Talent," Working Paper January 2020.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, and Raffaella Sadun, "CEO Behavior and Firm Performance," *Journal of Political Economy*, 2020, *128* (4), 1325–1369.

Becker, Gary S., *A Theory of Marriage*, University of Chicago Press,

Bennedsen, Morten, Francisco Perez-Gonzalez, and Daniel Wolfenzon, "Do CEOs Matter?," Working Paper September 2010.

— , — , **and** — , "Estimating the Value of The Boss: Evidence from CEO Hospitalization Events," Working Papers August 2011.

Bertrand, Marianne and Antoinette Schoar, "Managing with Style: The Effect of Managers on Firm Policies," *The Quarterly Journal of Economics*, 2003, *118* (4), 1169–1208.

— , **Robin Burgess, Arunish Chawla, and Guo Xu**, "The Glittering Prizes: Career Incentives and Bureaucrat Performance," *The Review of Economic Studies*, 05 2019, *87* (2), 626–655.

Best, Michael, Jonas Hjort, and David Szakonyi, "Individuals and Organizations as Sources of State Effectiveness," Working paper October 2019.

Bloom, Nicholas and John Van Reenen, "Measuring and Explaining Management Practices Across Firms and Countries," *The Quarterly Journal of Economics*, 2007, *122* (4), 1351–1408.

— , **Aprajit Mahajan, David McKenzie, and John Roberts**, "Do Management Interventions Last? Evidence from India," Working Paper 24249, National Bureau of Economic Research January 2018.

— , **Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts**, "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

— , **Renata Lemos, Raffaella Sadun, and John Van Reenen**, "Does Management Matter in schools?," *The Economic Journal*, 2015, *125* (584), 647–674.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar, "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico," *Journal of Political Economy*, 2018, *126* (2), 635–687.

Card, David, Jorg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 2013, *128* (3), 967–1015.

Cartocci, Roberto, *Mappe del tesoro*, Bologna: Il Mulino, 2007.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Productivity Dispersion in Medicine and Manufacturing," *American Economic Review*, May 2016, *106* (5), 99–103.

Choudhury, Prithwiraj, Tarun Khanna, and Christos A Makridis, "Do Managers Matter? A Natural Experiment from 42 RandD Labs in India," *The Journal of Law, Economics, and Organization*, 12 2019, *36* (1), 47–83.

Eeckhout, Jan and Philipp Kircher, "Identifying Sorting—In Theory," *The Review of Economic Studies*, 03 2011, *78* (3), 872–906.

Finan, F., B.A. Olken, and R. Pande, "Chapter 6 - The Personnel Economics of the Developing State," in Abhijit Vinayak Banerjee and Esther Duflo, eds., *Handbook of Economic Field Experiments*, Vol. 2 of *Handbook of Economic Field Experiments*, North-Holland, 2017, pp. 467 – 514.

Gaure, Simen, "Practical Correlation Bias Correction in Two-way Fixed Effects Linear Regression," Memorandum 21/2014, Oslo University, Department of Economics 2014.

Giorcelli, Michela, "The Long-Term Effects of Management and Technology Transfers," *American Economic Review*, January 2019, *109* (1), 121–52.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "The Role of Social Capital in Financial Development," *American Economic Review*, June 2004, *94* (3), 526–556.

Hoffman, Mitchell and Steven Tadelis, "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis," Working Paper 24360, National Bureau of Economic Research February 2018.

Holmstrom, Bengt and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law Economics and Organization*, 1991, *7*, 24–52.

Janke, Katharina, Carol Propper, and Raffaella Sadun, "The Impact of CEOs in the Public Sector: Evidence from the English NHS," Working Paper May 2019.

Khan, Adnan Q., Asim Ijaz Khwaja, and Benjamin A. Olken, "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings," *American Economic Review*, January 2019, *109* (1), 237–70.

Lazear, Edward P., Kathryn L. Shaw, and Christopher T. Stanton, "The Value of Bosses," *Journal of Labor Economics*, 2015, *33* (4), 823–861.

Limodio, Nicola, "Bureaucrat Allocation in the Public Sector: Evidence from the World Bank," Working Paper October 2019.

Lucas, Robert E., "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 1978, *9* (2), 508–523.

Perez-Gonzales, Francisco, "Inherited Control and Firm Performance," *American Economic Review*, December 2006, *96* (5), 1559–1588.

Rasul, Imran and Daniel Rogger, "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *The Economic Journal*, 2018, *128* (608), 413–446.

_ , _ , and Martin Williams, "Management and Bureaucratic Effectiveness: Evidence from the Ghanaian Civil Service," Working Paper July 2019.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter, "Firming Up Inequality," Working Paper 21199, National Bureau of Economic Research May 2015.

Stinebrickner, Todd R, Ralph Stinebrickner, and Paul J Sullivan, "Job Tasks and the Gender Wage Gap among College Graduates," Working Paper 24790, National Bureau of Economic Research July 2018.

Syverson, Chad, "Product Substitutability and Productivity Dispersion," *The Review of Economics and Statistics*, 2004, *86* (2), 534–550.

_ , "What Determines Productivity?," *Journal of Economic Literature*, June 2011, *49* (2), 326–65.

Tsai, Thomas C., Ashish K. Jha, Atul A. Gawande, Robert S. Huckman, Nicholas Bloom, and Raffaella Sadun, "Hospital Board And Management Practices Are Strongly Related To Hospital Performance On Clinical Quality Metrics," *Health Affairs*, 2015, *34* (8), 1304–1311. PMID: 26240243.

Xu, Guo, "The Costs of Patronage: Evidence from the British Empire," *Forthcoming American Economic Review*, 2018.

Tables

Table I: Manager Characteristics

	(1) Full Sample	(2) Movers
Demographic Characteristics		
Male	0.63	0.71
Age	54.42	52.98
Experience Public Sector	26.85	24.68
Region of Birth		
North-East	0.13	0.09
North-West	0.12	0.13
Center	0.15	0.12
South or Islands	0.59	0.67
Abroad	0.01	0.00
Highest Educational Attainment		
High-School Diploma	0.22	0.10
Econ, Business, and Admin	0.13	0.11
Sci, Engen, Math, and Stat	0.05	0.07
Social Sciences and Humanities	0.20	0.22
Law	0.33	0.44
Missing Educ	0.06	0.08
Observations	851	207

Note: The table reports the summary statistics of manager characteristics. The statistics are computed over the full sample of managers in column (1), and over the subsample of movers in column (2). Movers are defined as those managers who oversee at least two offices over my sample period. Experience in the public sector is defined as the number of years since the manager was first hired in any public sector institution.

Table II: Characteristics of Social Security Offices

	(1) Full Sample	(2) Main Offices	(3) Local Branches
Productivity	94.56	103.65	91.72
Output (Thousands)	10.24	29.18	4.33
FTE	39.95	115.39	16.41
Hours	31.66	91.76	12.91
Training	0.62	1.73	0.28
Overtime	0.70	2.10	0.26
Absenteeism Rate	0.21	0.21	0.21
Quality	100.52	101.31	100.27
Backlog (Thousands)	54.24	197.68	9.48
Demand (Thousands)	68.02	220.55	20.42
Hires	0.05	0.13	0.02
Separations	0.50	1.53	0.17
Fires	0.00	0.01	0.00
Inbound Transfers	0.72	2.08	0.29
Outbound Transfers	0.30	0.63	0.20
Retirement	0.24	0.72	0.09
Divorce	0.87	0.88	0.87
Blood Donations	0.03	0.03	0.03
Average Worker Age	52.57	52.69	52.53
Fraction Female Workers	0.58	0.57	0.58
Fraction Female Top-Officials	0.39	0.40	0.39
N	13212	3142	10070
Number of Managers	851	221	638
Number of Offices	494	111	383

Note: The table reports the summary statistics for social security offices. All statistics are calculated across office-quarter observations. The statistics are computed over the full sample of offices in column (1), and over the subsample of main offices and local branches in column (2) and (3), respectively. The number of office-quarter observations for the full sample, main offices and local branches for quality are 10943, 2658, and 8285, respectively; for divorce these statistics are 11052, 2622, and 8430, and for blood donations they are 11104, 2648, and 8456. Output, demand, and backlog are measured in thousands of hours, while FTE, training, hours, and overtime are measured in full-time equivalent units.

Table III: Sample Characteristics

	(1) Full Sample	(2) Balanced-Analysis Sample
# Managers	851	601
# Offices	494	282
# Managers >1 Office over the Sample Period	207	184
# Offices >1 Manager over the Sample Period	404	282
# Connected Sets	276	143
# Events	635	318
# Events in Main Offices	226	80
# Events in Local Branches	409	238
N	13,212	8,165

Note: The table reports the sample characteristics for the full sample of offices in column (1) and for the balanced-analysis sample in column (2). The latter includes the subset of offices for which I observe the outgoing manager being in charge for at least four quarters before the change in leadership and the incoming manager being assigned to the office for at least six quarters after that. "# Managers >1 Office over the Sample Period" represents the number of managers who serve in at least two sites over my sample period. "# Offices >1 Manager over the Sample Period" represents the number of offices that experience at least one change in leadership over my sample period. Each office has only one manager at each point in time but may have multiple managers over my sample period. Events are defined as changes in leadership. "N" represents the number of office-quarter observations.

Table IV: Analysis of Variance of Yearly Measures of Productivity per Worker at INPS Offices

	(1) Ln(P)	(2) Ln(P)	(3) Ln(P)	(4) Ln(P)	(5) Ln(P)
N	3,316	3,316	3,316	3,316	3,316
R sq.	0.325	0.727	0.835	0.789	0.839
Adj. R sq.	0.324	0.679	0.762	0.720	0.765
Time FE	Yes	Yes	Yes	Yes	Yes
Office FE	No	Yes	Yes	No	No
Manager FE	No	No	Yes	Yes	No
Manag-by-Office FE	No	No	No	No	Yes
Pvalue			0.000	0.000	

Note: This table investigates how much of the variance in log productivity is explained by the office, manager, and time components in the full sample. N represents the number of office-year observations. The p-value at the bottom of the table tests the null hypothesis that manager effects are jointly zero.

Table V: Can Observables Predict Incoming Manager FE?

	(1) Manager FE		(2) Change in Manager FE	
Main Office	-0.612	(0.095)	0.029	(0.040)
North or Center	0.161	(0.071)	-0.017	(0.027)
P_{2011}	0.000	(0.002)	-0.002	(0.001)
Y_{2011}	0.000	(0.000)	0.000	(0.000)
FTE_{2011}	-0.000	(0.003)	-0.001	(0.001)
P_{t-1}	0.001	(0.002)	-0.001	(0.001)
P_{t-2}	-0.001	(0.001)	-0.001	(0.001)
P_{t-3}	0.000	(0.001)	-0.000	(0.001)
P_{t-4}	0.000	(0.001)	-0.002	(0.001)
Y_{t-1}	-0.000	(0.000)	-0.000	(0.000)
Y_{t-2}	0.000	(0.000)	-0.000	(0.000)
Y_{t-3}	0.000	(0.000)	0.000	(0.000)
Y_{t-4}	-0.000	(0.000)	0.000	(0.000)
FTE_{t-1}	0.002	(0.002)	-0.002	(0.001)
FTE_{t-2}	0.001	(0.002)	0.001	(0.002)
FTE_{t-3}	0.001	(0.001)	-0.000	(0.001)
FTE_{t-4}	-0.003	(0.002)	0.001	(0.002)
Growth Rate P - 3 q	0.058	(0.078)	-0.032	(0.065)
Growth Rate P - 2 q	-0.008	(0.100)	0.061	(0.095)
Growth Rate P - 1 q	-0.125	(0.115)	-0.007	(0.078)
Growth Rate Y - 3 q	0.021	(0.077)	0.086	(0.059)
Growth Rate Y - 2 q	-0.002	(0.040)	-0.033	(0.043)
Growth Rate Y - 1 q	0.075	(0.101)	-0.014	(0.075)
Growth Rate FTE - 3 q	-0.084	(0.156)	-0.033	(0.149)
Growth Rate FTE - 2 q	0.138	(0.175)	0.074	(0.146)
Growth Rate FTE - 1 q	-0.115	(0.207)	0.093	(0.186)
N	521		521	
R sq.	0.482		0.605	
Connected Set FE	Yes		Yes	
P-value (All)	0.000		0.000	
P-value (Growth Rates)	0.807		0.864	

Note: This table investigates the extent to which office characteristics predict the incoming manager FE or the change in manager FE. The sample includes all events balanced on $[-4,0]$. The dependent variable in column (1) is the manager effect estimated using (3), while in column (2) it is the difference between the estimated effect of the incoming and outgoing manager. P, Y, and FTE represent productivity, output, and full-time equivalent employment, respectively. t indexes the time of the event, and P_{2011} represents the office productivity at baseline. Y_{2011} and FTE_{2011} are defined accordingly. "Growth Rate P - x q" is defined as the productivity growth rate of office i between $t-(x+1)$ and $t-1$. "N" represents the number of office-quarter observations. "P-value (All)" and "P-value (Growth Rates)" are the p-values for the null hypothesis that all regressors of interest are jointly statistically significant and that the growth rates are jointly significant, respectively. All regressions include connected set fixed effects. Standard errors are clustered at the office level and are reported in parentheses.

Table VI: Biased-Corrected Variance-Covariance Decomposition

	(1) Component	(2) Share of Total
Var(Ln(P))	0.1106	100 %
Var(Manager)	0.0102	9.22%
Var(Office)	0.0319	28.84 %
Var(Time)	0.0408	36.89%
Cov(Manager, Office)	-0.0096	-8.68%
Cov(Time, Manag. + Office)	0.0015	1.39%
N	2,735	

Note: This table presents the bias-corrected variance-covariance decomposition of log productivity in the largest connected set. The model includes dummies for manager, office, and quarter fixed effects.

Table VII: Manager Effects and Observable Characteristics

	(1) Manager FE	
Male	-0.06	(0.03)
Experience in the Public Sector	0.02	(0.01)
Experience in the Public Sector Squared	-0.00	(0.00)
Center	0.07	(0.06)
South or Islands	0.02	(0.04)
North-West	0.00	(0.05)
Abroad	0.00	(0.06)
Econ, Business, and Admin	0.04	(0.05)
Sci, Engen, Math, and Stat	-0.08	(0.06)
Social Sciences and Humanities	0.02	(0.04)
Law	-0.05	(0.04)
Missing Educ	-0.09	(0.07)
N	851	
R sq.	0.45	
Connected Set FE	Yes	

Note: This table presents the correlation between the manager effect estimated from (3) and manager characteristics. These characteristics include gender, experience, the region of birth, and highest educational attainment. N represents the number of managers in my sample. "Experience in the public sector" is defined as the number of years since the manager was first hired in any public sector institution. The omitted categories are "Female", "North-East", and "No college". Controls include connected set fixed effects. Robust SE in parentheses.

Table VIII: Estimated Effects of Changes in Managerial Talent on Office Composition

k	(1) A(Retirement)	(2) A(Hires)	(3) A(Fires)	(4) A(Inbound T)	(5) A(Outbound T)
-4	0.041	0.178	0.003	0.120	-0.039
	(0.125)	(0.110)	(0.004)	(0.121)	(0.146)
-3	-0.044	0.093	0.002	0.069	-0.111
	(0.088)	(0.090)	(0.004)	(0.090)	(0.133)
-2	-0.055	0.034	0.003	0.068	-0.122
	(0.059)	(0.073)	(0.004)	(0.082)	(0.112)
0	0.301	0.024	-0.008	0.031	-0.023
	(0.085)	(0.018)	(0.010)	(0.159)	(0.053)
1	0.393	0.027	-0.056	-0.033	-0.010
	(0.100)	(0.033)	(0.031)	(0.163)	(0.067)
2	0.381	0.024	-0.049	-0.196	-0.142
	(0.105)	(0.033)	(0.040)	(0.172)	(0.097)
3	0.392	0.006	-0.063	-0.327	-0.234
	(0.117)	(0.038)	(0.045)	(0.174)	(0.114)
4	0.438	-0.015	-0.040	-0.385	-0.231
	(0.116)	(0.039)	(0.038)	(0.171)	(0.123)
5	0.413	0.005	-0.061	-0.403	-0.303
	(0.120)	(0.039)	(0.041)	(0.174)	(0.125)
6	0.399	-0.082	-0.059	-0.537	-0.405
	(0.125)	(0.057)	(0.042)	(0.184)	(0.135)
N	318	318	318	318	318
Time FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the reduced-form impacts of managerial talent on office composition. More specifically, it reports the coefficients π_1^k obtained estimating (11) on the balanced-analysis sample. N represents the number of office-quarter observations. The dependent variable, cumulative y_{it}, is reported at the top of each column and $A(.)$ is short for asinh. All models include time fixed effects, main effects and two-way interactions between a dummy for Center-North, a dummy for main offices, a set of dummies for quartiles of baseline productivity, as well as time effects interacted with the dummy for Center-North. k indexes event time. Each coefficient is obtained from a separate regression. Bootstrapped standard errors are reported in parentheses.

Table IX: Estimated Effects of Changes in Managerial Talent on Quality, Backlog, and Demand

k	(1) Ln(Quality)	(2) Ln(Backlog)	(3) Ln(Demand)
-4	-0.036	0.150	0.124
	(0.025)	(0.117)	(0.114)
-3	-0.029	0.140	0.113
	(0.024)	(0.100)	(0.129)
-2	-0.049	0.053	0.071
	(0.026)	(0.081)	(0.105)
0	-0.058	-0.129	0.100
	(0.041)	(0.090)	(0.124)
1	0.064	-0.077	0.282
	(0.067)	(0.099)	(0.129)
2	-0.091	-0.248	-0.275
	(0.054)	(0.116)	(0.189)
3	0.049	-0.345	-0.075
	(0.041)	(0.120)	(0.130)
4	0.055	-0.258	0.061
	(0.035)	(0.172)	(0.165)
5	0.010	-0.071	0.176
	(0.031)	(0.160)	(0.175)
6	-0.008	-0.145	0.171
	(0.068)	(0.178)	(0.143)
N	300	318	313
Time FE	Yes	Yes	Yes

Note: This table reports the reduced-form impacts of managerial talent on quality, backlog, and demand. More specifically, it reports the coefficients π_1^k obtained estimating (11) on the balanced-analysis sample. N represents the number of office-quarter observations. The dependent variable, y_{it}, is reported at the top of each column. All models include time fixed effects, main effects and two-way interactions between a dummy for Center-North, a dummy for main offices, a set of dummies for quartiles of baseline productivity, as well as time effects interacted with the dummy for Center-North. k indexes event time. Each coefficient is obtained from a separate regression. Bootstrapped standard errors are reported in parentheses.

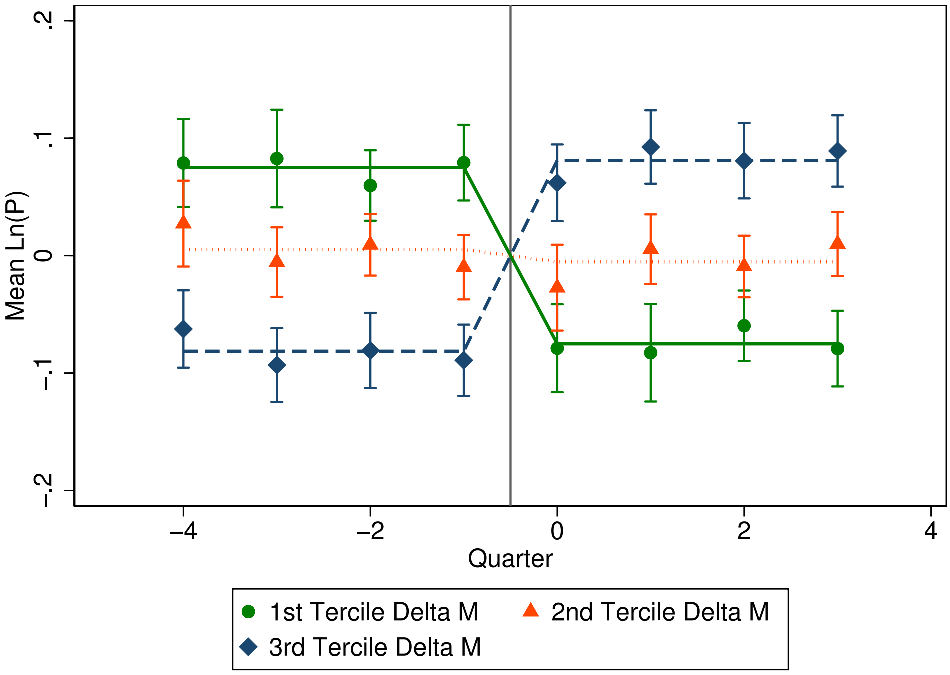
Table X: Counterfactual Exercises

		ΔY
Policy 1:	Reassign	6.9%
Policy 2:	Replace bottom 20%	2.9%
Policy 3:	Replace bottom 20% + Reassign	7.4%
Policy 4:	Random allocation	2%

Note: This table reports the counterfactual exercises that illustrate the efficiency gains from alternative managerial allocation schemes. I consider four counterfactual policies that the social planner can implement. Policy 1: she can reallocate existing managers according to the optimal rule. Policy 2: she can fire the bottom 20% of top-level bureaucrats and substitute them with the median manager (but allocate them as in the current environment). Policy 3: she can implement both Policy and 2. Policy 4: she can randomly assign existing managers to offices (1,000 iterations). The sample includes all the connected sets with at least five managers. Figure L.IIa, L.IIb, and L.IIc in the Online Supplement illustrate these concepts graphically.

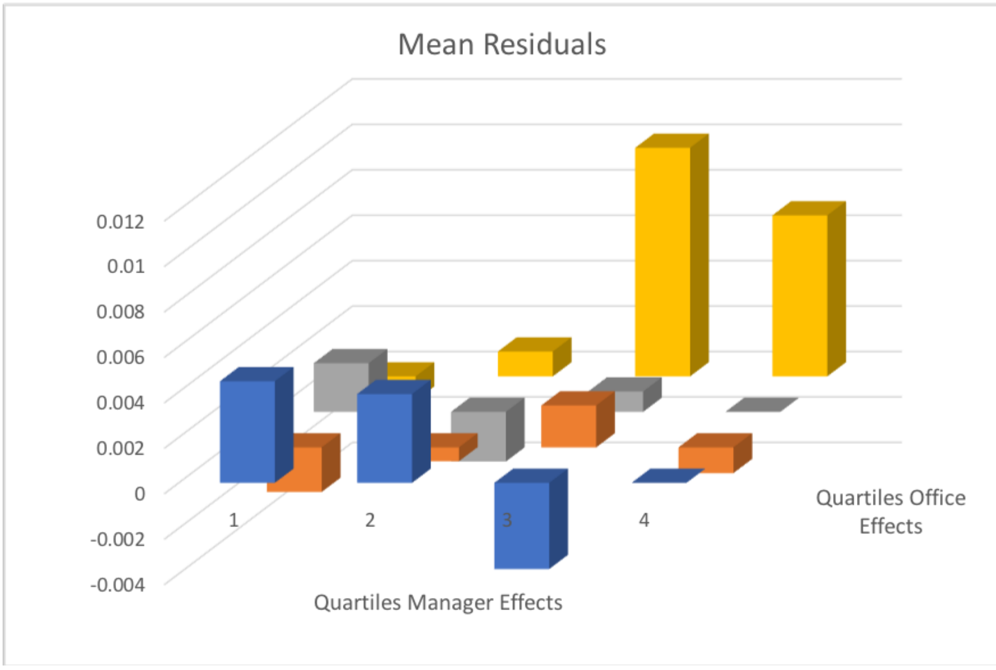
Figures

Figure I: Mean Productivity for Offices which Experience a Change in Leadership Classified by Tercile of Changes in Manager Effects



Note: This figure reports the event study for the mean (trend-adjusted) log office productivity and the associated 95% confidence intervals for three types of office transition associated with a change in leadership. These three types of transitions are an overall increase in manager ability (blue diamonds), a decrease in management quality (green circles) or no significant change (orange triangles). $\widehat{\Delta M_i}$ represents the change in the estimated manager fixed effects. The x-axis indexes event time.

Figure II: Mean Residual by Manager/Office Quartiles, 2011q1-2017q2

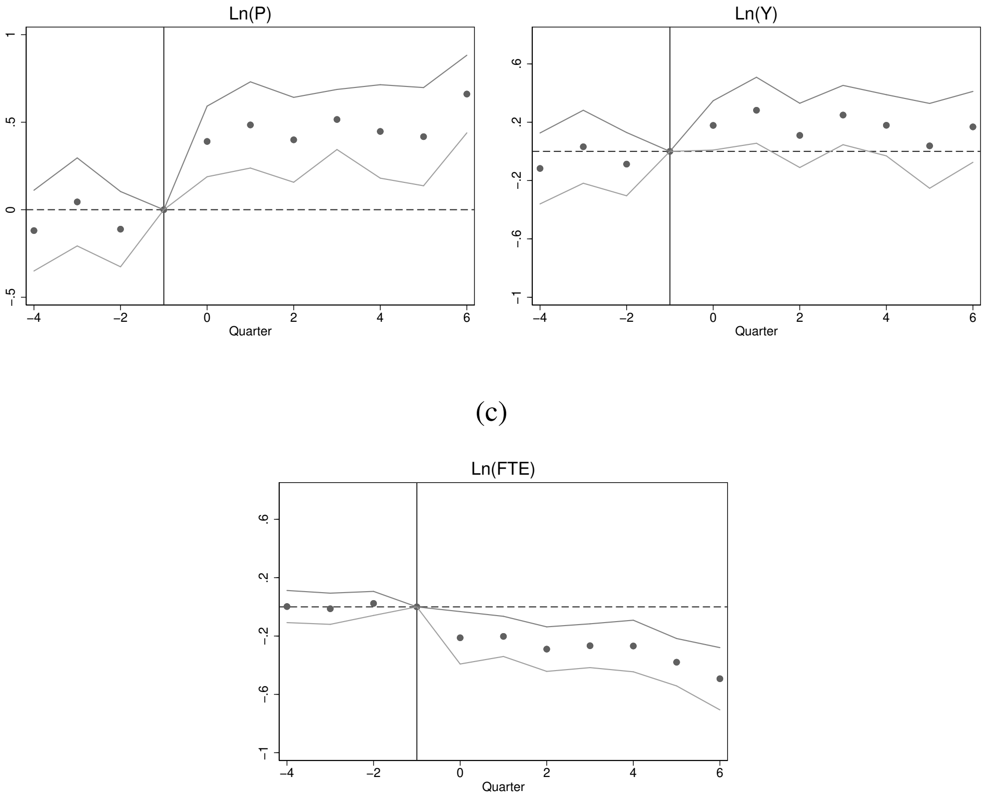


Note: This Figure shows mean residuals from model (3) with cells defined by quartiles of estimated manager effect, interacted with quartiles of estimated office effect.

Figure III: Decomposition of Productivity Effects

(a) (b)

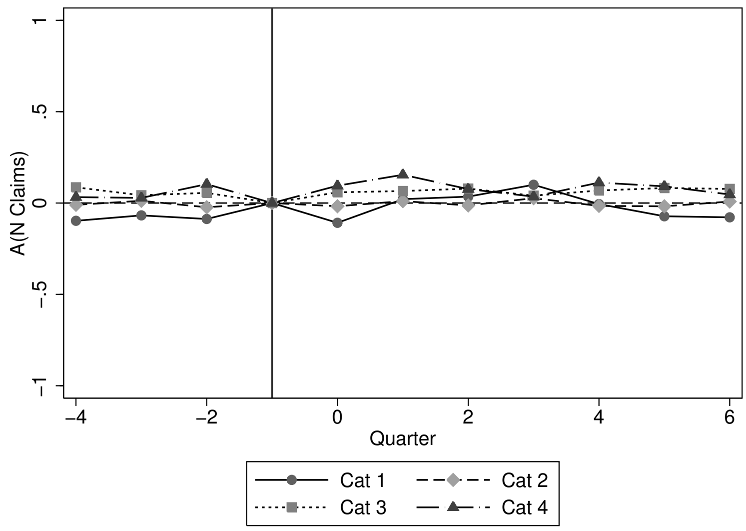
(c)



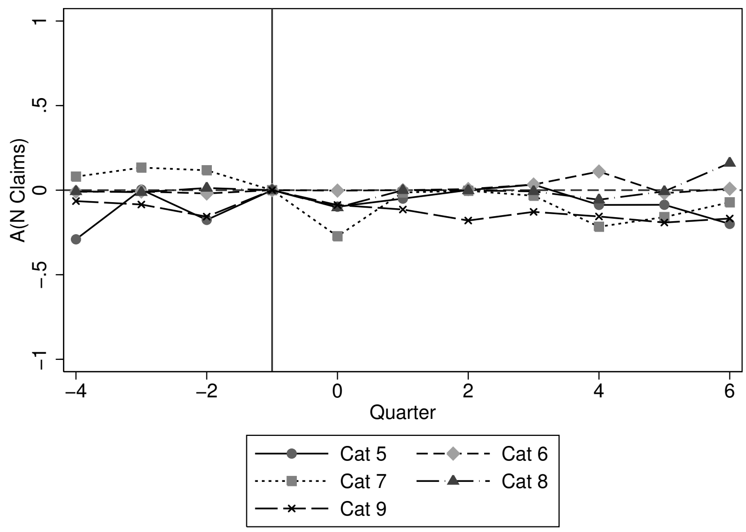
Note: Panels (a)-(c) report the regression coefficients and the associated 95% bootstrapped confidence intervals that identify the reduced-form impacts of managerial talent, i.e., $\hat{\pi}_1^k$ from (11). The coefficients at $k = -1$ are normalized to zero. The dependent variables are log productivity (Panel a), log output (Panel b), and log full-time equivalent employment (Panel c). The x-axis indexes event time. Refer to Table H.III in the Online Appendix for these results in table format.

Figure IV: More Productive Managers Do Not Shift Production

(a)



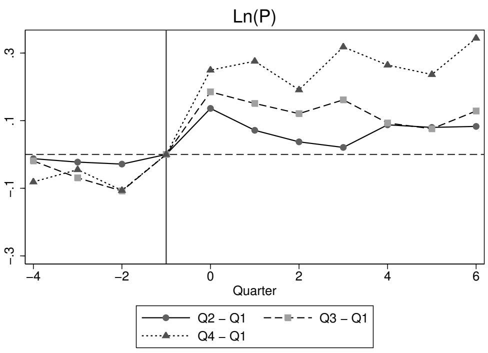
(b)



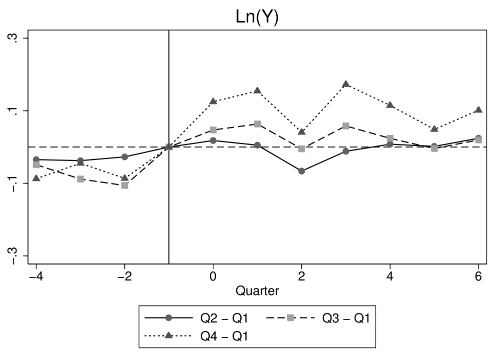
Note: Panels (a) and (b) report the regression coefficients that identify the reduced-form impacts of managerial talent on nine claim categories. The coefficients at $k = -1$ are normalized to zero. The dependent variables represent the number of claims belonging to product category n processed by each office. n=1, 2, ..., 9 and the categories are defined as follows: 1: Insurance and pensions, 2: Subsidies to the poor, 3: Services to contributors, 4: Social and medical services, 5: Specialized products, 6: Archives and data management, 7: Administrative cross-checks, 8: Checks on benefits, 9: Appeals. The x-axis indexes event time.

Figure V: Estimated Effect of Leadership Changes by Quartile of Changes in Manager Effects

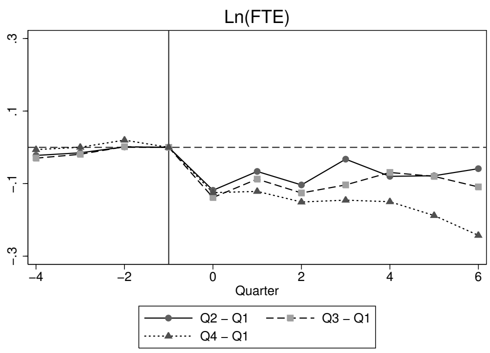
(a)



(b)



(c)



Note: Panels (a)-(c) report the regression coefficients that identify the reduced-form impacts of managerial talent in the quartile specification, i.e., $\hat{\beta}_v^k$ from (12). The coefficients at $k = -1$ are normalized to zero. The dependent variables are log productivity (Panel a), log output (Panel b), and log full-time equivalent employment (Panel c). As I omit the first quartile of $\widehat{\Delta M_i^{L,k}}$, the estimated coefficients identify the difference in the outcome of interest between the offices that are in the j-th quartile vs those in the first quartile, i.e., $Q_j - Q_1$ where $j = 2, 3, 4$. The x-axis indexes event time.

Online Appendix for
“Managers and Productivity in the Public Sector”

by Alessandra Fenizia

Appendix A : Excluding the Front Office

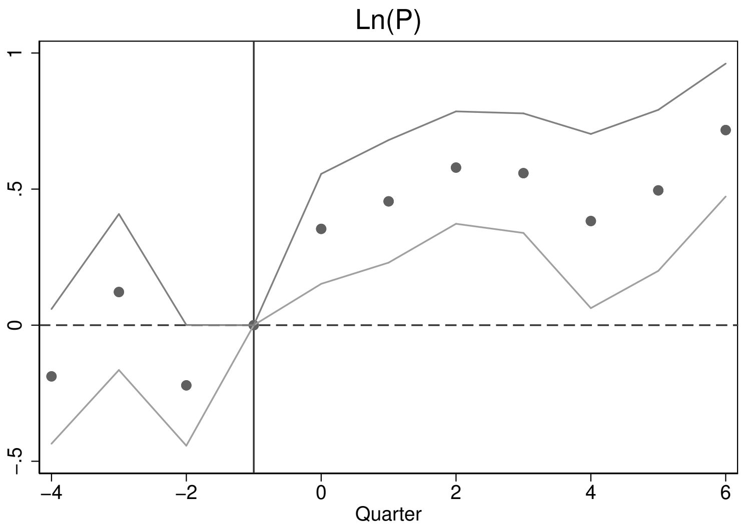
Front offices are holdovers from a time when people applied for benefits in person. These days beneficiaries either apply online or through tax consultants (*Centri di Assistenza Fiscale*). They can learn about application procedures and eligibility criteria on the INPS website and can check the status of their application through the online portal. As such, front office operations are quite limited. Front offices provide information that can be found on the INPS website to beneficiaries who are not non-technologically savvy. For example, a potential beneficiary may walk into the front-office to learn whether she is eligible for disability insurance and what documents she should include in her application. Alternatively, a citizen may have applied for a pension and may want to inquire about the status of her application. Finally, a beneficiary may be eligible for multiple welfare programs and may want to know whether applying to one program precludes her from applying to the others.

Measuring productivity in any customer facing setting is challenging. In contrast with the other elements of the productivity indices, INPS measures front office output using the inputs—the amount of time employees spend on front office duties. Thus, the measure bluntly captures the value of staffing the office without adjusting for the number of customers served or the complexity of their demands.

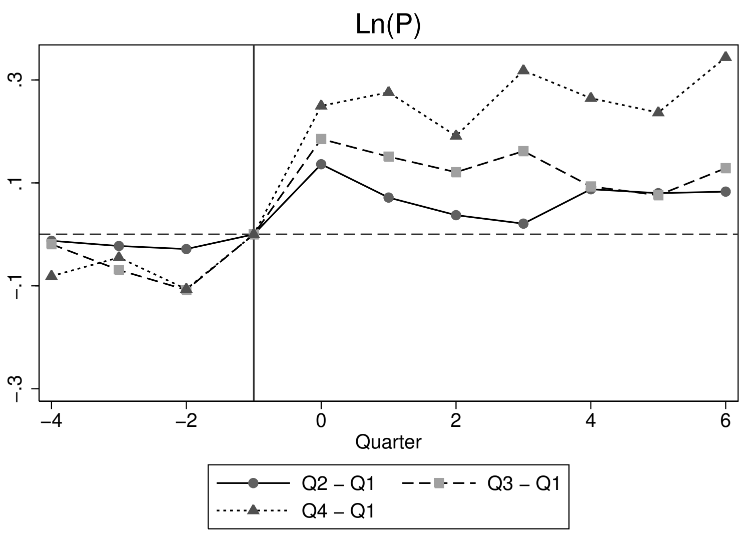
One may be concerned that because employees may sit idle in periods of low demand, my productivity measures may not be purely capturing output. Nonetheless, the difficulty of measuring front office productivity does not affect my estimates of manager productivity. First, front-office services depend on the demographic composition of the catchment area and macroeconomic shocks and they are unlikely to be correlated with manager effectiveness. Second, I construct an alternative measure of office productivity ($\ln P_{it}^{c}$) that does not include front-office output and I estimate (11) and (12) using this alternative measure. The pattern of results is unchanged (Figure A.Ia and A.Ib). As an additional robustness check, I re-estimate (3) using $\ln P_{it}^{c}$ as my dependent variable, and I correlate manager fixed effects obtained from this model with those from (3). The correlation coefficient is 92%. Overall, I conclude that changes in front office operations are not driving my results.

Figure A.I: Robustness to the Exclusion of Front Office Output

(a)



(b)



Note: This figure mirrors Figures IIIa and Va and reports the reduced-form effects of managerial talent on office productivity measured without front office output. Panel (a) reports the regression coefficients and the associated 95% bootstrapped confidence intervals for the linear specification, i.e., $\hat{\pi}_1^k$ from (11). Panel (b) reports the regression coefficients for the quartile specification, i.e., $\hat{\beta}_v^k$ from (12). The estimated coefficients identify the difference in the outcome of interest between the offices that are in the j-th quartile vs those in the first quartile of ΔM_i^L, i.e., $Q_j - Q_1$ where $j = 2,3,4$. The coefficients at $k = -1$ are normalized to zero. The x-axis indexes event time.

Appendix B : Equalizing Workloads Across Offices

INPS optimizes its resource allocation to meet a demand that has a strong seasonal component and often exhibits idiosyncratic shocks. Given the many constraints related to hiring and firing bureaucrats, the Social Security Agency can either reallocate employees or workloads across sites. While reassigning workers to offices seems an appealing strategy, in practice, this is often not feasible. Workers can not be forced to move from one site to another against their will and those who choose to move are relatively few as documented by the low number of inbound and outbound transfers in Table II. To equalize workloads across offices and use resources effectively, INPS highly encourages managers to trade claims (*principio di sussidiarietà*). Managers facing low demand are instructed to contact managers in high-demand offices and ask to be transferred a share of their claims. If the two managers agree on the trade, claims are transferred electronically and they count toward the production of the office that processes them. Equalizing workloads is beneficial for both offices as the traded claims increase the output of the low-demand office, and the high-demand office is not penalized by a decrease in the quality index due to longer processing time. This is also beneficial for citizens living in the catchment area of high-demand offices as their claims are processed faster than they would have been in the absence of the trade. As INPS encourages trades across sites and the pay-for-performance scheme incentivizes managers to transfer claims from low- to high-demand offices, trades are anecdotally very common.[28]

Appendix C : Sample Selection and Manager Rotation

In this section, I provide more details on the sample selection, discuss manager assignment, and document the patterns of manager rotation.

INPS is constituted of its Rome headquarters, twenty-five regional centres, 111 main offices and 383 local branches. As claim processing takes place in main offices and local branches only, these two types of offices are often referred to as "production sites". They now serve a similar purpose but local branches are holdovers from a time when people applied for benefits in person. The headquarters and the regional centres oversee the production sites but do not engage in claim processing directly. As a result, I can not construct my productivity measure for these offices and I exclude them from the analysis. My sample includes all production sites, namely 111 main offices and 383 local branches.

[28]As I do not have data on trades, I can not analyze this margin.

Next, I describe the within-office hierarchy of INPS production sites. In each local branch, a single manager oversees production line workers. Managers often task a few senior employees with additional responsibilities like monitoring specific areas of the production process or supervising other employees. These responsibilities are associated with higher status and additional monthly compensation. Main offices have the same structure as local branches with one exception: they also have two high-rank officials to whom the manager delegates some of her responsibilities. I abstract from the high-rank officials in my empirical analysis.

As discussed in Section II, while INPS has clear eligibility requirements associated with each job posting, there are no officials guidelines on how to choose among qualified applicants. The eligibility criteria vary across vacancies but typically entail a minimum level of educational attainment, job titles and in some cases passing a competitive examination. These requirements are carefully codified and followed scrupulously. However, if there are multiple candidates eligible for the same position, there are no official guidelines on how to select among them. Anecdotally, past performance is not a factor that is taken into account when evaluating candidates and more senior and experienced managers are not de facto given priority over their younger colleagues. In conclusion, human resources officers make a case-by-case assessment rather than follow a set of official or unofficial guidelines. Refer to Online Appendix D for a discussion on managers' incentives and how they can induce managers to sort into different office types.

Next, I examine the patterns of manager rotation. Table B.I reports the number of changes in leadership (events), the number of offices, and the average number of events per office by macro-region. On average each site experiences 1.2 to 1.4 events over my sample period and the distribution of these events looks fairly uniform across regions (columns 3). Columns 6 and 9 report the same statistics for main offices and local branches respectively. Main offices exhibit on average higher rotation than local branches, consistently with the mandatory rotation scheme discussed in Section II, yet no region appears to be originating a disproportionate number of moves. Managers stationed in main offices typically move to other main offices, while managers in local branches move to either other local offices or are (very rarely) promoted to main offices. No manager is demoted from a main office to a local branch over my sample period. Over the sample period, it is quite rare that a production worker is promoted to a managerial position at the same office. Not surprisingly, a large share of moves occurs within the same broad geographical region, while transfers across regions are less common. This feature is important as it describes the constraints I face in the reallocation of managers discussed in the counterfactual exercise (Section VII).

Finally, I examine the distribution of events per offices. Almost 80% of offices experience at least one rotation over my sample period. Importantly, as my identification strategy relies on manager rotation, the sites that do not experience any change in leadership do not contribute to my estimates. Overall manager rotation affects the vast majority of offices and most sites experience one to two changes in leadership. These stylized facts show that managers moves do not originate from a few peculiar sites but are a pervasive feature of the Italian Social Security Administration.

Table B.I: Manager Rotation by Macro-Region

	Full Sample			Main Offices			Local Branches		
	N Events (1)	N Offices (2)	Events /Office (3)	N Events (4)	N Offices (5)	Events /Office (6)	N Events (7)	N Offices (8)	Events /Office (9)
North-East	110	91	1.2	41	22	1.8	69	69	1.0
North-West	159	123	1.3	63	25	2.5	96	98	1.0
Center	114	92	1.2	45	25	1.8	69	67	1.0
South	155	120	1.3	56	26	2.1	99	94	1.1
Islands	97	68	1.4	21	13	1.6	76	55	1.4

Note: This table documents manager rotation. “N events” represents the number of changes in leadership over the sample period. The statistics are computed over the full sample in columns (1)–(3), over main offices in columns (4)–(6), and over local branches in columns (7)–(9).

Appendix D : Incentives

In this section, I provide an overview of the incentive-pay scheme for different classes of INPS workers and I discuss how financial and non-financial incentives may induce managers to sort into different office types.

The Incentive-Pay Scheme

As discussed in the main body of the paper, INPS implements an incentive-pay scheme to reward employees' performance. All employees' salary has a fixed component (*retribuzione tabellare*) and a bonus (*retribuzione accessoria*). The fixed component is tied to the job description. The bonus includes performance compensation and indemnities. I describe the bonus structure abstracting from indemnities for simplicity. Total bonuses are the sum of ordinary and special bonuses, which are determined by the worker's office performance relative to office targets. Performance is measured aggregating both the productivity and quality indicators to prevent managers from focusing on one dimension at the expense of the other. Production targets are defined as the maximum between the nationwide productivity indicator in the previous year and the previous year office achievement. Refer to Section I of the Online Supplements for the fomula used to compute bonuses.

Ordinary Bonuses: INPS provides quarterly bonuses to employees on the basis of the year-to-date performance of their office. The system is designed to incentivize workers to maintain high productivity and quality of service throughout the year but especially during the second semester of the year when the inflow of new claims is higher. Bonuses differ between managers and production line workers. Managers' bonuses are a function of the performance of both their office and the broader geographical region to which their office belongs. The geographic component of bonuses is intended to generate shared responsibility for public service provision in the region and ultimately foster cooperation among managers. Bonuses, also differ between managers stationed at main offices and those serving in local branches, which I turn to next.

Ordinary bonuses of managers stationed in main offices are a linear function of performance relative to production targets. In particular, 56% of the performance compensation depends on the performance of the office they are in charge of relative to its production target, 14% is based on the performance of the region her site belongs to relative to its production targets, and the remaining 30% is awarded according to performance evaluations by their superiors.[29]

Ordinary bonuses awarded to managers serving in local branches are a step function of office performance relative to its production target. Each manager is bumped up (down) one step if the region where the site is located outperforms (underperformers) its production target. One may be concerned that ordinary bonuses are more closely tied to the office performance for managers stationed in main offices than those assigned to local branches and that these difference in incentives may be driving the productivity gains I document in Section IV. In Online Appendix G I show that this is not the case and that productivity effects are, if anything, stronger in local branches than in main offices.

Ordinary bonuses for workers are an increasing step function of the performance of the region the office is located in relative to its production target. In principle, managers could differentiate bonuses between employees working at the same site, but this does not happen in practice.

[29]Managers are rated along ten dimensions: understanding of the big picture, innovation, performance relative to production targets, organizing and monitoring of their employees, customer satisfaction, networking, problem-solving, decision making, leadership, and resources utilization.

Special Bonuses: INPS also provides bonuses that directly reward improvements in the quality of the service provided. More specifically, special bonuses are an increasing linear function of the improvement in the office quality indicator relative to its previous year and are awarded to both managers and workers.

Incentives and Sorting

In this subsection, I discuss the financial and non-financial incentives for managers to sort into different office types.

INPS' incentive scheme is designed to not reward managers simply for overseeing an innately productive office. First, performance is measured relative to the office production targets. Second, bonuses are both a function of office performance level and its improvement relative to the previous year. Managers stationed at productive sites may benefit from the permanent component of office productivity, but they also have a much harder time improving the office performance relative to the prior year. Conversely, managers serving in unproductive offices may be negatively impacted by the poor overall performance of the office, but can more easily improve office productivity and quality. All managers with a given job title are paid the same fixed nominal wage, but the cost of living differs substantially across regions. Northern Italy is on average more expensive than the South and small towns are often cheaper than major cities.

Financial incentives are only a subset of the incentives managers face and not necessarily the most important ones in this context. Anecdotally, most managers move to be as close as possible to where their family lives (which typically coincides with their birthplace). Southern offices are in very high demand as 59% of managers are born in the South (Table I), but only 38% of the offices are located in this region (Table B.I). While career concerns represent powerful incentives in settings characterized by strong job security and lack of other incentives to perform, they are unlikely to play a major role in this context as most INPS managers in my sample happen to be toward the end of their career (Bertrand et al., 2019).

Systematic preferences of managers for a particular geographical region or more or less productive offices do not represent a threat to my empirical strategy, an argument I lay out more formally in IV.B. One may also worry that, as the bonus is a function of previous year achievements, this might generate cycles of high and low effort. However, this is not consistent with the evidence presented in Subsection V.B showing that the productivity gains are mostly driven by changes in the number of workers assigned to the office as opposed to changes in office output. Finally, one might be concerned that managers may sort to offices based on where they expect to receive the highest bonuses; this is akin to sorting based on comparative advantage. I test for this type of sorting in Subsection IV.D and I find no evidence of sorting on comparative advantage. Refer to Subsection IV.B, IV.C, and IV.D for a careful discussion of the threats to the identification assumption and corresponding tests of validity.

Appendix E : Autocorrelation

In this section, I use simulation analysis to show that a substantial degree of autocorrelation is needed to represent a serious threat to the leave-out strategy presented in Section V.A.

The leave-out strategy ensures that the outcomes of interest are not mechanically related to my measures of manager ability. However, if unobserved productivity shocks u_{it} are serially correlated, my leave-out measure of managerial effectiveness may still be spuriously, yet indirectly correlated with outcomes.

To evaluate what degree of autocorrelation would represent a serious threat to my empirical strategy, I hold constant the mobility structure from my sample and generate four simulated log productivity measures $y^1_{it}, y^2_{it}, ..., y^4_{it}$ suffering from different degrees of autocorrelation. I model log productivity as additive in an office permanent component (α_i), a manager component ($\theta_{m(i,t)}$), a time component (τ_t), and an idiosyncratic error term (u^d_{it}) as in equation (14).

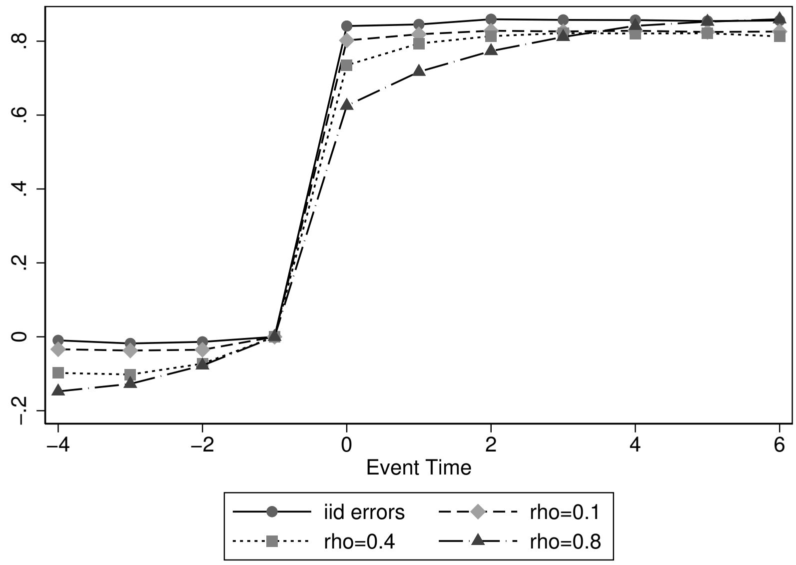
$$y^d_{it} = \alpha_i + \theta_{m(i,t)} + \tau_t + u^d_{it} \tag{14}$$

where $d = \{1,2,3,4\}$ indexes the DGP and all the other variables are defined as in Section IV. I draw the manager, office, and time effects from a normal distribution with mean zero and variance σ^2. The four DGP's only differ in the degree of autocorrelation present in the error term. The autocorrelation structure is specified as follows $u^d_{it} = \rho u^d_{i,t-1} + \xi^d_{it}$ where I choose $\sigma^2_{\xi^d}$ so that the variance of the error term is constant across DGPs $\xi^d_{it} \sim N(0, \sigma^2_{\xi^d})$ and the autocorrelation coefficient takes the following values $\rho = \{0, 0.1, 0.4, 0.8\}$. Therefore, u^1_{it} represents an i.i.d. error, while $u^2_{it}, ..., \varepsilon^4_{it}$ have an AR(1) structure.

When the errors are i.i.d., the leave-out procedure fully purges $\hat{\pi}^k_1$ from idiosyncratic productivity shocks. Therefore, I compare the performance of the leave-out procedure in the presence of serially correlated productivity shocks against the benchmark case of i.i.d. errors. I run the leave-out estimation strategy outlined in Section V.A on each of the seven simulated dependent variables, repeat this procedure 1,000 times, and report the average the estimated coefficients over the 1,000 replications in Figure E.I. This

Figure shows that modest levels of autocorrelation (i.e., $\rho \approx 0.1$) have little impact on the leave-out procedure. As my estimated autocorrelation coefficient is 0.04, I conclude that autocorrelated errors do not represent a serious threat to my empirical strategy.

Figure E.I: Impact of Serial Correlation on the Leave-Out Procedure



Note: This figure reports the point estimates ($\hat{\pi}_1^k$) from (11). The dependent variables are the simulated outcomes $y_{it}^1, y_{it}^2, ..., y_{it}^7$ constructed as in (14). The x-axis indexes event time.

Appendix F : Reduced Form Impacts of Managerial Talent on Time Allocation

In this section, I investigate whether the time allocation changes with the takeover of a more productive manager. Table F.I presents the reduced form impacts of managerial talent on absenteeism rates, training, overtime work, total hours, and the wage bill.

Better managers decrease the absenteeism rate of the office they oversee (column 1). Interestingly, this effect seems to be short-lived, as it peaks four quarters after the change in leadership and then it appears to fade. This is consistent with managers directly affecting the incentives for employees to show up to work, most likely by requesting audits for those workers whom they believe take sick days without a medical reason (Online Supplement J). Managers can enrol their employees in courses and workshops organized by INPS (i.e., formal training) and can also engage in on-site tutoring (i.e., informal training). I find no overall impact of managerial talent on the former (column

2) while I can not evaluate the latter as INPS does not collect information on informal training. Remarkably, more productive managers keep up production without resorting to more overtime hours to compensate for the reduction in full-time equivalent employment (column 3). I summarize these results studying the evolution of the total number of hours devoted to production. As a more productive manager takes charge, the number of hours devoted to production shows a modest decrease (column 4) mirroring the effect on FTE.

Next, I study the impact of managerial talent on total costs, which I proxy with the wage bill. As I do not observe the wage bill directly, I construct an index for it. Specifically, since overtime hours are remunerated 30% more than regular hours, I construct the wage bill index by weighing every regular hour (h) by one and every overtime hour by 1.3:

$$\text{wage bill}_{it} = 1 \times h_{it} + 1.3 \times \text{overtime}_{it}.$$

Since the wage index is a function of the number of hours, not surprisingly its behavior closely mirrors it (column 5). This wage index abstracts from seniority benefits and social contributions, thus, these estimates are likely to be an upper bound on the true impact on the wage bill.

Appendix G : Heterogeneity

In this Section, I assess whether managers have heterogeneous treatment effects that differ by geographical location, office size, office type, social capital, and baseline productivity. I estimate the following specification which builds on (11) and allows for heterogeneity in H_i (a pre-determined office characteristic):

$$\Delta \ln P_i^k = \pi_0^k + \pi_1^k \widehat{\Delta M}_i^{Lk} + \pi_1^{kH} \widehat{\Delta M}_i^{Lk} \times H_i + \Gamma^k X_i + \Delta \varepsilon_{it}^k. \tag{15}$$

All the variables are defined as above. π_1^{kH} captures the heterogeneous treatment effects and it is the main coefficients of interest.

Northern Italy is known for being a rich and entrepreneurial region, while Southern Italy is often depicted as poor and unproductive. One might think that sites located in one of these two regions may be more responsive to changes in managerial talent. Contrary to these expectations, Figure G.Ia shows that increases in managerial ability have no differential impact on offices located in the North compared to those in the South boradly defined. Along similar lines, it might be easier to improve performance in

Table F.I: Estimated Effects of Changes in Managerial Talent on Time Allocation

k	(1) Abs. Rate	(2) A(Training)	(3) A(Overtime)	(4) Ln(Hours)	(5) Ln(Wage Bill 30%)
-4	-0.014	0.180	0.031	0.000	-0.001
	(0.022)	(0.125)	(0.081)	(0.067)	(0.068)
-3	-0.023	0.134	0.061	0.016	0.015
	(0.016)	(0.130)	(0.069)	(0.059)	(0.059)
-2	-0.033	0.077	0.152	0.064	0.063
	(0.016)	(0.091)	(0.082)	(0.046)	(0.046)
0	-0.016	0.004	-0.009	-0.198	-0.199
	(0.020)	(0.108)	(0.082)	(0.100)	(0.103)
1	-0.007	-0.069	-0.082	-0.195	-0.194
	(0.016)	(0.123)	(0.089)	(0.075)	(0.075)
2	-0.046	-0.308	0.003	-0.208	-0.207
	(0.019)	(0.142)	(0.087)	(0.082)	(0.082)
3	-0.057	-0.376	-0.024	-0.172	-0.173
	(0.020)	(0.140)	(0.070)	(0.092)	(0.091)
4	-0.074	-0.133	0.113	-0.161	-0.161
	(0.024)	(0.133)	(0.083)	(0.103)	(0.102)
5	-0.037	-0.030	-0.003	-0.330	-0.331
	(0.020)	(0.138)	(0.079)	(0.101)	(0.100)
6	-0.045	-0.093	-0.017	-0.421	-0.446
	(0.026)	(0.135)	(0.091)	(0.115)	(0.126)
N	318	318	318	318	318
Time FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the reduced-form impacts of managerial talent on the office time allocation. More specifically, it reports the coefficients π_1^k obtained estimating (11) on the balanced-analysis sample. N represents the number of office-quarter observations. The dependent variable, y_{it}, is reported at the top of each column and $A(.)$ is short for asinh. All models include time fixed effects, main effects and two-way interactions between a dummy for Center-North, a dummy for main offices, a set of dummies for quartiles of baseline productivity, as well as time effects interacted with the dummy for Center-North. k indexes event time. Each coefficient is obtained from a separate regression. Bootstrapped standard errors are reported in parentheses.

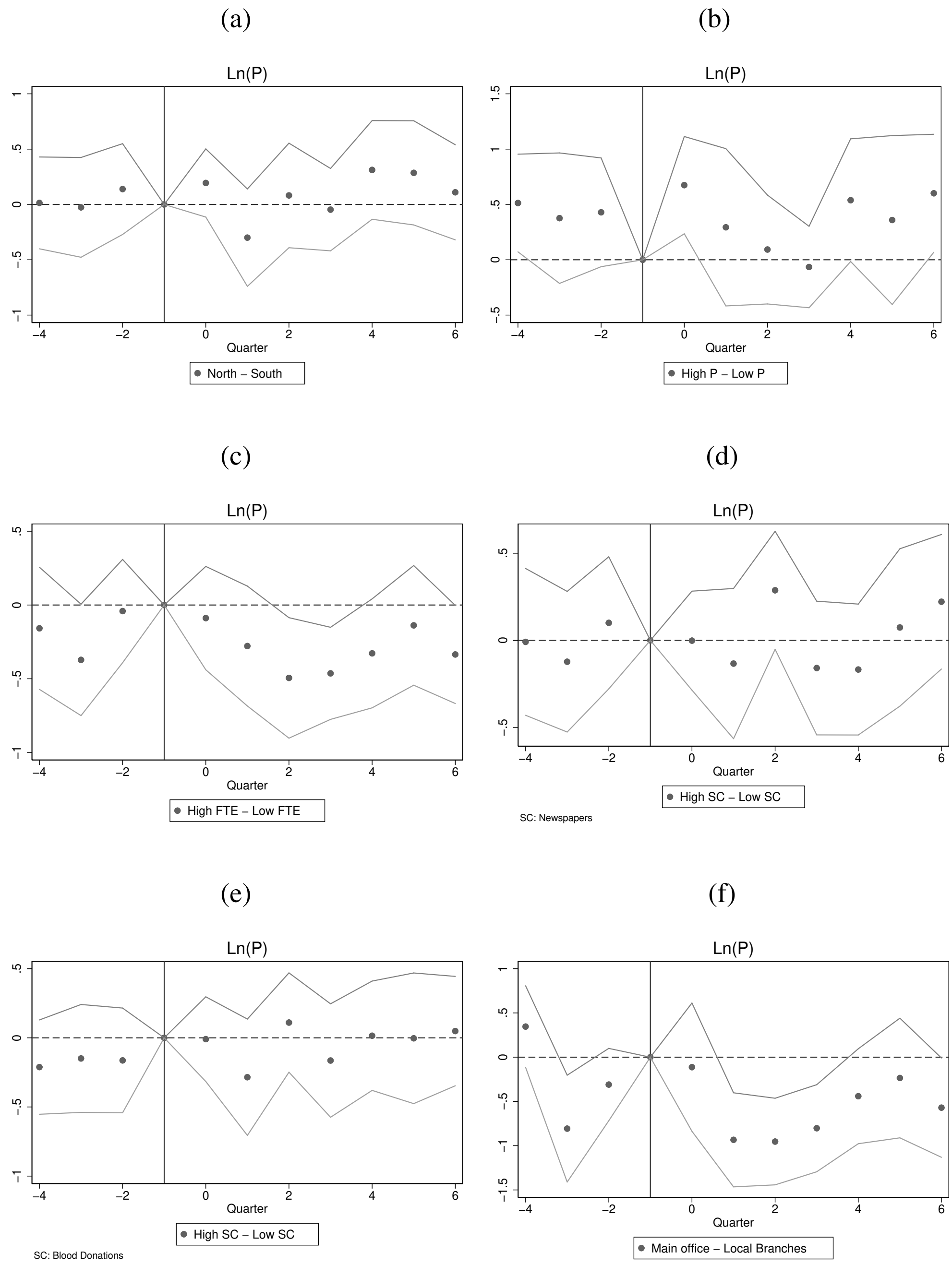
an unproductive, small office rather than in a large, productive site. I measure baseline productivity and office size as the average of these variables in 2011. I define "low" productivity offices as those below the median baseline productivity; likewise, small (i.e., "low" FTE) sites are those below median baseline size. Interestingly, there are no heterogeneous effects along these dimensions (Figures G.Ib and G.Ic).

As culture and social norms shape work ethics and everyday interactions, I test whether sites located in areas with higher social capital are more responsive to changes in manager ability. Social capital is a complex and multifaceted concept. I construct two province-level proxies for it using the daily number of non-sport newspapers sold per 1,000 inhabitants and the number of donated blood bags per million inhabitants in 1995 (Guiso et al., 2004; Cartocci, 2007). Figure G.Id and G.Ie display no evidence of heterogeneity in these two dimensions of social capital.

I also test whether productivity gains differ between main offices and local branches. Figure G.If shows some suggestive evidence that productivity gains may be lower in main offices than in local branches, although this effect is only temporary and imprecisely estimated. This pattern is consistent with managers directly overseeing employees in local branches, whereas the hierarchy is more complex in main offices (Online Appendix C).

Albeit I have admittedly limited power to detect heterogeneous treatment effects, productivity gains do not appear to differ by geographical location, office size, office type, baseline productivity, and social capital. These findings line up with the structure imposed by the two-way fixed effects models that do not allow for heterogeneous treatment effects.

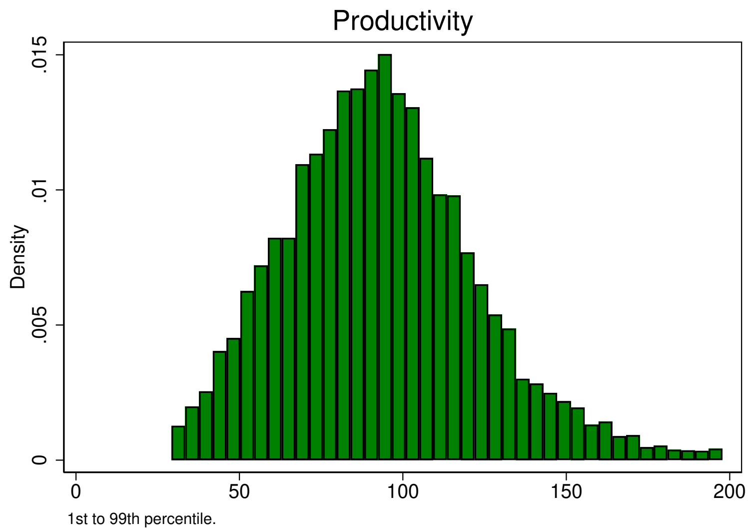
Figure G.I: Estimated Heterogeneous Effects of Leadership Changes



Note: Panels (a)-(f) report the regression coefficients and the associated 95% bootstrapped confidence intervals that identify the heterogeneity in the reduced-form impacts of managerial talent on office productivity, i.e., $\hat{\pi}_1^{kH}$ from (15). The coefficients at $k = -1$ are normalized to zero. The dependent variable is log productivity in all panels. H_i represent the pre-determined office characteristic that might be associated with heterogeneous impacts of managerial talent. The set of H_i includes office geographical location (panel a), baseline office productivity (panel b), baseline office size (panel c), social capital proxied by daily sales of non-sport-newspapers (panel d), social capital proxied by blood donations (panel e), and main office vs local branch (panel f). "High H_i" is defined as being above the median of baseline H_i. The x-axis indexes event time.

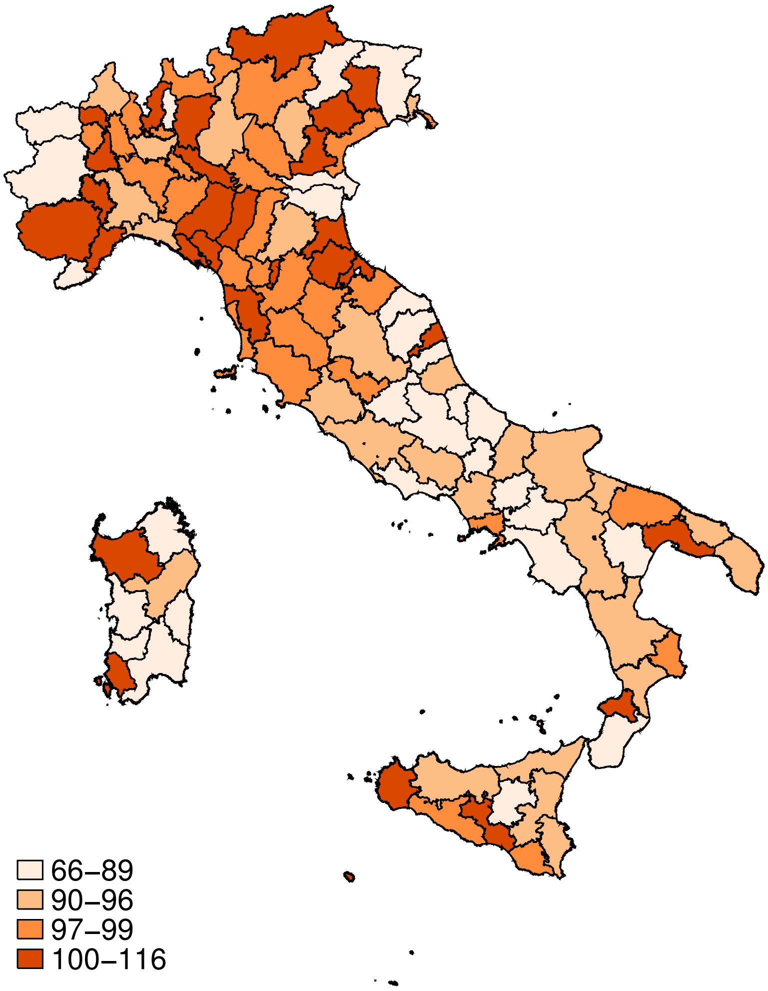
Appendix H Additional Figures and Tables

Figure H.I: Distribution of Productivity



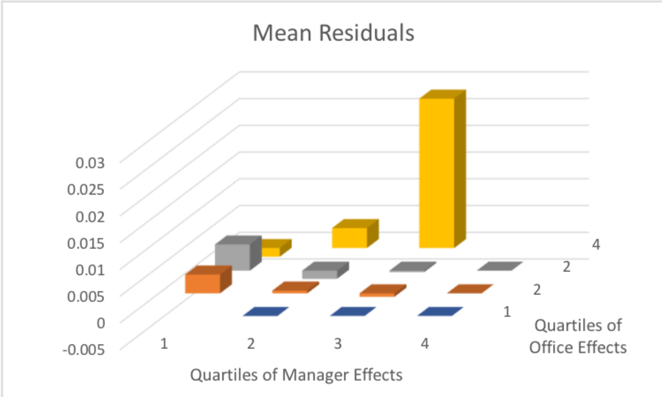
Note: This figure shows the distribution of quarterly productivity for the full sample of social security offices. Observations below the 1st percentile and above the 99th percentile have been excluded.

Figure H.II: Heat Map of Province Average Productivity



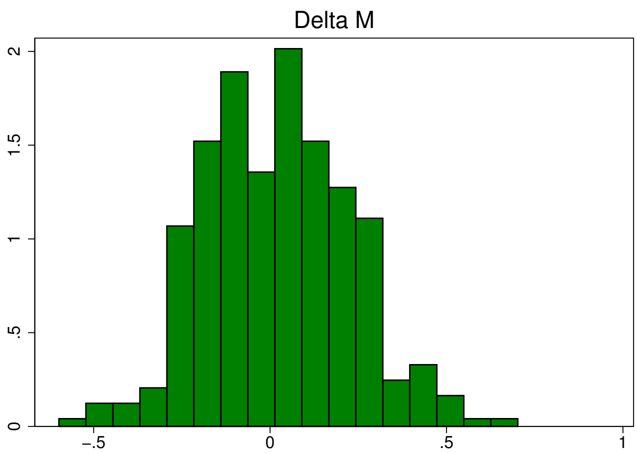
Note: This figure shows the average office productivity in each of the 100 Italian provinces. Darker shaded areas represent more productive provinces.

Figure H.III: Mean Residual by Manager/Office Quartiles (Largest Connected Set), 2011q1-2017q2



Note: This figure shows mean residuals from model (3) on the largest connected set with cells defined by quartiles of estimated manager effect, interacted with quartiles of estimated office effect.

Figure H.IV: Treatment Intensity



Note: This figure reports the distribution of the change in managerial talent ($\widehat{\Delta M_i}$) associated with the change in leadership for the events in the balanced-analysis sample.

Table H.I: Dispersion in Productivity

Productivity Measure	Within-Industry Productivity Moment	
Panel A: My Measure		
Labor productivity:	Median	4.524
log(weighted claims/employee)	IQ range	0.426
	90-10 percentile range	0.860
	95-5 percentile range	1.161
	St. deviation	0.366
	N	13,212
Panel B: Syverson (2004)		
Labor productivity:	Median	3.174
log(value added/employee)	IQ range	0.662
	90-10 percentile range	1.417
	95-5 percentile range	2.014

Note: Panel A reports the statistics of interest for my productivity measure calculated over the full sample. N represents office-quarter observations. Panel B is taken from Table 1 of Syverson (2004) and reports plant-level productivity distribution moments across 433 (four-digit SIC) manufacturing industries.

Table H.II: Analysis of Variance of Quarterly Measures of Productivity per Worker at INPS Offices

	(1) Ln(P)	(2) Ln(P)	(3) Ln(P)	(4) Ln(P)	(5) Ln(P)
N	11643	11643	11643	11643	11643
R sq.	0.352	0.579	0.640	0.615	0.643
Adj. R sq.	0.350	0.560	0.603	0.584	0.604
Time FE	Yes	Yes	Yes	Yes	Yes
Office FE	No	Yes	Yes	No	No
Manager FE	No	No	Yes	Yes	No
Manag-by-Office FE	No	No	No	No	Yes
Pvalue			0.000	0.000	

Note: This table investigates how much of the variance in log productivity is explained by the office, manager, and time components in the full sample. N represents the number of office-quarter observations. The p-value at the bottom of the table tests the null hypothesis that manager effects are jointly zero.

Table H.III: Estimated Effects of Changes in Managerial Talent

k	(1) Ln(Productivity)	(2) Ln(Output)	(3) Ln(FTE)
-4	-0.119	-0.117	0.002
	(0.118)	(0.124)	(0.056)
-3	0.045	0.032	-0.013
	(0.128)	(0.128)	(0.055)
-2	-0.111	-0.087	0.024
	(0.110)	(0.111)	(0.042)
0	0.390	0.178	-0.212
	(0.103)	(0.086)	(0.092)
1	0.484	0.282	-0.202
	(0.126)	(0.115)	(0.070)
2	0.399	0.110	-0.290
	(0.124)	(0.112)	(0.078)
3	0.516	0.249	-0.266
	(0.088)	(0.104)	(0.077)
4	0.447	0.179	-0.268
	(0.136)	(0.107)	(0.090)
5	0.417	0.038	-0.379
	(0.143)	(0.148)	(0.083)
6	0.661	0.168	-0.493
	(0.113)	(0.124)	(0.109)
N	318	318	318
Time FE	Yes	Yes	Yes

Note: This table reports the reduced-form impacts of managerial talent on productivity, output, and full-time equivalent employment. More specifically, it reports the coefficients π_1^k obtained estimating (11) on the balanced-analysis sample. N represents the number of office-quarter observations. The dependent variable, y_{it}, is reported at the top of each column. All models include time fixed effects, main effects and two-way interactions between a dummy for Center-North, a dummy for main offices, a set of dummies for quartiles of baseline productivity, as well as time effects interacted with the dummy for Center-North. k indexes event time. Each coefficient is obtained from a separate regression. Bootstrapped standard errors are reported in parentheses. Results in graph format are reported in Figures IIIa, IIIb, and IIIc.

Table H.IV: Counterfactual Exercises (Largest CS)

		ΔY
Policy 1:	Reassign	7.7%
Policy 2:	Replace bottom 20%	2.6%
Policy 3:	Replace bottom 20% + Reassign	8.1 %
Policy 4:	Random allocation	2%

Note: This table reports the counterfactual exercises that illustrate the efficiency gains from alternative managerial allocation schemes in the largest connected set (CS). I consider four counterfactual policies that the social planner can implement. Policy 1: she can reallocate existing managers according to the optimal rule. Policy 2: she can fire the bottom 20% of top-level bureaucrats and substitute them with the median manager (but allocate them as in the current environment). Policy 3: she can implement both Policy and 2. Policy 4: she can randomly assign existing managers to offices (1,000 iterations). The sample includes only the largest connected set. Figure L.IIIa, L.IIIb, and L.IIIc in the Online Supplement illustrate these concepts graphically.