Managers and Offices: Location-Based Determinants of Bureaucratic Effectiveness in India*

Prabhat Barnwal† Aaditya Dar ‡ Aprajit Mahajan§ Maitri Punjabi¶

Parminder Pal Singh Sandhu||

March 2026

Abstract

Variation in government performance is well documented across countries but much less is known about variation at finer geographies or its proximate causes. Using both observational and experimental data from a single Indian state, we document substantial variation in the delivery of key citizen services across offices and decompose it into contributions from managers and locations. First, bias-corrected AKM (Abowd et al., 1999) variance decompositions show that location effects are substantially larger than manager effects: location variance is nearly five times manager variance and accounts for over half of variation in service delivery time. Second, a movers design reveals that destination office productivity remains unchanged when managers transfer in. Consistent with this, support-staff productivity is invariant to the identity of the supervising manager, suggesting that relatively immobile support staff are a key location-specific factor. Finally, a state-wide RCT that permanently tightened delivery deadlines shows that these location-based constraints are not immutable: low-performing offices—where pre-existing deadlines were binding—substantially improved productivity, regardless of manager quality. Taken together, our results suggest an important role for location-based factors in service delivery.

*We thank Gulshan Banas, Torsha Chakravorty, Rishita Sankrit, Shivangi Chaturvedi and Kanchana Chiring for excellent research assistance. We are grateful to the Department of Governance Reform, Government of Punjab, India for sharing data. The views expressed in this paper are those of the authors alone and do not necessarily reflect those of the Government of Punjab. The use of Government of Punjab data does not imply government endorsement of the findings or conclusions. Financial assistance from J-PAL Governance Initiative, International Growth Center and the Digital Identity Research Initiative at the Indian School of Business is gratefully acknowledged. The RCT for this project was registered in the AEA RCT Registry (AEARCTR-0009621) and received IRB approval from Michigan State University (STUDY00006452).

†Michigan State University, prabhat@msu.edu.

‡Independent researcher, aadityadar@gmail.com.

§Department of Agricultural & Resource Economics, UC Berkeley, aprajit@gmail.com.

¶Michigan State University, mpunjabi@msu.edu.

||Indian Administrative Service, Government of Punjab, and Indian School of Business, parminder.sandhu@ias.gov.in.

1 Introduction

Government performance varies widely across countries, and there is growing evidence of substantial variation even across large sub-national units such as states (e.g. infant mortality rates vary by a factor of 12 across the best and worst performing Indian states while the analogous gap for functional literacy rates is 1.5).[1,2] However, much less is known about the extent of variation across even finer geographies—across districts, offices, and individual service providers within a single state—or the proximate sources of such variation.

Understanding the sources of this fine-grained variation is also particularly policy-relevant because it is at this scale that governments can most plausibly and regularly intervene—by reassigning managers, changing office-level resources, or reforming administrative procedures. If variation across jurisdictions is driven largely by managers, then policies that directly affect manager incentives and performance could yield large benefits. Conversely, if variation primarily reflects location-based differences—like institutional capacity, political economy considerations or resource constraints—such policies by themselves may have limited effects, and more innovative approaches may be required. At least two fundamental challenges complicate the study of this question. First, granular data on individual managers' and local staff performance is typically unavailable. Second, the experimental ideal of randomized manager assignment across locations is practically infeasible.

We address these challenges by combining detailed administrative performance data, manager transfer data, and a state-wide RCT to assess the relative importance of manager- and location-specific components of government performance within a single Indian state (Punjab).[3] Punjab's e-governance platform provides uniquely granular data on bureaucratic processes, digitally tracking citizen service requests through the system. This allows us to observe–for over 10 million time-stamped citizen applications– which managers and clerical staff handle citizen service requests and how long these individual actors take to provide these services. We combine this data with manager transfer records over a five-year period (2021–2025). These services are also highly salient

[1]Infant mortality (per 1000 live births) in Uttar Pradesh is 50 and 4 in Kerala. Under-five mortality rate shows a comparable gap (Source: National Family Health Survey, 2019–21). Similarly, literacy rates range from 94 percent in Kerala to 62 percent in Bihar (Source: Census of India, 2011).

[2]The underlying causes of such sweeping differences are actively debated. Scholars have pointed to variations in political institutions such as state developmental orientation (Kohli, 2004), party system fragmentation (Chhibber and Nooruddin, 2004), bureaucratic reshuffles (Iyer and Mani, 2012), electoral incentives (Ferraz and Finan, 2011; T. Ryan Rogers and Andrew Hall, 2020) and ethnic favoritism (Burgess et al., 2015) as potential contributors. Others have emphasized variation in implementation capacity, documenting that under-resourced and overloaded local bureaucracies undermine effective service delivery Dasgupta and Kapur (2020). See also Ahuja (2019); Vivek (2014) for state-level case studies in India. In part this reflects the difficulty of measuring government functioning cleanly and drawing credible causal links.

[3] In the Indian context, a *transfer* refers to the re-assignment of an officer (hereafter, manager) from one office location to another. The potential reasons for a transfer include periodic rotation norms, administrative and public interest grounds (e.g., filling vacancies or project-specific needs), personnel management (e.g., promotions, training), welfare or hardship grounds (e.g., family or health reasons), and when ordered by legal or regulatory institutions (e.g., higher courts, the Election Commission). See Section 2.1 for more details on transfers.

to citizens: an analysis of a representative sample of 10,000 citizen grievances state-wide reveals that almost half were related to the services we study, with non-delivery by mandated deadlines being the most common complaint.[4]

We have three main findings. First, using bias-corrected AKM-style (Abowd et al., 1999; Kline, 2024) variance decompositions of service processing time, we show that location effects are substantially larger than manager effects: location variance is nearly five times manager variance and accounts for over half of total outcome variation. Second, a movers design reveals that destination office productivity remains unchanged when managers transfer in. Consistent with this, support-staff productivity is invariant to the identity of the supervising manager, suggesting that relatively immobile support staff are a key location-specific factor. Finally, we complement these observational findings with evidence from a state-wide randomized controlled trial (RCT) that permanently tightened service delivery deadlines. Using the estimated AKM fixed effects from the pre-experimental period—separately for manager processing time and support-staff processing time—to classify offices and managers by productivity, we show that these location-based constraints are not immutable: shorter deadlines substantially reduced processing times in low-performing offices regardless of manager productivity.

Taken together, our results are consistent with models in which frontline workers exercise real discretion in service delivery (Lipsky, 1980), transient supervisors cannot credibly monitor permanent teams (Holmström, 1982; Tirole, 1986), and persistent organizational norms resist change by rotating managers (Akerlof and Kranton, 2000; Kreps, 1990). In addition to providing credible evidence on the empirical relevance of these proposed mechanisms we contribute to the growing literature on government performance. First, we add to the body of work using mover designs to decompose performance variation.[5] Unlike prior work, which typically observes only aggregate office output, our e-governance data provide separate time-stamped measures of manager and support-staff processing time. This allows us to estimate AKM decompositions separately for each tier and to identify support-staff inertia as a proximate source of the dominant location effect.[6] Second, we provide novel experiment-based evidence on the explanatory power of office- vs manager-level factors in the effectiveness of policy interventions.[7]

[4]Based on authors' analysis of data from the Department of Grievance Redressal for 2021.

[5]This approach has been applied, most prominently, to understand variation in wages (see Abowd et al., 1999; Card et al., 2013), healthcare utilization (Finkelstein et al., 2016) and teacher effectiveness (Chetty and Hendren, 2018). In the public sector context, Fenizia (2022) studies the effect of managers on overall office performance in Italy, Barteska and Lee (2025) use the rotation of Korean export promotion officers across countries, and Best et al. (2023b) examine bureaucratic ability in Russian public procurement. See also Banerjee et al. (2020); Dodge et al. (2025) and Rogger and Schuster (2023) for a review of the use of e-governance data to track government effectiveness.

[6]This complements the literature on the role of selection, incentives and monitoring in bureaucracies (see e.g. the review in Finan et al., 2015) and the work on public sector management and productivity (see e.g. Fenske et al., 2023; Genicot et al., 2026; Rasul and Rogger, 2018).

[7]A distinct literature examines government performance in episodic "mission mode" activities—time-bound tasks with automatic exit once complete, such as elections, vaccination campaigns, limiting seasonal crop residue burning or census operations (see e.g. Dipoppa and Gulzar, 2024; Kapur, 2020; Khan, 2025)—arguing that the state may perform better on such episodic activities than on routine service delivery. We complement this literature by examining bureaucratic performance in non-mission mode tasks where organizational effectiveness may operate through different

Our findings have at least two important implications. First, the dominance of location effects suggests increasing the research and policy focus on the institutional environment and constraints in which bureaucrats operate. Second, reforms targeting office-level factors—such as technology, management practices, local political economy or resource allocation—may be important complements to manager-level policies in improving service delivery. More generally, further unpacking the sources of office-level heterogeneity is an important topic for future research.

The paper proceeds as follows. Section 2 describes the institutional context and data. Section 3 outlines our empirical strategy, Section 4 discusses identification, Section 5 presents the results while Section 6 discusses implications and concludes.

2 Context and Data

2.1 Context

We focus on mid-level bureaucrats in Punjab, a state in Northern India with a population of approximately 28 million (2011 Census). The state has implemented a series of e-governance reforms that digitized a broad set of citizen services and created detailed administrative records enabling precise measurement of service delivery times and responsible managers (Kumar and Singh, 2024). In 2018, the state enacted the Punjab Transparency and Accountability in Delivery of Public Service Act which mandates time-bound delivery of government services with specific stipulated time limits (Government of Punjab, 2018). The Act empowers the government to specify the designated managers responsible and time limits for the delivery of specified public services. The Act also mandated electronic service delivery when technically feasible. This legislation creates the legal framework that generates the time-stamped administrative data for our analysis.

Our data covers approximately 1,000 offices across five major government departments over a five-year period (January 2021–December 2025). These departments—Revenue (e.g. caste and residence certificates), Health (e.g. birth and death registrations), Social Security (e.g. pensions and welfare), Rural Development (e.g. rural area certificates), and Local Government (e.g. special cases of birth and death certificates)—collectively handle a large volume of citizen-government interactions processed via the state's e-governance platform.[8] Table 1a summarizes the number of services and applications processed by officials in each department during our sample period.

The managers in our study occupy mid-level positions in the state bureaucracy (senior within their district or sub-district jurisdictions, but mid-career relative to the statewide civil-service hierarchy) with substantial discretion over service delivery. They include Sub-Divisional Magistrates (SDMs), Tehsildars, Medical Officers, Block Development Officers, and Executive Officers of Mu-

mechanisms.

[8] Best et al. (2023a) provide examples of three specific government functions (social security, public procurement, and tax collection) from where existing administrative case data can be repurposed to construct objective measures of bureaucratic performance. In our context, a common e-Governance platform called eSewa has standardized the processing of citizen applications on a large number of services across multiple state government departments.

nicipal Committees. These positions typically require specialized training and are filled through competitive examinations and/or internal promotions.

Transfers are common (see Section 2.2.2 for more details).[9] Iyer and Mani (2012) find that elite Indian civil servants average about 16 months in a posting (well below the recommended tenure guidelines of 2-3 years) with extensive transfers occurring when new political leadership assumes office.[10] In our setting, we focus primarily on lateral transfers for these mid-level managers across equivalent positions.

2.2 Data

2.2.1 Service (Application) Data

The state government has been rapidly moving public service delivery online.[11] We focus on these online services as they form an increasingly important medium of service delivery. Our primary dataset contains the universe of over 10 million citizen applications filed between January 2021 and December 2025 across 111 different services in the five departments.[12] These applications encompass a wide range of citizen services—including residence and health certificates and social security benefit verifications (including caste certificates required for affirmative action purposes) —representing core administrative processes that form a significant portion of routine citizen-government interactions.

We use the number of days to process an application (from filing to final approval) as our primary measure of productivity for at least four reasons. First, it is objectively defined and every handoff is electronically time-stamped, allowing consistent aggregation at the manager and office level. Second, it is relatively high frequency (we assess outcomes at the week level) allowing for finer examination of productivity changes. Third, it is legally meaningful because time-bound service laws set explicit delivery deadlines. Finally, as noted above, it is highly salient to citizens—non-delivery by mandated deadlines is the most common citizen grievance.

For each application we observe the filing date, approval date, the approving authority (the officer in charge of processing the application, hereafter *manager*), and processing office. The data tracks processing time for each actor along the chain, allowing us to identify separately the application's time with the office staff and time with the manager. While processing time is our primary measure, we also construct an alternative measure, the delay rate, using information on

[9] Recognizing the administrative disruption caused by frequent transfers, the Indian Supreme Court ordered tenure protections for civil servants Supreme Court of India (2013). However, implementation has been limited, with state governments continuing to exercise discretionary transfer powers.

[10] Using data on 2,800 Indian Administrative Service officers over 1980-2004, Iyer and Mani argue that politicians use transfers as a tool to control officers and find that officers who develop expertise-based reputations experience fewer transfers and lower variability in post importance (relative to those who develop loyalty-based reputations).

[11] Despite this push, some citizen-government interactions require physical presence or complex multi-party coordination and continue to be handled offline.

[12] Our administrative data covers the same five departments throughout this period. The number of digitized services has grown over time as the state expanded its e-governance platform.

whether the service was delivered within the mandated deadline (see Appendix C for details on the construction of our key outcomes).

Service complexity varies significantly both within and across departments. Simple services like income certificate issuance require minimal documentation and single-manager approval, while complex services such as a delayed registration of death involve multiple verification stages and coordination between officials in different departments. Processing deadlines reflect this complexity, ranging from 7 days for straightforward certificate services to 45 days for multi-stage services. The average application in our sample takes 1.93 log days to process at the manager-week level, corresponding to a geometric mean of approximately 7 days.[13] Substantial variation exists across offices: the 90th percentile office processes applications in half the time of the 10th percentile office, highlighting the productivity differences our analysis aims to explain.

To illustrate the service workflow, consider a citizen applying for a Caste Certificate, a document verifying membership in a caste category (Scheduled Caste, Scheduled Tribe or Other Backward Class) required for accessing quotas (affirmative action) in education or employment.[14] The citizen submits an application with supporting documentation by either visiting a service center (called a Sewa Kendra) or online—in both cases the application is digitized and assigned to the relevant Tehsildar office based on the applicant's residence. The office then completes the verification process which may require field investigations by junior officials (patwaris), cross-referencing with government documents or community interviews. Finally, the manager—the Tehsildar in this case— approves or rejects the application or, in rare cases, returns it to the applicant with a request for additional documentation.

While the workflow is standardized and uniform across the state (and largely so across the country), there is a large variation in processing time across tehsil offices. Figure 1 shows tehsil-level variation in days taken to process applications for a Scheduled Caste Certificate after controlling for district-level observables (district population, average income, and fraction Scheduled Caste). We find substantial residual variation, motivating our investigation.

Outcomes: The richness of our data allows us to construct three productivity measures. Our primary outcome, y_{it}, is a manager-week measure of productivity for manager i in week t. For each manager-week we take all applications approved by manager i in week t, compute the average number of calendar days taken to process each application and then take the natural logarithm. Second, $\log(y_{it}) = \log(y^{o}_{j(i,t)})$, the log of the mean days taken *by manager* i per application in office $j(i,t)$. Finally, $\log(y_{it}) = \log(y^{s}_{j(i,t)})$ is the log of the mean days per application taken by *subordinate staff* in $j(i,t)$. In all cases y_{it} is an *average* across applications within a manager-week—an intensive-margin efficiency measure, not a workload total.[15] Since each manager handles multiple service

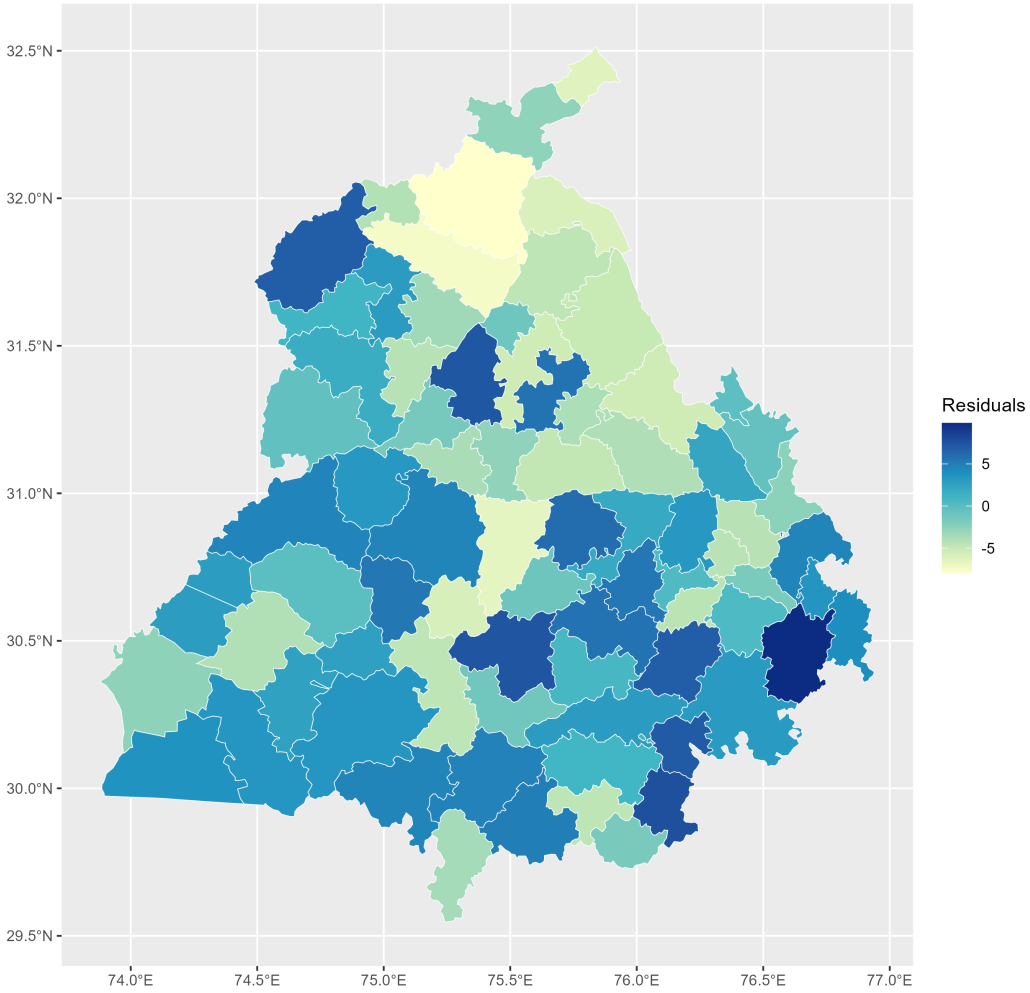
[13] The geometric mean ($\exp(\overline{\log y}) \approx 6.9$ days) is substantially lower than the arithmetic mean of approximately 12 days (Table 1b) because the processing time distribution is right-skewed. We work throughout with log processing days, so the geometric mean is the relevant summary statistic. Approximately 18% of applications exceed their mandated processing deadline.

[14] Punjab's population has the highest fraction of Scheduled Castes among all Indian states (Census 2011).

[15] At the application level, total processing time equals manager time plus staff time by construction. However, the

types, y_{it} pools across all services processed by manager i in week t.[16]

Figure 1: Geographic variation in public service delivery time



Note: Figure shows a tehsil-level (i.e. sub-district) map of Punjab plotting the residuals from a regression of application processing days for a Caste Certificate (Schedule Caste group) on district-level observables, including average income, population, and fraction of Scheduled Caste population. The deadline for this service is 17 days. The service is provided by the Department of Personnel and the approving authority is a Tehsildar. Map is generated using application-level data for the 2021–2025 period (n=1,939,589 observations).

component-level variables (staff time, manager time) are not observed for all application types: staff processing time is missing for approximately 9% of manager-week observations whose service types do not track separate staff time. Because the summary statistics in Table 1b restrict to the common sample where all three measures are observed, the composition of services in that subsample differs from the full sample, and the component means need not sum to the overall mean.

[16]This is the relevant unit of analysis for the AKM decomposition. The RCT analysis is at the service-week level (see below for additional discussion).

2.2.2 Manager and Transfer Data

We merge the services data with comprehensive manager records identifying the approximately 2,300 managers in charge of processing applications during this period (Table OA-2 provides a year-by-year breakdown).[17]

The manager transfer data is derived from eSewa administrative records: district technical coordinators enter transfer order information directly into the platform when managers are reassigned. This data includes each manager's start and end dates in each posting. We focus on managers that moved across offices during 2021–2025 subject to the following selection criteria: we require managers to serve at least two weeks in both origin and destination postings to allow sufficient time for productivity assessment. Second, we exclude assignments marked as "additional charge" where managers temporarily oversee multiple positions, to focus on substantive postings.

This leaves us with a sample of 994 managers who moved across 804 offices during our study period. Table 1b provides a summary of manager characteristics and productivity in our entire sample and the subsample of movers. In addition to dates, the transfer data also identifies origin and destination offices, and designation changes. The majority of transfers (876 out of 994) were within a designation (i.e. lateral moves) and we find similar results when excluding across-designation transfers (see Figure OA-13). In our sample, the average observed posting duration is approximately 47 weeks for the full sample and 35 weeks for movers (Table 1b); these are lower bounds on actual tenure since postings extending beyond our panel are right-censored. It is important to note that officers of the same designation are responsible for the same set of services, though service volume across offices may vary.

Transfers occur throughout the year but show some patterns, with larger reshuffles typically occurring mid-year or after elections (see Figure OA-2; eliminating post-election transfers does not change our results). In the analysis below, we do not find evidence that transfers are performance driven (see e.g. Figure 2(c)). Our relatively sparse mobility network reflects our institutional context where mid-level managers move much less frequently than private sector workers.

Table 1b compares movers with the full sample of managers (the table reports 959 of the 994 movers for whom non-missing productivity measures are available). While movers differ somewhat—they process slightly fewer applications and have somewhat different delay rates—these differences are relatively modest substantively. We also test whether transfers are correlated with destination productivity and find that manager productivity is weakly negatively correlated

[17] Online services are one component of a manager's duties. Multitasking models (e.g. Holmström and Milgrom, 1991) suggest that differential incentives (between online and offline services) across locations could create location effects through effort substitution between different tasks. Our response is two fold: first, our institutional setting mitigates this concern. Punjab's bureaucratic system features uniform wage schedules, standardized performance evaluation criteria, and career advancement processes that do not vary systematically across office locations, making effort substitution an unlikely driver of our observed location effects. Second, to the extent that such substitution exists across locations we interpret it as a *source* of the location-based effect rather than a confound, and a topic for further investigation.

with office productivity, providing limited evidence for negative assortative matching. In addition, we show below that there are no "pre-trends" in manager productivity in the weeks leading up to a transfer. The negative correlation between manager and office productivity is not inconsistent with the lack of anticipatory effects. It may reflect (moderate) systematic administrative assignment, while the absence of pre-trends suggests that transfer timing is unrelated to performance shocks. This is consistent with transfer timing reflecting tenure rather than performance.

Table 1: Summary Statistics

(a) Department Summary

			Full Sample			Largest Connected Component		
Department	Services	Applications	Managers	Offices	Obs*	Managers	Offices	Obs*
Rural Development	1	197,810	374	153	16,781	371	151	16,641
Revenue	45	4,213,178	810	336	39,732	783	321	39,116
Local Government	38	2,543,626	286	157	20,089	0	0	0
Health	59	2,938,584	781	324	47,645	265	87	15,879
Social Security	5	599,721	39	22	2,698	0	0	0
Total	111	10,492,919	2,290	992	126,945	1,419	559	71,636

(b) Manager Characteristics and Productivity

	Full Sample Mean (SD)	Movers Mean (SD)
Manager tenure		
Posting duration (weeks)	47.1	34.9
	(55.57)	(37.32)
Productivity measures		
Applications processed per week	82.66	80.23
	(147.33)	(128.86)
Processing days	11.90	13.17
	(21.81)	(23.53)
Processing days by office staff	11.88	11.35
	(21.08)	(21.30)
Log(Processing days)	1.93	2.03
	(0.96)	(0.97)
Number of Managers	2,290	959
Manager-Week Observations	126,945	73,808

Notes: Table 1a summarizes each department's applications (from application-level data) along with the number of managers and offices. The total number of unique services is 111 (37 services are handled by more than one department so the cross-department total is 148). *Obs = manager × week observations. A small number of managers serve in multiple departments; each such manager is assigned to the department in which they have the most observations. "Largest Connected Component" shows the subset of managers and offices in the largest connected component of the manager-office mobility network, used for AKM variance decomposition (Table 2). Local Government and Social Security lack sufficient cross-department mobility to connect to this component. Table 1b summarizes manager posting durations and productivity measures for the full sample and the subsample of movers used in the event study (Figure 2). Both panels restrict to manager-weeks with non-missing time-varying controls (staff counts, vacancy rates, workload measures), matching the estimation sample in Table 2. Productivity measures are restricted to manager-weeks where all three outcomes (total, manager, and staff time) are observed (corresponding to 959 of the 994 movers).

3 Empirical Specification

We use two approaches to examine the variation in office productivity across the state. We begin with the standard "AKM" model (see e.g. Kline, 2024, for a recent exposition) that decomposes productivity using a set of office and manager fixed effects. Next, we use an event-study (Finkelstein et al., 2016) of changes in productivity around manager transfers. Both sets of results complement each other and we outline each in turn.

3.1 AKM Model

Our first estimating equation follows the standard AKM decomposition:

$$\log(y_{it}) = \alpha_i + \psi_{j(i,t)} + x'_{it}\beta + \nu_{it} \tag{1}$$

where $\{\alpha_i\}_{i \in I}$ are manager fixed effects, $j(i,t) \in J$ is the office location of manager i in period t, $\psi_{\cdot}$ is a set of office fixed effects and x_{it} is a vector of controls.[18,19] The productivity measures y_{it} are those described above on p.5. Because each office has a single supervising manager at any given time, the mapping from (i,t) to j is one-to-one in our data; we retain the office subscript $j(i,t)$ to make the fixed-effect structure explicit.

3.2 Event Study

Our second estimating equation uses an event-study of changes in productivity around manager transfers. The key idea is that if manager characteristics drive performance, a manager's productivity should remain unchanged across transfers. Conversely, if office characteristics dominate, managers should adopt their destination office's productivity level upon transfer.

For each moving manager i, we define the "treatment intensity" as the difference in average productivity between destination and origin offices:

$$\delta_i = \log(\bar{y}_{d(i)}) - \log(\bar{y}_{o(i)})$$

where $\bar{y}_{d(i)}$ and $\bar{y}_{o(i)}$ are the mean productivity levels in manager i's destination and origin offices, respectively, calculated using application-level data excluding manager i's own observations. When appropriate, we compute mean productivity levels separately for the manager and the remaining staff.

[18] We will abuse notation and use (I, J, T) to refer both to the sets of managers, offices and time-periods respectively as well as their cardinality.

[19] The controls are week indicators, staffing and workload variables: total staff count, subordinate staff vacancy indicator, manager vacancy indicator, log number of applications received, and number of service types handled by the office.

Our estimating equation, estimated only on the sample of transferred managers, is:

$$\log(y_{it}) = \alpha_i + \sum_{r \neq -1} \theta_r \delta_i \mathbb{I}_{\{r = t - t_i\}} + \tau_t + \epsilon_{it} \tag{2}$$

where $\{\alpha_i\}$ are manager fixed effects, $\{\tau_t\}$ are time fixed effects, and θ_r are coefficients on the indicators for leads and lags relative to period of the transfer t_i and measures changes in productivity in weeks around the transfer relative to δ_i.

The primary coefficient of interest is θ_0, representing the immediate jump in productivity upon transfer. Under purely location-driven productivity variation, we expect $\theta_0 = 1$, indicating managers immediately adopt their new office's productivity level. Conversely, under purely manager-driven variation, we expect $\theta_0 = 0$. In Appendix A we show that Equation (2) can be derived from a stylized utility maximization framework where managers face both effort costs and conformity pressures.

4 Identification

4.1 Identification and Estimation in the AKM Model

The AKM framework requires a strict exogeneity assumption for consistent estimation via OLS: $\mathbb{E}\left(\nu_{it} | j(i,s) = k\right) = 0$ for all i, all $(s,t) \in T^2$, and all $k \in J$. These moment conditions imply that managers' entire transfer histories are orthogonal to productivity shocks.[20]

Connectedness. Manager and office effects in Equation (1) are jointly identified only within connected components of the mobility network—groups of offices linked through chains of manager transfers. Correspondingly, we restrict our main analysis to the largest connected component, which contains 1,419 managers across 559 offices. Appendix F.3.1 examines robustness to alternative sample choices. We also estimate Equation (1) separately by department (managers overwhelmingly transfer within rather than across departments), focusing on Rural Development and Revenue where offices are well-connected (see Table 1a).

As an alternative to restricting attention to the largest connected component, we also implement a multiple connected components (MCC) approach, following Best et al. (2023b). We retain all connected components meeting minimum size thresholds—at least 6 managers and 3 offices—rather than discarding smaller components. This yields substantially more managers and offices with productivity estimates, which proves useful for the RCT heterogeneity analysis (Section 5.3). We examine robustness to alternative threshold choices in Appendix I.5.

Normalization. Even within a connected component, manager and office effects are identified only up to a constant. For the largest connected component analysis, we adopt the conventional

[20] The AKM specification also imposes that manager and office effects combine additively (in logs). See Kline (2024) for discussion of this "transitivity" assumption.

normalization by setting one office effect to zero (the normalization does not affect variance estimates). For the MCC approach, we normalize within each component separately and include component-level intercepts.

Limited Mobility Bias. With infrequent manager moves, plug-in variance estimates from OLS fixed effects are upward biased (Andrews et al., 2008). We address this using the covariance shrinkage approach of Best et al. (2023b), which jointly shrinks estimated effects toward zero while accounting for the correlation structure of estimation errors. As robustness checks, we also implement the Andrews et al. (2008) homoskedasticity-based bias correction, the leave-one-out estimator of Kline et al. (2020), and the split-sample estimator from Best et al. (2023b) (all in Appendix F.3.2).

4.2 Event Study Identification

Identification for the event study relies on a strict exogeneity assumption—the same spirit as the AKM assumption above, but applied specifically to the sample of transferred managers. Here we emphasize its implications for transfer timing and the potential for testing pre-trends.

Transfer Timing Orthogonality: We require that transfer timing is unrelated to destination productivity shocks (recall that we only use the sample of transferred managers). This is plausible in our setting with transfers typically being based on tenure and routine (see also Appendix E) rather than performance. We can allow for some forms of endogenous transfers–for instance movers are allowed to differ from non-movers in their levels of productivity (through manager fixed effects) and we can allow for common shocks in productivity around the move (through relative week fixed effects).

Testing the Assumption: The event study framework allows us to test for systematic pre-trends that would indicate endogenous timing (see also Appendix G on matching). Figure 2 plots the estimates of $\{\theta_r\}$ from Equation (2). There is no systematic change in manager productivity in the weeks leading up to the transfer suggesting that transfers are not systematically timed to coincide with productivity shocks. The relatively flat pre-trend pattern is also inconsistent with strategic transfer timing based on performance considerations.

We formally test for pre-trends using a joint F-test of all pre-transfer coefficients (from -12 to -2 in relative weeks from transfer) following the suggestions in Roth (2022). The F-statistic is 1.57 (p=0.104), failing to reject the null of no pre-trends. The mean absolute pre-trend coefficient is small relative to the jump at transfer of 1.030.

5 Results

5.1 AKM Variance Decomposition

We begin with a bias-corrected (via shrinkage) AKM variance decomposition that quantifies the relative importance of manager versus office effects. A distinctive feature of our data is that we observe processing time separately for the manager and subordinate staff, allowing us to estimate the AKM decomposition for each component independently as well as for total processing time.Table 2 shows that office effects account for 83% of the combined manager-office variance in total processing time. This pattern persists for office staff time (79% office share) but is more balanced for manager time (66% office share).

Robustness. Our preferred estimates restrict the sample to the largest connected component (1,419 managers, 559 offices) and uses covariance shrinkage (Best et al., 2023b) to address limited mobility bias. We obtain similar office shares using alternative samples and bias-correction approaches. For instance, using multiple connected components (which retains more observations at the cost of cross-component comparability; Table OA-5) yields similar estimates. Similarly, alternative bias-correction methods including the correction of Andrews et al. (2008), split-sample estimation, and the leave-one-out estimator of Kline et al. (2020) (Tables OA-7 and OA-8) all yield very similar estimates and conclusions. Finally, department-specific decompositions also yield similar results (Table OA-4).

Manager and Office Effects Correlation. AKM decompositions with limited mobility typically estimate negative correlations between the two sets of fixed effects (e.g., $\rho = -0.32$ for Manager Time). This could reflect either genuine negative assortative matching or mechanical bias from estimating both effects in the same regression. We can use our manager- and staff-time specific AKM regressions to explore this issue. Specifically, we correlate manager fixed effects from the Manager Time regression (α^M) with office fixed effects from the Staff Time regression (ψ^S). Because these come from separate regressions, the correlations are not purely mechanical. We find $\widehat{\text{Corr}}(\alpha^M, \psi^S) = 0.17$ with a standard error of 0.07. This suggests that the negative within-column correlations are (at least in part) estimation artifacts rather than genuine negative sorting. The positive cross-stage correlation points to modest positive assortative matching, where faster managers tend to work in offices with more efficient staff.

Table 2: Variance Decomposition: Manager vs. Office Effects

	Total Time	Manager Time	Staff Time
Sample			
Observations	71,636	71,636	71,636
Managers	1,419	1,419	1,419
Offices	559	559	559
Variances			
$\text{Var}(y)$	1.03	0.52	1.41
$\widehat{\text{Var}}(\psi)$	0.60 (0.088)	0.28 (0.041)	0.77 (0.093)
$\widehat{\text{Var}}(\alpha)$	0.13 (0.083)	0.15 (0.038)	0.20 (0.075)
Ratios to Outcome Variance			
$\widehat{\text{Var}}(\psi)/\text{Var}(y)$	0.58 (0.085)	0.55 (0.079)	0.55 (0.066)
$\widehat{\text{Var}}(\alpha)/\text{Var}(y)$	0.12 (0.081)	0.29 (0.073)	0.14 (0.053)
Correlation and Shares			
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.07 (0.090)	-0.32 (0.078)	-0.13 (0.078)
$\widehat{\text{Corr}}(\alpha^{\text{M}}, \psi^{\text{S}})$	0.17 (0.069)		
Office Share	0.83 (0.051)	0.66 (0.052)	0.79 (0.048)
Manager Share	0.17 (0.051)	0.34 (0.052)	0.21 (0.048)

Notes: This table reports variance decomposition estimates from a two-way fixed effects (AKM) model: $\log(y_{it}) = \alpha_i + \psi_{j(i,t)} + \tau_t + x'_{it}\beta + \varepsilon_{it}$, where α_i are manager fixed effects, ψ_j are office fixed effects, τ_t are week fixed effects, and x_{it} are time-varying office-level controls. Each column estimates the model with a different outcome y_{it}, defined as the mean processing days per application for manager i in week t: **Total Time** uses overall processing days, **Manager Time** uses days attributable to the manager (approving authority), and **Staff Time** uses days attributable to subordinate office staff (see p. 5 for details).

Sample: Analysis is restricted to the largest connected component of the manager-office mobility network, which ensures joint identification of manager and office effects. Connected components are computed using the pooled sample across all departments. **Controls:** All regressions include week indicators, total staff count, subordinate staff vacancy indicator, manager vacancy indicator, log of the number of applications received, and the number of services handled by the office. **Estimation:** We report covariance shrinkage estimates to address limited mobility bias. Point estimates use 100 bootstrap iterations to estimate the error covariance matrix required for shrinkage. **Standard errors:** Standard errors (in parentheses) computed via double bootstrap: 200 outer bootstrap iterations $\times$ 100 inner iterations, using match-level block resampling. **Variance ratios:** $\widehat{\text{Var}}(\psi)/\text{Var}(y)$ and $\widehat{\text{Var}}(\alpha)/\text{Var}(y)$ report the estimated share of outcome variance attributable to office and manager effects respectively. **Correlation:** $\widehat{\text{Corr}}(\alpha, \psi)$ is the estimated correlation between manager and office fixed effects after covariance shrinkage. **Cross-stage correlation:** $\widehat{\text{Corr}}(\alpha^{\text{M}}, \psi^{\text{S}})$ correlates shrinkage-corrected manager fixed effects from the Manager Time regression with shrinkage-corrected office fixed effects from the Staff Time regression. The standard error is computed via bootstrap. **Variance shares:** Office Share $= \widehat{\text{Var}}(\psi)/[\widehat{\text{Var}}(\alpha) + \widehat{\text{Var}}(\psi)]$ measures the estimated fraction of combined manager-office variance attributable to offices. The standard errors for these two rows are identical since Manager Share $= 1-$ Office Share.

5.2 Event Study

We complement the variance decomposition with an event study that examines productivity changes around manager transfers. When a manager transfers from one office to another, total productivity at the destination office remains, to first order, completely unchanged (Figure 2). $\hat{\theta}_0$ is 1.030 for (log) processing days (CI: $[0.85, 1.21]$).[21] We extend the event study window to 16, 20, and 24 weeks and find similar estimates of θ_0 (Figure OA-9).

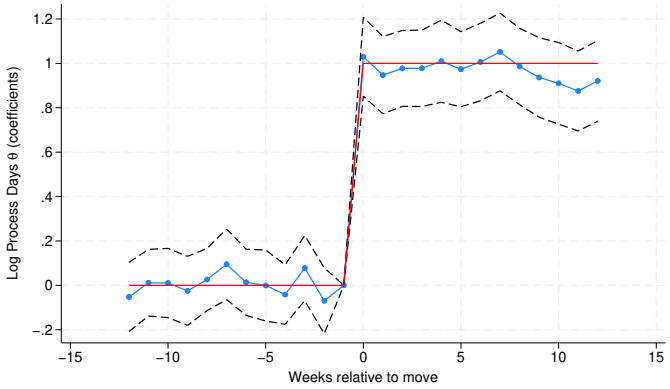
Dynamics: In panel (a), the pre-transfer period (weeks -12 to -1), the estimates hover around zero, consistent with stable office productivity prior to the transfer. In the post-transfer period (weeks 1 to 12), the coefficients likewise remain strikingly flat, showing no evidence of gradual adjustment in either direction—the new manager neither improves nor worsens the destination office's pre-existing productivity.[22]

[21] The corresponding figure is similar for delay rates; see Appendix F.1.

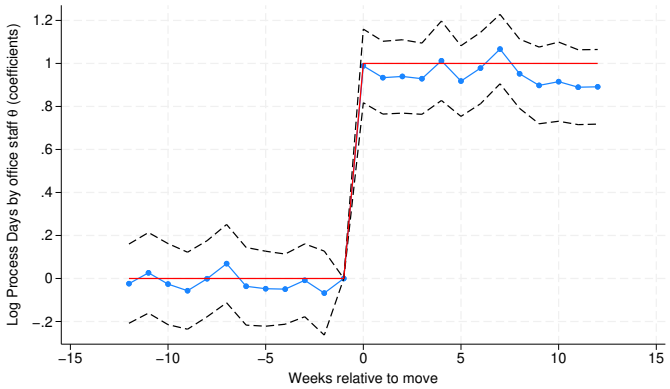
[22] A potential concern is that managers might inherit backlogs from their predecessors, confounding our interpretation. However, the convergence pattern with—$\theta_0 \approx 1$ and the flat transfer dynamics are inconsistent with a backlog story which would predict more gradual adjustment as inherited applications are processed.

Figure 2: Location-based share of application processing time

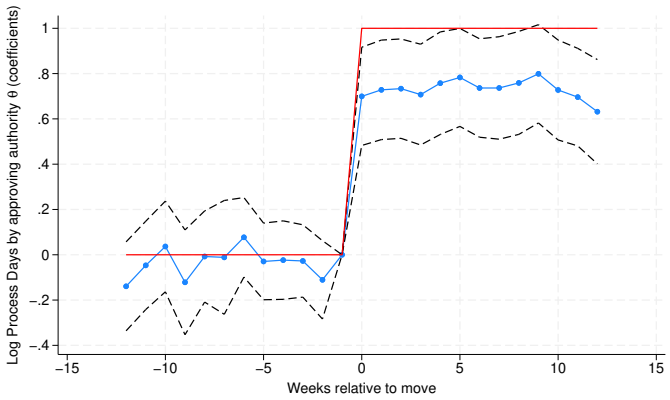
(a) Total Processing Time



(b) Office Staff Time



(c) Manager Time



Notes: Figure shows the $\theta_{r(i,t)}$ coefficients from our main estimating equation, normalized to zero at $r = -1$. Dashed lines are pointwise 95% CIs. **(a)** $\hat{\theta}_0 = 1.030$ with CI [0.85, 1.21] for total processing days. **(b)** $\hat{\theta}_0 = 0.988$ with CI [0.82, 1.16] for staff time. **(c)** $\hat{\theta}_0 = 0.699$ with CI [0.48, 0.92] for manager time. The solid red line shows the benchmark of immediate full convergence ($\theta_0 = 1$). Each panel uses the corresponding outcome to calculate delta (leave-one-out, Section 3). 994 movers, 13,041 observations.

Proximate Mechanisms: We attempt to identify the channels through which location effects operate by disaggregating application processing time into components attributable to different actors in the approval process—time taken by the manager and time taken by other office staff (the manager's subordinates in the destination office).

We re-estimate Equation (2) using subordinate staff processing time as the left-hand side variable and re-defining δ_i correspondingly using staff time. Figure 2(b) displays the estimates of $\{\theta_r\}$ from this regression. We estimate $\hat{\theta}_0 = 0.988$ (CI: $[0.82, 1.16]$), indicating that nearly all of the staff-time gap between origin and destination offices is location-based. The confidence interval includes 1, consistent with the hypothesis that staff productivity is entirely location-determined and invariant to the identity of the supervising manager. The remaining coefficients are very similar during the entire post-transfer period, with no evidence of gradual adjustment. Since staff time comprises the bulk of processing time (Table 1) it is perhaps unsurprising that our estimate of $\hat{\theta}_0$ is similar to that from our main specification. Perhaps most strikingly, when we carry out the same exercise using only the manager's own processing time (Figure 2(c)), $\hat{\theta}_0 = 0.699$ (CI: $[0.48, 0.92]$), demonstrating that location-based constraints bind even for activities over which managers have presumably the most control (their own time).

These results suggest that location effects operate in part through staff maintaining pre-existing productivity levels regardless of the supervising manager. This institutional "stickiness" may reflect several factors: staff may be more familiar with local procedures and systems, face constant resource constraints or local political pressures that persist regardless of the supervising manager. Subordinate staff are substantially less mobile than supervising managers and thus may be more sensitive to local institutional norms as well. Using eSewa administrative records on staff postings, we find that only 23% of subordinate staff (dealing clerks, patwaris, ANMs) transferred across offices during our sample period, compared to 69% of managers (Table OA-1). Subordinate staff who do transfer also move less frequently (1.3 postings on average vs. 3.3 for managers) and serve longer in each posting (2.4 years vs. 0.8 years). However, we lack data to test specific hypotheses about the underlying mechanisms through which staff immobility translates into location effects.

Office Level Heterogeneity: Finally, we examine whether our results differ according to the type of office by examining productivity dynamics for moves between offices at different points in the productivity distribution (Appendix F.2). We classify offices as high or low productivity using their average performance over the entire sample period. Managers initially in high-productivity offices who transfer to low-productivity offices immediately experience substantial declines in productivity, despite their strong prior performance. The symmetric pattern holds for managers originating in low-productivity offices: those transferred to high-productivity offices experience immediate performance improvements. These patterns are robust across different productivity thresholds for defining high and low performance offices (see Appendix G.2).

5.3 Experimental Evidence: Deadline Reductions

The evidence above suggests subordinate staff productivity plays an important role in determining overall office productivity and managerial turnover has limited effects in changing staff performance. We explore this in greater detail using experimental evidence from a randomized controlled trial that changed service delivery deadlines.

The experiment permanently reduced delivery deadlines for 15 randomized services (treated group) while maintaining the pre-existing original deadlines for another group of 15 services (control).[23] Treatment deadlines were set at the 60th-75th percentile of pre-treatment processing time per service, representing an average mandated reduction of 2.9 days (approximately 16%) across all 15 services. The intervention was implemented on December 15, 2022. The analysis sample comprises 394,626 manager-service-week observations across 30 services and 1,909 managers.[24] The post-treatment period contains 249,159 observations (134,410 treatment, 114,749 control).

We estimate the treatment effect using two approaches. First, a simple difference-in-differences specification:[25]

$$y_{isjt} = \psi_{sj} + \alpha_{si} + \tau_t + \beta \cdot (\text{Treat}_s \times \text{Post}_t) + \epsilon_{isjt} \tag{3}$$

where y_{isjt} is a measure of processing time in office j with manager i in charge of service s in week t, ψ_{sj} and α_{si} are service-by-office and service-by-manager fixed effects respectively, and τ_t are week fixed effects. β captures the average treatment effect on processing time. Standard errors are two-way clustered by service and week.

Panel A of Table 3 presents the basic results. The average treatment effect is small and statistically insignificant (-0.61 days, $p = 0.46$ for total time; -0.40 days for staff time). One possible explanation, consistent with our observational findings that office characteristics dominate processing time variation, is that low-productivity offices have room to improve while high-productivity offices do not, so that the average effect masks offsetting responses. We test this directly below. Figure OA-19 displays the corresponding event study, confirming parallel pre-trends and showing that treatment effects emerge shortly after implementation.

[23] The 30 study services were selected based on sufficient pre-treatment volume and stratified by department and online/offline status prior to randomization. An additional 76 services were excluded from the experiment; we examine these non-study services in Appendix I.1.

[24] The initial sample contains 414,598 observations (1,999 managers). Absorbing service×office and service×manager fixed effects iteratively drops 19,972 singleton observations (approximately 5%), yielding the regression sample of 394,626.

[25] Recent work has highlighted potential biases in two-way fixed effects estimators when treatment timing is staggered and effects are heterogeneous across cohorts (Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021). These concerns are less pressing to our setting since treatment is randomly assigned and all treated units receive treatment simultaneously (so e.g. there are no comparisons using already-treated units as controls).

Table 3: RCT Treatment Effects by Manager and Office Productivity

Panel A: Main Effects (Full Sample)

	Total Time	Manager Time	Staff Time
Treatment × Post	−0.613	−0.192	−0.399
	(0.829)	(0.126)	(0.732)
Observations	394,626	394,626	394,626
Dep. var. mean	10.19	1.89	7.73

Panel B: Main Effects (Heterogeneity Sample)

	Total Time	Manager Time	Staff Time
Treatment × Post	−0.891	−0.199	−0.673
	(0.727)	(0.129)	(0.634)
Observations	221,718	221,718	221,718
Dep. var. mean	9.74	1.81	7.34

Panel C: Implied Treatment Effects by Productivity Type

	Total Time	Manager Time	Staff Time
Low Manager, Low Office	-3.34^{***}	-0.72^{***}	-2.53^{***}
	(0.89)	(0.19)	(0.75)
Low Manager, High Office	1.13	-0.29^{***}	1.38^{**}
	(0.72)	(0.08)	(0.66)
High Manager, Low Office	-3.11^{***}	−0.00	-3.00^{***}
	(0.83)	(0.20)	(0.68)
High Manager, High Office	1.81^{**}	0.20	1.51^{**}
	(0.71)	(0.12)	(0.64)
Observations	221,718	221,718	221,718

Notes: Panel A shows the average treatment effect of shortened deadlines on processing time (days). Panel B restricts to the subsample with both manager and office AKM fixed effects. Panel C shows implied treatment effects at one standard deviation above/below the mean of manager and office productivity (see Appendix I.8 for underlying regression coefficients). Standard errors two-way clustered by service and week in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The AKM and movers design suggest that office productivity plays a central role in processing time variation, raising the question of whether treatment effectiveness also varies with office productivity. Ranking offices by their pre-intervention AKM fixed effects estimated from staff processing time,[26] we find that offices more than one standard deviation above the mean already

[26]Office productivity is the within-component normalized AKM fixed effect estimated from staff processing time using pre-RCT data (2021 through December 14, 2022). Higher values indicate faster processing. See Appendix I.7 for details.

met the new RCT-mandated deadlines comfortably before the intervention (93% compliance, mean slack of 9.8 days), while those one standard deviation below the mean were much closer to the new deadlines (46% compliance, mean slack of -1.2 days; Appendix I.7). A natural question, then, is whether treatment effects vary with office—or manager—productivity.

Heterogeneity: We exploit our decomposition of processing time into manager and staff components to construct *actor-specific* productivity measures: manager productivity is the AKM fixed effect estimated from manager processing time, while office productivity is the AKM fixed effect estimated from staff processing time. This approach ensures that manager and office productivity capture distinct dimensions of performance—a manager's decision-making speed versus the efficiency of office-level administrative processes.[27] Both measures are estimated using pre-RCT data using the covariance shrinkage multiple component approach (Best et al., 2023b). We estimate

$$y_{isjt} = \psi_{sj} + \alpha_{si} + \tau_t + \alpha M_i O_j + D_{st}\big(\beta_0 + \beta_1 M_i + \beta_2 O_j + \beta_3 M_i O_j\big) + \epsilon_{isjt} \tag{4}$$

where $D_{st} = \text{Treat}_s \times \text{Post}_t$, M_i is manager i's productivity, and O_j is office j's productivity. Both measures are standardized (mean zero, unit standard deviation), with higher values indicating faster processing. All other terms are as in Equation (3).

Panel C reveals that office productivity is the key moderator of treatment response. In low-productivity offices, deadline reductions produce large and persistent declines in total processing time regardless of manager quality: -3.34 days ($p = 0.001$) with low-productivity managers and -3.11 days ($p = 0.001$) with high-productivity managers. In contrast, high-productivity offices—where pre-treatment slack averaged nearly 10 days (Appendix I.7)—show no improvement under either manager type, with point estimates of 1.13 days ($p = 0.130$) and 1.81 days ($p = 0.016$). Since these offices were already processing well within the new deadlines, the treatment provided little scope for improvement. The increase in processing times for high productivity offices with high productivity managers ($+1.81$ days, $p = 0.016$) is notable and may reflect reallocation of effort toward newly binding treated services.[28] Strikingly, holding office type fixed, manager productivity has essentially no bearing on the magnitude of the treatment effect: both high- and low-productivity managers achieve similar total reductions in slow offices.

While the total effects in low-productivity offices are similar across manager types, the channels differ. Low-productivity managers speed up both their own processing (-0.72 days, $p = 0.001$) and staff time (-2.53 days, $p = 0.002$). High-productivity managers—already fast themselves—show no change in their own time ($p = 0.984$), but their staff respond just as strongly (-3.00 days, $p = 0.0001$). In both cases, the bulk of the improvement comes from staff times.[29]

[27] AKM estimation requires connected components through manager mobility. Using the pre-RCT period (2021 through December 14, 2022) and multiple connected components, we identify both measures for 54% of observations (N=221,718). See Appendix I.6 for a comparison of the AKM and non-AKM samples and Table OA-6 for the variance decomposition.

[28] For comparison with Panel A, Panel B of Table 3 shows the main treatment effect in the heterogeneity sample. The point estimates are somewhat larger but remain insignificant at conventional levels (-0.89 days, $p = 0.230$).

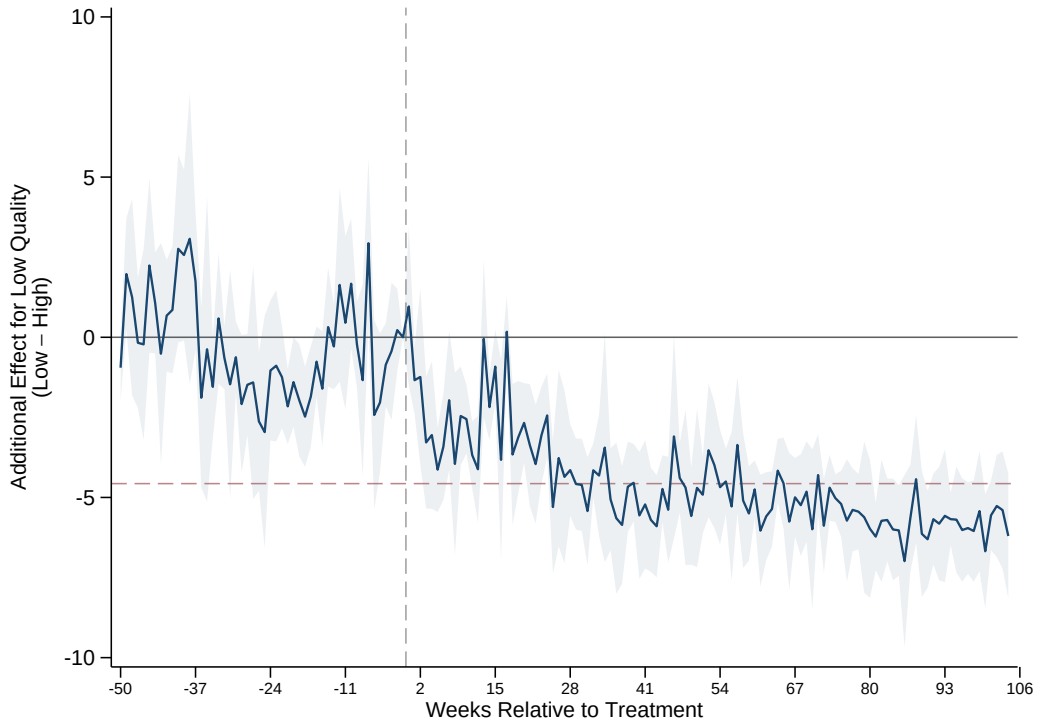
[29] Using private sector data, Kala and McKelway (2026) show that increasing managerial discretion in the allocation

To examine the dynamics of this heterogeneity, we extend the event study in Equation (13) by interacting treatment timing with an indicator for low-quality offices:

$$y_{isjt} = \psi_{sj} + \alpha_{si} + \tau_t + \beta_s \cdot t + \sum_{r \neq -1} \left(\gamma_r + \delta_r \cdot \text{LowQ}_j\right) \cdot \mathbf{1}\{t - t^* = r\} \cdot \text{Treat}_s + \epsilon_{isjt} \tag{5}$$

where $\text{LowQ}_j = \mathbf{1}\{\hat{O}_j \leq 0\}$ is an indicator for offices with below-mean AKM office productivity, estimated using pre-RCT data, τ_t are week fixed effects, and $\beta_s \cdot t$ are service-specific linear time trends. Here γ_r captures the treatment effect for high-quality offices at relative week r, and δ_r captures the *additional* effect for low-quality offices.

Figure 3: Heterogeneous Treatment Effects by Office Quality: Event Study



Notes: Figure shows $\{\hat{\delta}_r\}$ from Equation (5), the additional treatment effect in low-quality offices ($\hat{O}_j \leq 0$) relative to high-quality offices. Reference period is $r = -1$. The outcome is total processing days. Shaded area shows 95% pointwise confidence intervals. Specification includes week fixed effects, service-specific linear time trends, service-by-office, and service-by-manager fixed effects. Standard errors two-way clustered by service and week. Sample truncated at 104 weeks post-treatment. The average post-treatment $\hat{\delta}$ is -4.57 days (SE $= 0.59$, $t = -7.71$, $p < 0.001$).

Figure 3 plots the $\hat{\delta}_r$ coefficients. Pre-treatment, the differential treatment effect between low- and high-quality offices is centered around zero, consistent with parallel pre-trends. At treatment,

of rewards to staff can help increase productivity, but it may come at the cost of reduced workplace interactions and cohesion. In the public sector, such costs may be even higher, and organizational norms may be stronger.

there is a sharp negative break: low-quality offices improve by approximately 4.6 days more than high-quality offices, and this differential persists through the post-treatment period. This confirms that the heterogeneity documented in Panel C of Table 3 is not driven by differential pre-trends and emerges sharply at the time of treatment.

Spillovers: An important caveat is that 74% of managers in our sample process applications for both treatment and control services. If managers learn from processing treatment services and apply faster processing habits to control services, this would attenuate our estimates. Alternatively, we would expect the obverse if managers focus on treatment services at the expense of control services. We present evidence on spillovers in Appendix I.3 suggesting positive spillovers that reduce control service processing times.

Robustness. Panel C uses estimated AKM fixed effects (from pre-RCT data) as regressors. Cross-fitted AKM fixed effects (following Chernozhukov et al., 2018; Kline, 2024, see Appendix I.12) yield very similar standard errors, suggesting that the generated-regressor correction is negligible (Appendix I.11). Since treatment is randomized across only 30 services, we also report randomization inference p-values that do not rely on asymptotic cluster approximations; the key results remain significant. The heterogeneity results are robust to using non-RCT services to estimate AKM effects (Appendix I.10.1) and to using pre-treatment processing speed as an alternative productivity measure that does not require AKM estimation (Appendix I.10.2). The results are also robust to alternative standard error clustering (Table OA-14), to different transformations of the outcome variable ($\log(1+y)$ and $\text{asinh}(y)$, Table OA-15), and to the selective inclusion of the various fixed effects (Appendix I.4).[30]

5.4 Discussion

Office effects dominate. Our core finding is that office effects explain the bulk of productivity variation in Punjab—more so than in comparable settings. Office effects account for 83% of the combined manager-office variance in total processing time, compared to 76% in Italian social security offices (Fenizia, 2022) and 55% in Russian procurement (Best et al., 2023b).[31] A distinctive feature of our data is the ability to decompose processing time into the component attributable to the manager and that attributable to subordinate staff. This reveals that office effects constrain even activities over which managers have the most control: while the office share drops from 79% for staff time to 66% for manager time, office effects still explain about two-thirds of the combined variance for managers' own processing. By contrast, Bertrand and Schoar (2003) find that C-suite executives explain over 4 percentage points of R^2 on average (and up to 37pp for discretionary policies like cost-cutting)—substantially larger individual effects than we observe even for managers' own time.

[30] Treatment effects on other outcomes (delay rates and the number of applications processed) are presented in Table OA-16.

[31] In terms of total outcome variance, offices explain 58% in our setting, compared to 29% (offices) and 9% (managers) in Fenizia (2022) and 26% (organizations) and 21% (bureaucrats) in Best et al. (2023b).

Immediate convergence. The event study reinforces this finding. When managers transfer, they close essentially all of the origin–destination productivity gap immediately, with no evidence of gradual adjustment in either direction. This immediacy suggests that location-based factors—resources, administrative procedures, institutional norms—directly constrain productivity rather than operating through slow channels like learning or relationship-building.[32] The pattern extends to subordinate staff ($\hat{\theta}_0 = 0.988$), confirming that staff productivity does not vary with the identity of the incoming manager.

Limited positive assortative matching The cross-stage correlation between manager effects (from Manager Time) and office effects (from Staff Time) is positive ($\rho = 0.17$, $p < 0.05$), suggesting modest positive assortative matching.[33] However, even favorable manager–office matches have limited impact: offices receiving top-quartile improvements in manager productivity through transfers show gains of only approximately 0.2 log points in processing time (Appendix G.2).

Experimental evidence underscores importance of location The RCT provides complementary evidence on the role of office-level constraints. While high-productivity offices were already operating well below new deadlines (Appendix I.7), it was not *a priori* clear that low-productivity offices would respond to mandated deadline reductions at all—these offices could be slow for deep structural reasons (resource constraints, local political economy) impervious to administrative mandates. That they improve by approximately 3.2 days is informative; that the improvement comes primarily through *staff* processing time is more so. High-productivity managers in slow offices—already fast themselves—show no change in their own time, but their staff respond strongly (-3.00 days), while low-productivity managers speed up both their own processing (-0.72 days) and staff time (-2.53 days). This channel decomposition is not predicted by a purely mechanical ceiling interpretation: the ceiling story can explain why fast offices do not respond, but has nothing to say about which actor responds in slow offices. The fact that staff time—over which the transferring manager has limited direct control—drives the bulk of the improvement is consistent with our observational finding that location-based factors, operating in part through relatively immobile staff, are a central determinant of office productivity. Suggestively, point estimates for high-productivity offices are positive rather than zero, hinting at behavioral responses to deadline changes even where the new mandates are not binding.

6 Conclusion

Understanding the sources of the wide variation in government performance is a first-order policy and research question. Using a comprehensive dataset of digital service requests and manager transfers across Punjab's digital governance system, we employ both observational and ex-

[32] The immediacy of convergence also supports our identification assumptions. If transfers were strategically timed to coincide with performance shocks, we would expect gradual adjustments rather than the sharp, sustained shifts we observe.

[33] This is more similar to the positive sorting found in Russian procurement (Best et al., 2023b) ($\rho = 0.31$) than the negative sorting in Italian social security offices (Fenizia, 2022) ($\rho = -0.53$).

perimental approaches to decompose government productivity variation into manager and office components.

Bias-corrected AKM variance decompositions show that location effects are substantially larger than manager effects: location variance is nearly five times manager variance and accounts for over half of variation in service delivery time. A movers design shows that total office productivity remains largely unchanged when managers transfer. Consistent with this, support-staff productivity is invariant to the identity of the supervising manager, suggesting that relatively immobile support staff are a key location-specific factor. Identifying the constraints to their performance is an important next step in uncovering the sources of the strong location-based effects.

Our experimental evidence reinforces these observational findings. A state-wide RCT that permanently tightened delivery deadlines substantially improved productivity in low-performing offices, regardless of manager quality. This provides direct evidence that office-level factors mediate the effectiveness of policy interventions, consistent with the dominance of location-based effects in our variance decomposition and movers analysis.

Our results also have potentially important implications for public administration reform. The dominance of location effects suggests that policies focused on managers (e.g., incentives, training) may benefit from being complemented with policies addressing underlying office-level institutional constraints.[34]

[34] Khan (2025) suggests one approach for government bodies with well defined missions.

References

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333. 0, 2

AHUJA, A. (2019): *Mobilizing the Marginalized: Ethnic Parties without Ethnic Movements*, Modern South Asia, Oxford, New York: Oxford University Press. 1

AKERLOF, G. A. AND R. E. KRANTON (2000): "Economics and Identity," *Quarterly Journal of Economics*, 115, 715–753. 2, OA - 1

ANDREWS, M. J., L. GILL, T. SCHANK, AND R. UPWARD (2008): "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society Series A: Statistics in Society*, 171, 673–697. 12, 13, OA - 15

BANERJEE, A., E. DUFLO, C. IMBERT, S. MATHEW, AND R. PANDE (2020): "E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in India," *American Economic Journal: Applied Economics*, 12, 39–72. 2

BARTESKA, P. AND J. E. LEE (2025): "Personnel is Policy (Implementation): Bureaucrats and the Korean Export Miracle," CEP Discussion Paper dp2099, Centre for Economic Performance, LSE. 2

BATTAGLIA, L., T. CHRISTENSEN, S. HANSEN, AND S. SACHER (2025): "Inference for Regression with Variables Generated by AI or Machine Learning," Discussion Paper 2421, Cowles Foundation for Research in Economics, Yale University. OA - 43, OA - 44

BERNHEIM, B. D. (1994): "A Theory of Conformity," *Journal of Political Economy*, 102, 841–877. OA - 1

BERTRAND, M. AND A. SCHOAR (2003): "Managing with Style: The Effect of Managers on Firm Policies," *The Quarterly Journal of Economics*, 118, 1169–1208. 22

BEST, M., A. FENIZIA, AND A. Q. KHAN (2023a): "Government analytics using administrative case data," *The Government Analytics Handbook: Leveraging Data to Strengthen Public Administration. World Bank*. 3

BEST, M. C., J. HJORT, AND D. SZAKONYI (2023b): "Individuals and Organizations as Sources of State Effectiveness," *American Economic Review*, 113, 2121–2167. 2, 11, 12, 13, 20, 22, 23, OA - 15, OA - 35, OA - 36

BLUME, L. E., W. A. BROCK, S. N. DURLAUF, AND Y. M. IOANNIDES (2011): "Identification of Social Interactions," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, North-Holland, vol. 1, 853–964. OA - 1

BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009): "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 150, 41–55. OA - 1

BURGESS, R., R. JEDWAB, E. MIGUEL, A. MORJARIA, AND G. P. I MIQUEL (2015): "The Value of Democracy: Evidence from Road Building in Kenya," *American Economic Review*, 105, 1817–1851. 1

CALLAWAY, B. AND P. H. C. SANT'ANNA (2021): "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225, 200–230. 18

CARD, D., P. KLINE, AND J. HEINING (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128, 967–1015. 2, OA - 36

CHEN, J. AND J. ROTH (2024): "Logs with Zeros? Some Problems and Solutions," *Quarterly Journal of Economics*, 139, 891–936. OA - 34

CHERNOZHUKOV, V., D. CHETVERIKOV, M. DEMIRER, E. DUFLO, C. HANSEN, W. NEWEY, AND J. ROBINS (2018): "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal*, 21, C1–C68. 22, OA - 43

CHETTY, R. AND N. HENDREN (2018): "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *Quarterly Journal of Economics*, 133, 1107–1162. 2

CHHIBBER, P. AND I. NOORUDDIN (2004): "Do Party Systems Count? The Number of Parties and Government Performance in the Indian States," *Comparative Political Studies*, 37, 152–187. 1

CHUNG, E. AND J. P. ROMANO (2013): "Exact and Asymptotically Robust Permutation Tests," *The Annals of Statistics*, 41, 484–507. OA - 45

DASGUPTA, A. AND D. KAPUR (2020): "The Political Economy of Bureaucratic Overload: Evidence from Rural Development Officials in India," *American Political Science Review*, 114, 1316–1334. 1, OA - 6

DE CHAISEMARTIN, C. AND X. D'HAULTFŒUILLE (2020): "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–2996. 18

DIPOPPA, G. AND S. GULZAR (2024): "Bureaucrat Incentives Reduce Crop Burning and Child Mortality in South Asia," *Nature*, 634, 1125–1131. 2

DODGE, E., Y. NEGGERS, R. PANDE, AND C. M. T. MOORE (2025): "From Delay to Payday: Easing Bureaucrat Access to Implementation Information Strengthens Social Protection Delivery," Tech. rep., National Bureau of Economic Research. 2

FENIZIA, A. (2022): "Managers And Productivity In The Public Sector," *Econometrica*, 22. 2, 22, 23, OA - 28

FENSKE, J., M. HASEEB, AND N. KALA (2023): "How rules and compliance impact organizational outcomes: Evidence from delegation in environmental regulation," Tech. rep., National Bureau of Economic Research. 2

FERRAZ, C. AND F. FINAN (2011): "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments," *American Economic Review*, 101, 1274–1311. 1

FINAN, F., B. A. OLKEN, AND R. PANDE (2015): "The Personnel Economics of the State," . 2

FINKELSTEIN, A., M. GENTZKOW, AND H. WILLIAMS (2016): "Sources of Geographic Variation in Health Care: Evidence From PatientMigration," *The Quarterly Journal of Economics*, 131, 1681–1726. 2, 10

GENICOT, G., Z. MANSOOR, AND G. MANSURI (2026): "Hierarchy and Performance Pay: Experimental Evidence from the Public Sector," Tech. rep., National Bureau of Economic Research. 2

GOODMAN-BACON, A. (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225, 254–277. 18

GOVERNMENT OF PUNJAB (2018): "The Punjab Transparency and Accountability in Delivery of Public Services Act, 2018," https://prsindia.org/files/bills_acts/acts_states/punjab/2018/Act16of2018.pdf, punjab Act No. 16 of 2018, enacted on July 17, 2018. 3

HOLMSTRÖM, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340. 2

HOLMSTRÖM, B. AND P. MILGROM (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, And Job Design," *Journal of Law, Economics, & Organization*, 7, 24–52. 7

IYER, L. AND A. MANI (2012): "Traveling Agents: Political Change and Bureaucratic Turnover in India," *Review of Economics and Statistics*, 94, 723–739. 1, 4

KALA, N. AND M. MCKELWAY (2026): "Power to the Personnel? The Impacts of Managerial Discretion vs. Worker Democracy in Employee Recognition," . 20

KAPUR, D. (2020): "Why Does the Indian State Both Fail and Succeed?" *Journal of Economic Perspectives*, 34, 31–54. 2

KHAN, M. Y. (2025): "Mission Motivation and Public Sector Performance: Experimental Evidence from Pakistan," *American Economic Review*, 115, 2343–2375. 2, 24

KLINE, P., R. SAGGIO, AND M. SØLVSTEN (2020): "Leave-Out Estimation of Variance Components," *Econometrica*, 88, 1859–1898. 12, 13, OA - 15, OA - 17, OA - 43

KLINE, P. M. (2024): "Firm Wage Effects," . 2, 10, 11, 22, OA - 43, OA - 44

——— (2025): "Branching Fixed Effects," Working Paper, UC Berkeley. OA - 43

KOHLI, A. (2004): *State-Directed Development: Political Power and Industrialization in the Global Periphery*, New York: Cambridge University Press. 1

KREPS, D. M. (1990): "Corporate Culture and Economic Theory," in *Perspectives on Positive Political Economy*, ed. by J. E. Alt and K. A. Shepsle, Cambridge University Press, 90–143. 2

KUMAR, S. AND M. SINGH (2024): "E-GOVERNANCE IN PUNJAB: AN ANALYSIS OF STRATEGIES AND ITS IMPACT ON CITIZEN SERVICES," *ShodhKosh: Journal of Visual and Performing Arts*, 5, 2941–2950. 3

LIPSKY, M. (1980): *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*, New York: Russell Sage Foundation. 2

PAGAN, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247. OA - 43

RASUL, I. AND D. ROGGER (2018): "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *The Economic Journal*, 128, 413–446. 2

ROGGER, D. AND C. SCHUSTER (2023): *The government analytics handbook: Leveraging data to strengthen public administration*, World Bank Publications. 2

ROTH, J. (2022): "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, 4, 305–22. 12

RUBIN, D. B. (1987): *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley. OA - 45

SUN, L. AND S. ABRAHAM (2021): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225, 175–199. 18

SUPREME COURT OF INDIA (2013): "TSR Subramanian & Ors vs Union of India & Ors," Supreme Court of India, civil Appeal No. 6736 of 2007. 4

T. RYAN ROGERS AND ANDREW HALL (2020): "Political Competition and Public Service Delivery: Evidence from Electricity Provision in India," *American Economic Journal: Applied Economics*, 12, 71–97. 1

TIROLE, J. (1986): "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, and Organization*, 2, 181–214. 2

VIVEK, S. (2014): *Delivering Public Services Effectively: Tamil Nadu and Beyond*, Oxford, New York: Oxford University Press. 1

Managers and Offices: Location-Based Determinants of Bureaucratic Effectiveness in India

Prabhat Barnwal, Aaditya Dar , Aprajit Mahajan, Maitri Punjabi , Parminder Pal Singh Sandhu

Appendix – For Online Publication

Contents

A Deriving the Estimating Equation . . . . . **OA - 1**

B Context: Officer Cadres . . . . . **OA - 2**
B.1 Staff Mobility OA - 3

C Data Description . . . . . **OA - 3**

D Variation in Service Delivery across India . . . . . **OA - 5**

E Transfer Timing and Variation in Delta . . . . . **OA - 7**

F AKM Robustness . . . . . **OA - 10**
F.1 Event Study: Delay Rates OA - 10
F.2 Event Study Heterogeneity by Office Productivity OA - 11
F.3 Variance Decomposition by Department OA - 12

G Assessing Assortative Matching . . . . . **OA - 17**
G.1 AKM Correlations with Differing Outcomes OA - 17
G.2 Office Productivity by Changes in Manager Productivity OA - 17

H Counterfactual Analysis: Framework and Assumptions . . . . . **OA - 28**
H.1 Framework OA - 28

I RCT: Experimental Evidence on Deadline Reductions . . . . . **OA - 30**
I.1 RCT Design and Sample OA - 30
I.2 Main Treatment Effect Event Study OA - 31
I.3 Spillover Analysis OA - 32
I.4 Robustness OA - 33
I.5 Constructing Manager and Office Productivity Measures OA - 35
I.6 AKM vs Non-AKM Sample Comparison OA - 37
I.7 Pre-Treatment Slack by Office Productivity OA - 38
I.8 Heterogeneity Regression Coefficients OA - 39
I.9 Heterogeneity Event Study OA - 39
I.10 Robustness of Productivity Measures OA - 39
I.11 Alternative Inference for Table 3 Panel C OA - 42
I.12 Cross-Fitted AKM Fixed Effects OA - 43
I.13 Bootstrap Procedure for Table 3 Standard Errors OA - 46

A Deriving the Estimating Equation

We present a stylized characterization of the manager's choice of productivity level in period t while in office j by assuming utility is of the form:

$$U_{ijt}(y_{ijt}) \equiv w - \eta_{ijt}(y_{ijt}) - C_j(y_{ijt}) \tag{6}$$

where w is the (fixed) wage.

$\eta_{ijt}(\cdot)$ is the (potentially period- and location-specific) disutility of effort required for productivity y (e.g. number of applications processed in the period) and captures the notion that productivity requires costly effort. We parameterize these utility costs as convex: $\eta_{ijt}(y) = (\tilde{\eta}_i + \tilde{\tau}_t + \tilde{\epsilon}_{ijt})y + \frac{y^2}{2}$

$C_{jt}(y)$ captures the office-level component of the cost of productivity level y and is a reduced form approach to capturing the location effect on productivity. It could reflect conformity pressures in office j—a penalty for deviating from the office norm. This could reflect institutional pressures or social norms within offices (e.g. managers could face social repercussions for being "too slow" or "too fast" relative to the office norm).[35] It could also reflect location specific constraints to productivity (e.g. resources). Since we cannot distinguish between these components of the location effect, we subsumed them into a single measure. We parameterize the penalty function as $C_{jt}(y) = \frac{\tilde{\phi}}{2}\mathbb{E}_i^e\left(y - \bar{y}_{-ij}\right))^2$ where $\tilde{\phi} > 0$ and $\mathbb{E}_i^e(\cdot)$ is manager i's belief about the average productivity level of all other managers in office j (i.e. excluding manager i's productivity) $\bar{y}_{-ij}$. This quadratic penalty specification (and using the group mean as reference) aligns with canonical models in the social norms (Akerlof and Kranton, 2000; Bernheim, 1994) and social networks (Blume et al., 2011; Bramoullé et al., 2009) literatures.

Appendix A1 in Blume et al. (2011) shows that our estimating equation can be derived as the Bayes Nash Equilibrium of a game in which each manager i in office j maximizes eq. (6) given their beliefs (and that in equilibrium, agent beliefs will agree and be correct).[36] Assuming such a BNE with correct beliefs, we can obtain our estimating equation by characterizing the first-order conditions for maximizing $U_{ijt}(y)$ over y. In particular,

$$\begin{aligned} y_{ijt} &= \frac{1}{1+\tilde{\phi}}\left(\tilde{\phi}\bar{y}_{-ij} - \tilde{\eta}_i - \tilde{\tau}_t - \tilde{\epsilon}_{ijt}\right) \\ &= \eta_i + \phi\bar{y}_{-ij} + \tau_t + \epsilon_{ijt} \end{aligned}$$

where we have relabeled terms to ease exposition.

For a mover i who moves from origin office $o(i)$ to destination office $d(i)$, their productivity (for the relevant period) can be written (where we suppress the exclusion of i from the office means):

$$y_{ijt} = \eta_i + \phi\bar{y}_{o(i)} + \tau_t + \epsilon_{ijt} \quad \text{(for } t \text{ before move)} \tag{7}$$

$$y_{ijt} = \eta_i + \phi\bar{y}_{d(i)} + \tau_t + \epsilon_{ijt} \quad \text{(for } t \text{ after move).} \tag{8}$$

Define δ_i as the difference in average productivity between destination and origin offices :

$$\delta_i \equiv \bar{y}_{d(i)} - \bar{y}_{o(i)} \tag{9}$$

[35] Wages are fixed here so the only relevant trade-off is between the cost of effort $\eta(\cdot)$ and the conformity penalty $C(\cdot)$.

[36] Empirically, we will approximate this belief with the office's long-run productivity (excluding manager i's own observations.

We can re-write the equation in a unified form as:

$$y_{it} = \eta_i + \phi \bar{y}_{o(i)} + \mathbb{I}_{\{r(i,t)>0\}} \phi \delta_i + \tau_t + \epsilon_{ijt} \tag{10}$$

where $\mathbb{I}_{\{r(i,t)>0\}}$ is an indicator for being post-move. Replacing the single indicator function with a set of indicators for leads and lags relative to period of the transfer we obtain our main estimation equation:

$$y_{it} = \alpha_i + \theta_{r(i,t)} \delta_i + \tau_t + \epsilon_{ijt} \tag{11}$$

and where α_i combines η_i and $\phi \bar{y}_{o(i)}$. We can augment this model with covariates x_{ijt} by adding them to $\eta_{ijt}(\cdot)$. Identification for the $\theta_{r(i,t)}$ parameters are discussed in the main text.

B Context: Officer Cadres

The managers in our study represent distinct administrative cadres with varying hierarchical positions and transfer patterns.

Deputy Commissioners (DCs) constitute the senior-most district-level administrators, typically drawn from the Indian Administrative Service (IAS) in their first 6-8 years of service or are promoted from Provincial Civil Service ranks. DCs possess state-wide transfer liability and can be posted across any of Punjab's 23 districts.

Sub-Divisional Magistrates (SDMs) serve as the Deputy Commissioner's principal agents at the sub-divisional level (within a district), managing administration, development activities, law and order, and public grievances across 93 sub-divisions statewide. SDMs possess broad magisterial powers and coordinate multiple departments within their jurisdiction.

Tehsildars and Naib Tehsildars (T/NT) are revenue officers operating at the tehsil and sub-tehsil levels respectively. They focus primarily on revenue administration including land records management, property mutations, revenue collection, and recovery of government dues. Currently, about 80 Tehsildars operate across 101 tehsil offices while 160 Naib Tehsildars serve across 180 sub-tehsil offices.

Medical Officers and **Block Development Officers** represent specialized functional cadres with more limited geographic mobility. Medical Officers typically transfer within regional health jurisdictions, while Block Development Officers usually serve within specific development blocks or related rural development administrative units.

Executive Officers of Municipal Committees are generally confined to urban local body jurisdictions within their respective districts or municipal regions.

Transfers are intended to reflect a trade-off between preventing local entrenchment and maintaining service continuity. Managers rarely exit the administrative sector voluntarily (dismissals and resignations are uncommon), with most serving until statutory retirement. In our sample, the average observed posting duration is approximately 47 weeks (Table 1b); this is a lower bound on actual tenure since postings extending beyond our panel are right-censored.

B.1 Staff Mobility

In contrast to managers, subordinate staff—the dealing clerks, patwaris, and health workers who process the bulk of citizen applications—transfer far less frequently. Table OA-1 compares transfer rates for managers and four categories of subordinate staff using the same eSewa administrative records. Only 23% of staff transferred during our sample period, compared to 69% of managers. Among staff who do transfer, the average number of postings is 1.3 (vs. 3.3 for managers) and average posting duration is 2.4 years (vs. 0.9 years for managers). ANMs are the least mobile (6% transfer rate), while patwaris—revenue field staff who, like tehsildars, are part of the revenue department—transfer most frequently among staff categories (34%).

The manager count in Table OA-1 (3,407) exceeds the main analysis sample (2,290 in Table 1a) because the transfer data includes all managers with eSewa administrative assignments, including those posted to offices that do not process citizen services (e.g., purely administrative or coordinating roles). The main analysis sample restricts to managers observed processing at least one application.

Table OA-1: Transfer Rates: Managers vs. Subordinate Staff

	Mean Postings	Transfer Rate (%)	Mean Posting Duration (yrs)	N
Managers	3.28	69	0.83	3,407
Staff				
Dealing clerks	1.30	21	2.48	4,149
Patwaris	1.52	34	0.91	5,605
ANMs	1.08	6	4.92	3,685
Seva sahayaks	1.27	21	0.62	401
All staff	1.33	23	2.38	13,840

Notes: Data from eSewa administrative records on manager and staff postings. A "transfer" is defined as serving in two or more distinct offices during the sample period (2021–2025). We restrict to substantive postings (excluding additional charge assignments) and drop records with missing or inconsistent dates. Each row reports statistics at the person level after collapsing to one observation per person-office spell. Managers are defined as in the main analysis (Section 2.2.2). Staff categories include dealing clerks (clerical staff processing applications), patwaris (revenue field staff), ANMs (auxiliary nurse midwives in health offices), and seva sahayaks (service assistants). We exclude Sewa Kendra operators (who work across departments as platform operators rather than office staff) and elected or appointed village-level officials (sarpanch, nambardars, municipal councilors).

C Data Description

Manager Transfer Data: This dataset is drawn from eSewa, Punjab's backend e-governance system, which records each manager's administrative assignments. It contains an observation for each manager's assignment, including the start and end dates of each assignment and the office ID of the associated office. We calculate each manager's posting duration by counting the days between the start and end dates for each posting (Table 1b).

Productivity Data: There are two primary datasets that we use to calculate our final productivity measures. Both of these datasets also come from the eSewa database. We use **application-stage level** data, which includes a timestamped log of each actor's work on an application during its processing history. For example, if an

application was seen by a total of four actors, this dataset includes four observations containing the exact time that each actor received and forwarded the application, allowing us to calculate time taken by each of the four actors along the processing chain. We calculate the number of days taken by the final approving authority and the (collective) number of days taken by all other office staff. About 15% of applications in our full sample were for services that did not involve any office staff, implying this measure is set to missing. We also calculate the *total* number of days taken to process the application using the date the application was filed (and therefore received by the first actor in the chain) and the date it was approved by the manager. We reshape this data to be at the application level and merge it with a separate **service-level** dataset that includes each service's processing deadline. This allows us to measure whether the application took longer to process than the stipulated deadline, giving us a delay indicator for each application.

For our main estimation, we collapse the application-level data to the manager-week level, averaging all productivity measures (processing days, total and by actor type, and delay rate). We replace the averaged productivity measures with 1 if they equal 0 before generating the log variables.

Table OA-2 provides a year-by-year breakdown of the manager-week analysis sample. The sample is roughly balanced across years, with approximately 24,000–27,000 manager-week observations per year. Mean processing days are somewhat lower in 2021 (9.3 days) than in subsequent years (11–13 days), while manager time remains stable at 2.1–2.8 days throughout.

Table OA-2: Summary Statistics by Year

					Mean Processing Days		
Year	Obs	Managers	Offices	Movers	Total	Manager	Staff
2021	26,483	1,114	850	538	10.1	2.4	10.1
2022	26,965	1,236	884	670	13.5	3.0	13.4
2023	25,568	1,022	840	610	12.0	2.8	11.9
2024	23,472	985	770	595	12.0	2.6	12.1
2025	24,457	1,111	834	602	11.9	2.2	11.9
Total	126,945	2,290	992	953	11.9	2.6	11.9

Note: Obs = manager × week observations. Movers = managers observed in two or more offices over the full sample period. Mean processing days are in business days, averaged at the manager-week level. The sample restricts to manager-weeks with non-missing time-varying controls (staff counts, vacancy rates, workload measures), matching the estimation sample in Tables 1a and 2. Mean processing days for each component (total, manager, staff) are computed conditional on that component being observed; because staff and manager time breakdowns are unavailable for some service types (approximately 9% of observations), the component means need not sum to the total.

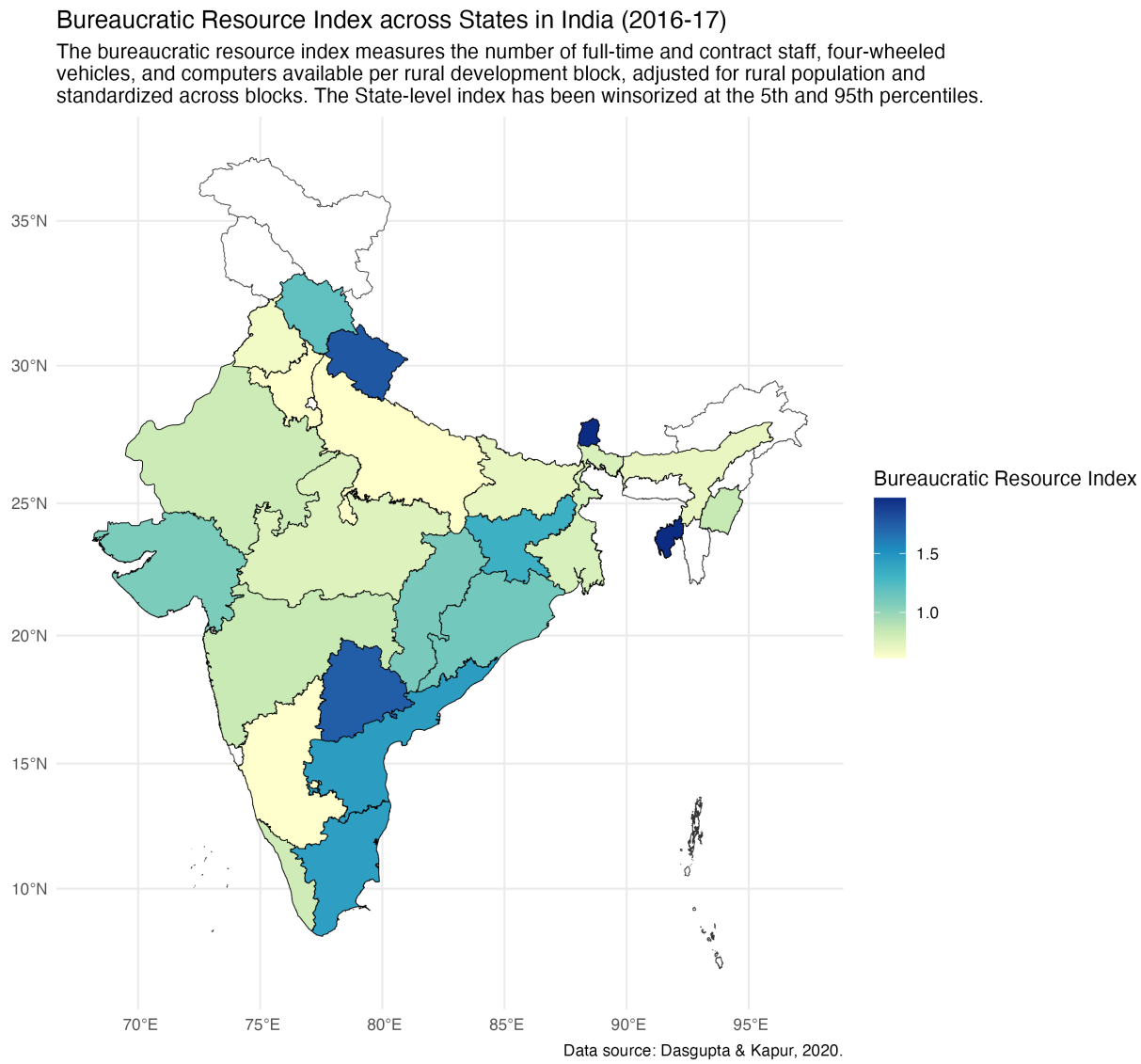
D Variation in Service Delivery across India

Table OA-3: Delivery deadlines matrix by state and document type

State	Domicile	BPL	Caste	Disability	Birth	Death	Income	Driving	Average
Andhra Pradesh		30	0	6	6	7	15	11	11
Assam		15	15	15	15	15		15	15
Bihar	21	60	60	21	10	14	21	21	29
Delhi	14	45	44	7	7	7	14	3	18
Karnataka	7	15	21	7	7	7	21	21	13
Kerala	3	0	3		7	0	6	5	3
Madhya Pradesh	7	30	30	15	15	15	3	15	16
Maharashtra	15	15	17	7	10	21	15	7	13
Punjab	15		7	9	9	9	15		11
Rajasthan	5	7	5	21	10	10	7	2	8
Telangana	6		30		3	3	7		10

Note: Empty cells indicate data not available.

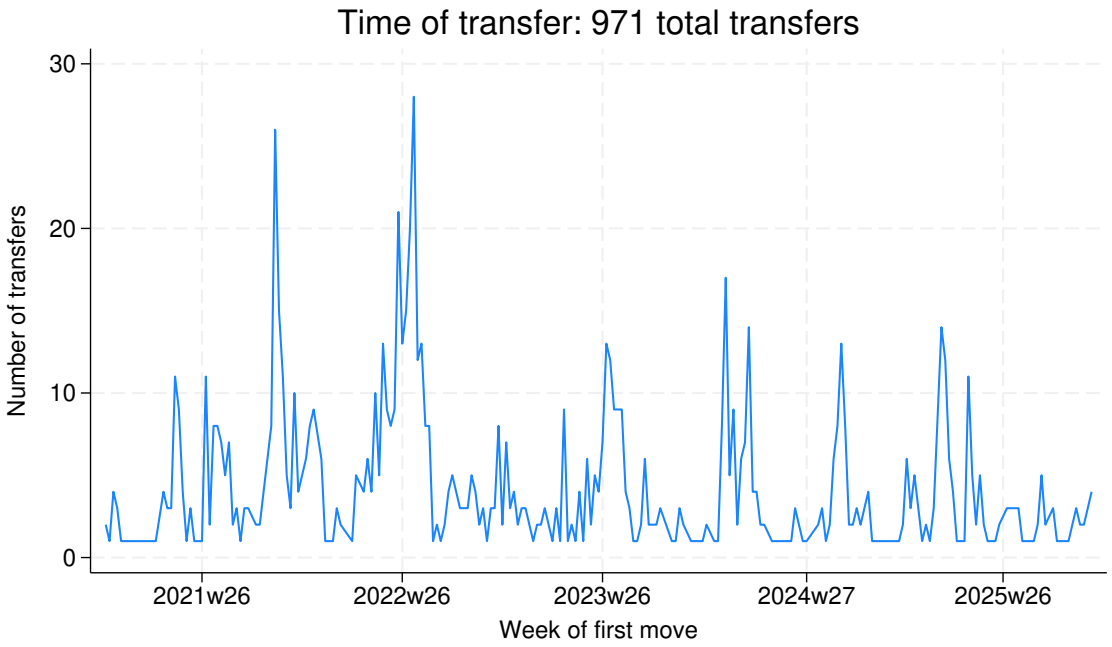
Figure OA-1: Bureaucratic Resource Index Across States



Note: The Bureaucratic Resource Index (BRI) was constructed by Dasgupta and Kapur (2020) to measure the under-resourcing of civil officers across blocks in India, using the number of active full-time employees, contract employees, four-wheel vehicles, and computers in each Block Development Office across 423 blocks. These measures were combined via principal component analysis and normalized by the in-sample standard deviation. **Sign convention:** A *higher* index value indicates greater bureaucrat overload (i.e., fewer resources per officer), not more resources. Punjab's blocks score above the national average, indicating relatively under-resourced local offices.

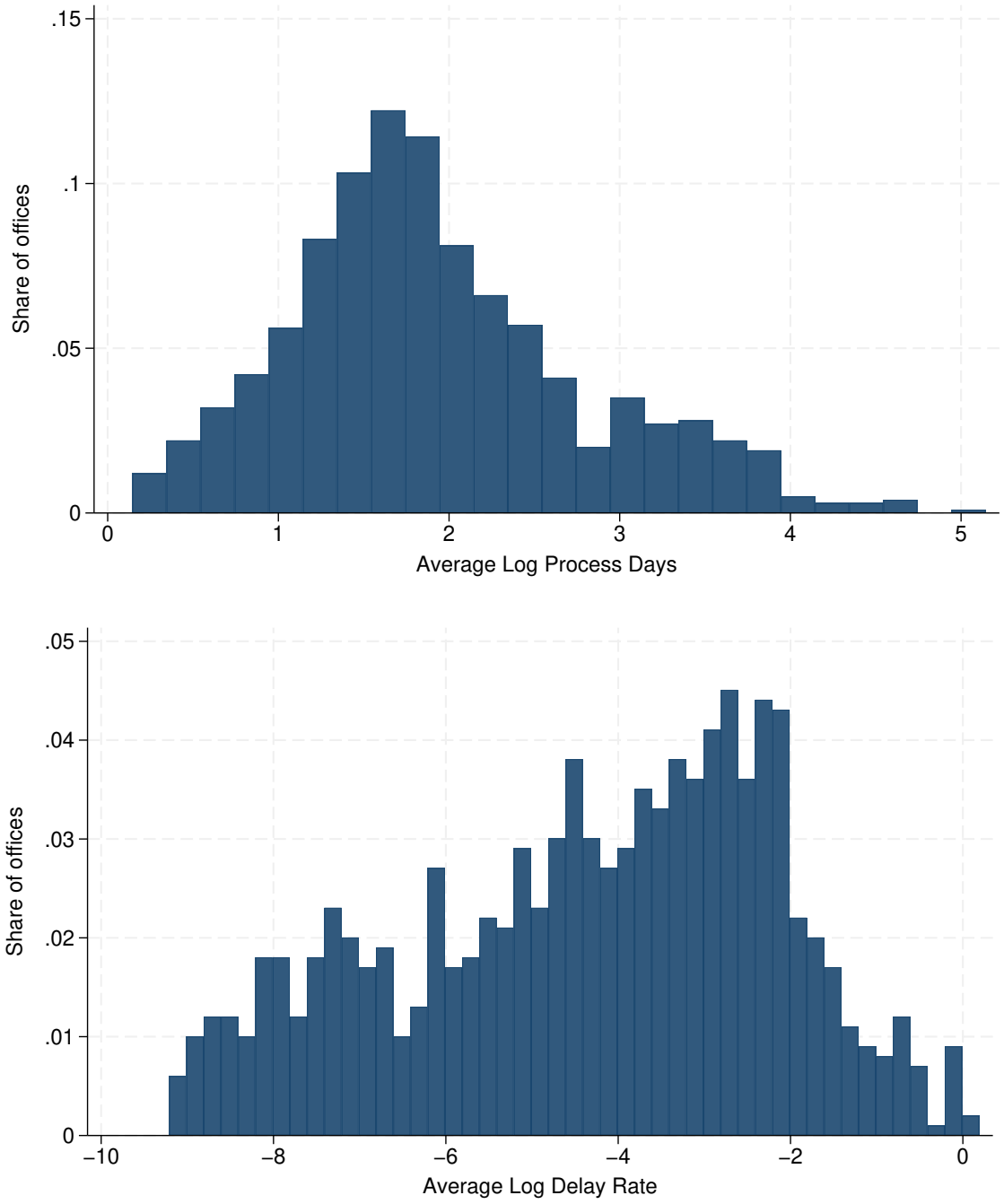
E Transfer Timing and Variation in Delta

Figure OA-2: Transfer Timing



Note: During our sample period (2021–2025), many transfers occurred in the summer of 2022 after a change in the state administration. Excluding this period, there are about 5 transfers per week. There are 994 movers in our sample.

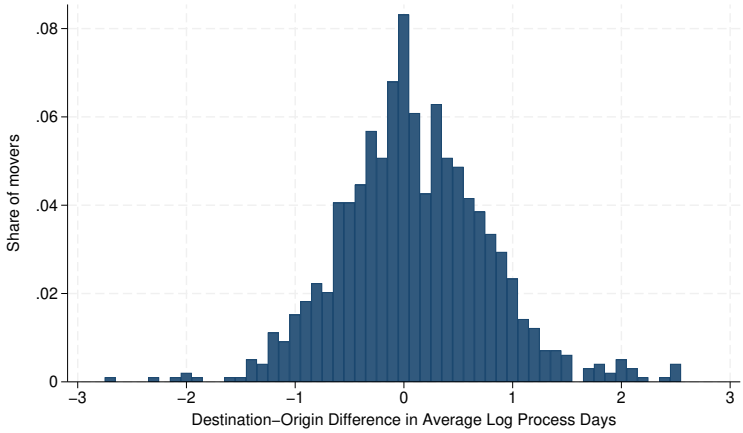
Figure OA-3: Distribution of overall office productivity



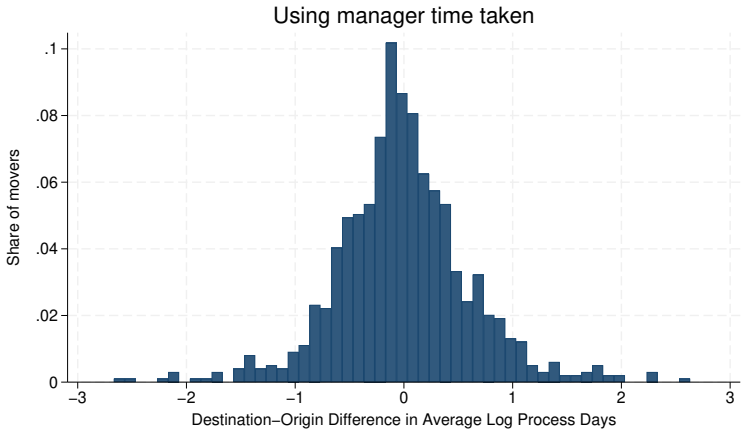
Note: Plots show the distribution of office-level productivity measures used to calculate delta for each mover.

Figure OA-4: Distribution of mover productivity changes (δ_i) by outcome

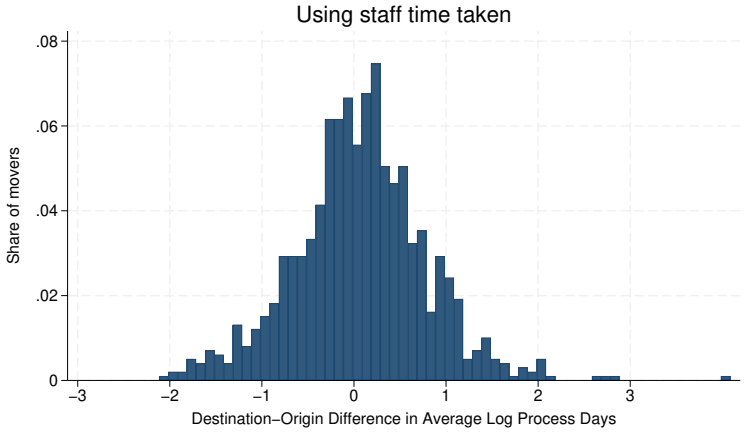
(a) Total processing time



(b) Manager (AA) time

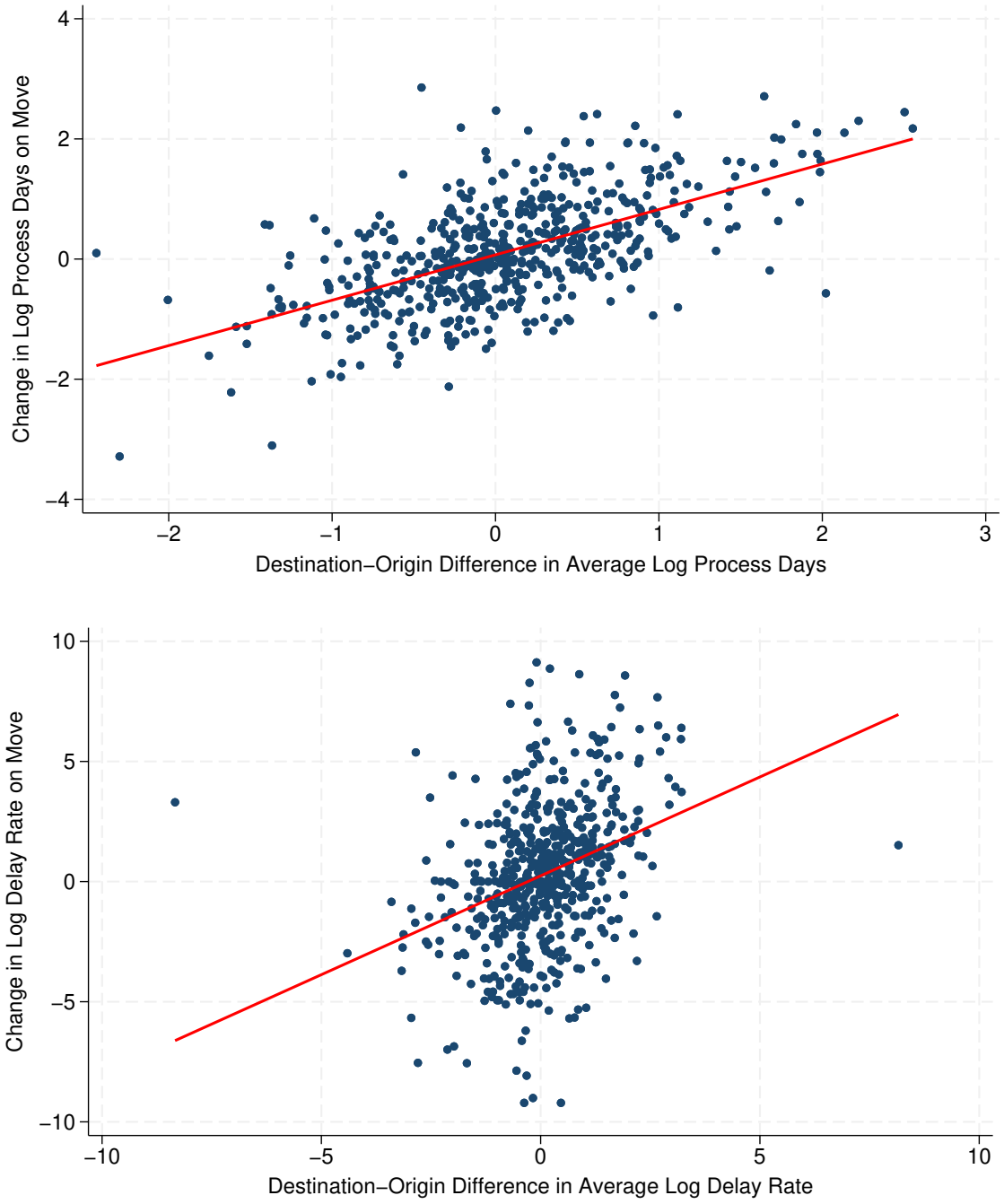


(c) Staff time



Note: Each panel shows the distribution of $\delta_i = \bar{y}_{d(i)} - \bar{y}_{o(i)}$, the difference in mean log processing time between the destination and origin office for each mover i. Panel (a) uses total processing time, panel (b) uses manager (Approving Authority) processing time, and panel (c) uses subordinate staff processing time. Positive values indicate moves to slower offices. The sample includes 994 movers observed in both origin and destination offices during 2021–2025.

Figure OA-5: Relationship between δ_i and manager productivity after transfer



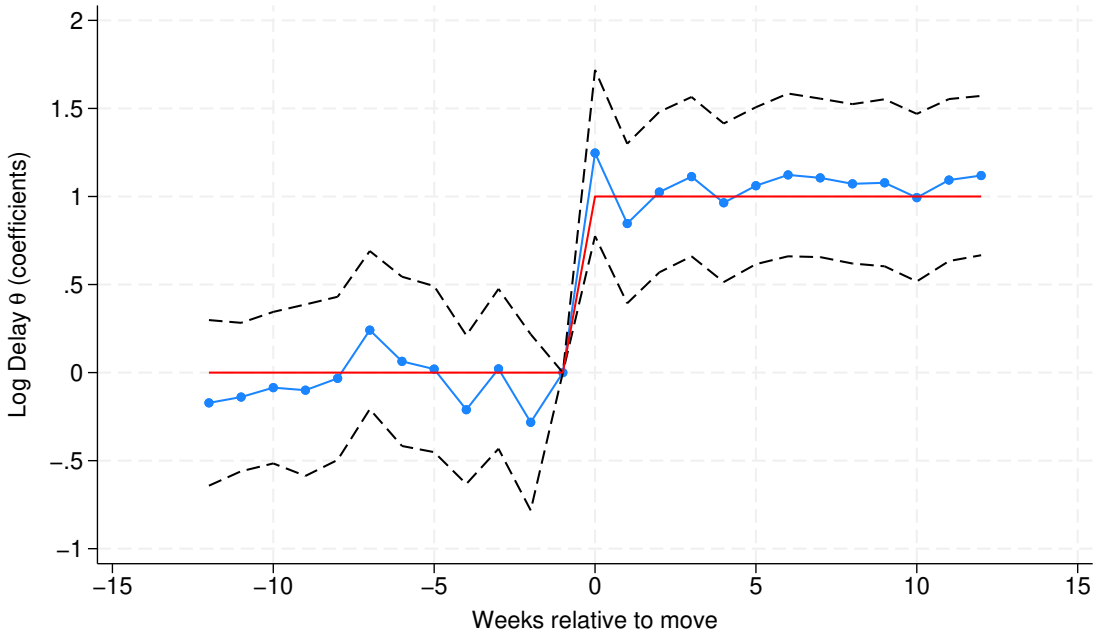
Note: Scatter plots show the first stage relationship between δ_i and the change in the manager's average productivity after the move.

F AKM Robustness

F.1 Event Study: Delay Rates

Figure OA-6 presents the event study using (log) delay rates as the outcome. The pattern is similar to the main event study for (log) processing days (Figure 2): managers' delay rates converge immediately to the destination office level ($\hat{\theta}_0 = 1.247$), with no evidence of gradual adjustment.

Figure OA-6: Location-based share: Log Delay Rate



Note: Figure shows the estimates of the $\theta_{r(i,t)}$ coefficients from our main estimating equation, using log delay rate as the outcome. We normalize the coefficient for the week prior to the transfer to zero. The dashed lines represent 95% pointwise confidence intervals. $\hat{\theta}_0 = 1.247$. There are 994 movers in our sample (n=13,041 bureaucrat-week observations). Delta is calculated using the difference in (log) delay rates in destination and origin offices, excluding manager i's own observations (leave-one-out, as defined in Section 3).

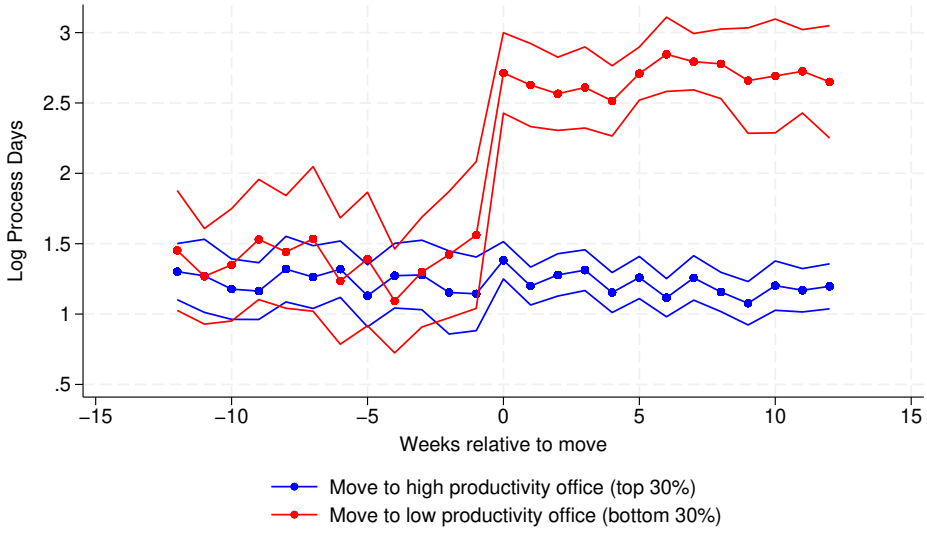
F.2 Event Study Heterogeneity by Office Productivity

Figure OA-7 examines productivity dynamics for moves between offices at different points in the productivity distribution. We classify offices as high or low productivity using their average performance (top and bottom 30%) over the entire sample period.

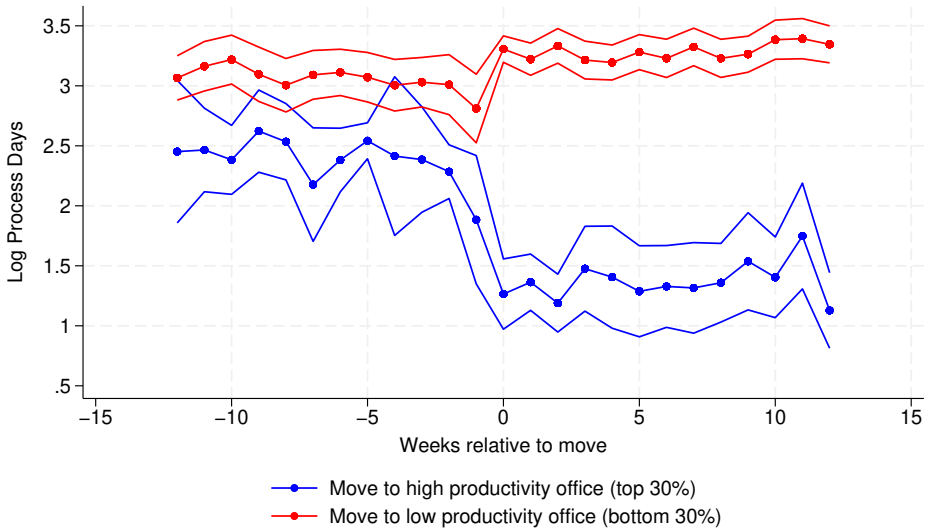
Figure OA-7(a) focuses on managers initially working in high-productivity offices. Before their transfers, all managers exhibit similar productivity levels and trends, regardless of their eventual destination. However, their post-transfer trajectories diverge sharply: those transferred to high-productivity offices (blue line) maintain their strong performance, while those moved to low-productivity offices (red line) immediately experience substantial declines in productivity. Figure OA-7(b) demonstrates the symmetric pattern for managers originating in low-productivity offices.

Figure OA-7: Manager productivity by origin and destination office productivity

(a) Origin: High productivity office



(b) Origin: Low productivity office



Note: These graphs plot weekly average manager productivity over time for managers that move across offices of different productivity levels. We define high productive offices as those that fall in the top 30% in rankings of offices based on $\log(\bar{y})$ for the entire sample period and low productive offices as those that fall in the bottom 30%. Figure OA-7(a) uses 1,136 observations and Figure OA-7(b) uses 1,662 observations. These are unadjusted means, not regression-based estimates.

F.3 Variance Decomposition by Department

Table OA-4 presents variance decomposition results separately for the four departments with sufficiently large connected components (Rural Development, Revenue, Local Government, and Health). Social Welfare is excluded due to insufficient size (40 managers, 22 offices). Office effects dominate in all four departments, with office shares ranging from 75–86% for total processing time. Revenue shows the highest office share (86%), while Local Government and Rural Development show the lowest (75–77%). The pattern is consistent across outcome measures: office shares are highest for total and staff time and more balanced for manager time, mirroring the

pooled results in Table 2.

Table OA-4: Variance Decomposition by Department

	Total Time	Manager Time	Staff Time
Panel A: Rural Development			
Observations	16,641	16,641	16,641
Managers	372	372	372
Offices	151	151	151
$\text{Var}(y)$	0.54	0.16	0.75
$\widehat{\text{Var}}(\psi)$	0.15	0.02	0.19
$\widehat{\text{Var}}(\alpha)$	0.05	0.01	0.06
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.14	-0.30	-0.13
Office Share	0.76	0.61	0.76
Panel B: Revenue			
Observations	39,116	39,116	39,116
Managers	784	784	784
Offices	321	321	321
$\text{Var}(y)$	0.96	0.63	1.25
$\widehat{\text{Var}}(\psi)$	0.60	0.28	0.74
$\widehat{\text{Var}}(\alpha)$	0.10	0.12	0.13
$\widehat{\text{Corr}}(\alpha, \psi)$	0.02	-0.03	0.02
Office Share	0.86	0.70	0.85
Panel C: Local Government			
Observations	16,677	16,677	16,677
Managers	235	235	235
Offices	133	133	133
$\text{Var}(y)$	0.46	0.31	0.70
$\widehat{\text{Var}}(\psi)$	0.21	0.13	0.35
$\widehat{\text{Var}}(\alpha)$	0.07	0.07	0.11
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.36	-0.34	-0.41
Office Share	0.75	0.64	0.76
Panel D: Health			
Observations	15,879	15,879	15,879
Managers	265	265	265
Offices	87	87	87
$\text{Var}(y)$	0.85	0.43	1.20
$\widehat{\text{Var}}(\psi)$	0.53	0.31	0.53
$\widehat{\text{Var}}(\alpha)$	0.13	0.22	0.23
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.32	-0.58	-0.43
Office Share	0.80	0.58	0.70

Notes: Each column uses a different outcome (see Table 2 notes for definitions). Each panel reports covariance shrinkage estimates of the variance decomposition for a single department, using the largest connected component within that department. Departments with fewer than 5,000 observations in the largest connected component are excluded (Social Welfare: 2,432 observations, 34 managers). All regressions include week fixed effects, total staff count, subordinate staff and manager vacancy indicators, log applications received, and number of services handled. Shrinkage uses 100 inner bootstrap iterations. Office Share $= \widehat{\text{Var}}(\psi)/[\widehat{\text{Var}}(\alpha) + \widehat{\text{Var}}(\psi)]$.

We next assess the robustness of our variance decomposition results along two dimensions: sample definition and bias-correction method.

F.3.1 Alternative Sample Definitions

Table OA-5 compares results across three sample definitions. The "Naive" approach uses all data without restricting to connected components. The "Largest CC" approach restricts to the single largest connected component (our preferred approach). The "Multiple CC" approach uses all components meeting minimum size thresholds (at least 6 managers and 3 offices) and adds component fixed effects.

Table OA-5: Variance Decomposition: Robustness to Sample Definition

	Naive	Largest CC	Multiple CC
Panel A: Total Processing Time			
Observations	126,945	71,636	94,904
Managers	2,290	1,419	1,755
Offices	992	559	736
Components	1	1	7
$\widehat{\text{Var}}(\psi)$	0.54	0.60	0.55
$\widehat{\text{Var}}(\alpha)$	0.25	0.13	0.12
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.28	-0.07	-0.17
Office Share	0.69	0.83	0.82
Panel B: Manager Processing Time			
Observations	126,945	71,636	94,904
Managers	2,290	1,419	1,755
Offices	992	559	736
Components	1	1	7
$\widehat{\text{Var}}(\psi)$	0.25	0.28	0.25
$\widehat{\text{Var}}(\alpha)$	0.22	0.15	0.15
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.42	-0.32	-0.32
Office Share	0.53	0.66	0.63
Panel C: Staff Processing Time			
Observations	126,945	71,636	94,904
Managers	2,290	1,419	1,755
Offices	992	559	736
Components	1	1	7
$\widehat{\text{Var}}(\psi)$	0.69	0.77	0.69
$\widehat{\text{Var}}(\alpha)$	0.47	0.20	0.18
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.35	-0.13	-0.17
Office Share	0.60	0.79	0.80

Notes: Each column uses a different outcome (see Table 2 notes for definitions). This table reports covariance shrinkage estimates of the variance decomposition under three different sample definitions. **Naive:** No connected set restriction (all observations). **Largest CC:** Restricted to the largest connected component of the manager-office mobility network. **Multiple CC:** All connected components with at least 6 managers and 3 offices. All regressions include week fixed effects, total staff count, subordinate staff and manager vacancy indicators, log applications received, and number of services handled. Shrinkage uses 100 inner bootstrap iterations. Office Share $= \widehat{\text{Var}}(\psi)/[\widehat{\text{Var}}(\alpha) + \widehat{\text{Var}}(\psi)]$.

Table OA-6 presents the full MCC decomposition with bootstrap standard errors. The MCC office shares (80–82% for total and staff time, 61% for manager time) closely match the largest connected component estimates in Table 2, confirming that the dominance of office effects is not driven by the specific sample definition.

Table OA-6: Variance Decomposition: Multiple Connected Components

	Total Time	Manager Time	Staff Time
Sample			
Observations	94,904	94,904	94,904
Managers	1,755	1,755	1,755
Offices	736	736	736
Connected Components	7	7	7
Variances (Shrinkage-Corrected)			
$\widehat{\text{Var}}(\psi)$	0.55 (0.09)	0.25 (0.03)	0.69 (0.09)
$\widehat{\text{Var}}(\alpha)$	0.12 (0.08)	0.15 (0.03)	0.18 (0.07)
Correlation and Shares			
$\widehat{\text{Corr}}(\alpha, \psi)$	−0.17 (0.08)	−0.32 (0.07)	−0.17 (0.06)
$\widehat{\text{Corr}}(\alpha^{\text{M}}, \psi^{\text{S}})$	0.13 (0.06)		
Office Share	0.82 (0.04)	0.63 (0.04)	0.80 (0.04)
Manager Share	0.18 (0.04)	0.37 (0.04)	0.20 (0.04)

Notes: Each column uses a different outcome (see Table 2 notes for definitions). This table replicates the variance decomposition from Table 2 using the multiple connected components (MCC) sample employed in the heterogeneity analysis of Table 3. The MCC approach retains all connected components with at least 6 managers and 3 offices, rather than restricting to the largest connected component. This yields a larger sample (1,755 managers vs. 1,419; 736 offices vs. 559) while requiring component fixed effects to absorb cross-component level differences.

Comparison with main table: Office shares (80–82%) are closely aligned with the largest connected component estimates, confirming that the dominance of office effects is not driven by the specific sample definition. The cross-stage correlation $\widehat{\text{Corr}}(\alpha^{\text{M}}, \psi^{\text{S}}) = 0.13$ (SE = 0.06) is weakly positive, consistent with the main analysis.

Estimation: All specifications include week fixed effects and time-varying controls (vacancies, staff count, log applications, number of services). Covariance shrinkage uses 100 inner bootstrap iterations. **Standard errors:** Standard errors (in parentheses) computed via double bootstrap: 200 outer bootstrap iterations × 100 inner iterations, using match-level block resampling.

F.3.2 Alternative Bias-Correction Methods

Table OA-7 compares plug-in estimates with three alternative bias-correction approaches: the Monte Carlo correction of Andrews et al. (2008), covariance shrinkage (Best et al., 2023b), and split-sample estimation (Best et al., 2023b). All four methods use the largest connected component and confirm that office effects dominate.

Table OA-8 reports results using the KSS leave-out estimator (Kline et al., 2020), applied to the largest connected component of the pooled mobility network. This approach requires the "leave-out connected set"—offices that remain connected after removing any single match—which may reduce the usable sample. The leave-out connected set retains 37,451 observations (465 movers, 299 offices) from the full largest CC (71,636 observations). The KSS leave-out office shares (0.64–0.76 depending on the outcome) are broadly consistent with the shrinkage estimates in Table 2.

Table OA-7: Variance Decomposition: Alternative Bias Correction Methods

	Plug-in	Andrews†	Cov-Shrinkage	Split-Sample
Panel A: Total Processing Time				
Observations			71,636	
Managers			1,419	
Offices			559	
$\widehat{\text{Var}}(\psi)$	1.05	0.99	0.60	0.98
$\widehat{\text{Var}}(\alpha)$	0.46	0.39	0.13	0.38
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.60	-0.56	-0.07	-0.58
Office Share	0.70	0.72	0.83	0.72
Panel B: Manager Processing Time				
Observations			71,636	
Managers			1,419	
Offices			559	
$\widehat{\text{Var}}(\psi)$	0.59	0.54	0.28	0.56
$\widehat{\text{Var}}(\alpha)$	0.34	0.28	0.15	0.30
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.70	-0.68	-0.32	-0.68
Office Share	0.64	0.66	0.66	0.65
Panel C: Staff Processing Time				
Observations			71,636	
Managers			1,419	
Offices			559	
$\widehat{\text{Var}}(\psi)$	1.24	1.14	0.77	1.12
$\widehat{\text{Var}}(\alpha)$	0.75	0.64	0.20	0.62
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.58	-0.54	-0.13	-0.55
Office Share	0.62	0.64	0.79	0.64

Notes: Each column uses a different outcome (see Table 2 notes for definitions). This table compares four approaches to address limited mobility bias in two-way fixed effects variance decomposition. All estimates use the largest connected component of the manager-office mobility network. **Plug-in:** Uncorrected variance components from the AKM regression. **Andrews:** Andrews et al. (2008) Monte Carlo bias correction using 10,000 Rademacher weight simulations. **Cov-Shrinkage:** Covariance shrinkage following Best et al. (2023), with 100 inner bootstrap iterations. **Split-Sample:** Split-sample covariance estimator with 50 random match-level splits. All regressions include week fixed effects, total staff count, subordinate staff and manager vacancy indicators, log applications received, and number of services handled. Office Share $= \widehat{\text{Var}}(\psi)/[\widehat{\text{Var}}(\alpha) + \widehat{\text{Var}}(\psi)]$. †The Andrews estimator requires $(X'X)^{-1}$, which is sensitive to near-collinearity. Staffing controls (clerk and approving-authority vacancy indicators, total staff count) are nearly spanned by the fixed effect structure, making $X'X$ near-singular. We therefore exclude staffing controls from the Andrews specification while retaining workload controls (log applications, number of services). The remaining estimators use point estimates rather than the inverse and are unaffected.

Table OA-8: Variance Decomposition: KSS Leave-One-Out Estimator

	Total Time	Manager Time	Staff Time
Leave-Out Connected Set			
Observations	37,451	37,451	37,451
Movers	465	465	465
Offices	299	299	299
Leave-Observation-Out Estimates			
$\text{Var}(y)$	0.61	0.48	0.85
$\widehat{\text{Var}}(\psi)$	0.42	0.15	0.68
$\widehat{\text{Var}}(\alpha)$	0.14	0.08	0.21
$\widehat{\text{Corr}}(\alpha, \psi)$	-0.51	-0.08	-0.59
Office Share	0.76	0.64	0.76

Notes: Each column uses a different outcome (see Table 2 notes for definitions). This table reports leave-observation-out (LOO) variance decomposition estimates following Kline et al. (2020), using the largest connected component of the pooled manager-office mobility network. The LOO estimator removes each observation in turn and provides heteroskedasticity-robust bias-corrected variance components.

Leave-out connected set: The LOO estimator requires a "leave-out connected set"—the subset of observations where removing any single observation does not disconnect the mobility graph. This may reduce the usable sample relative to the full largest connected component reported in Table 2.

Controls: All regressions include week fixed effects, total staff count, subordinate staff and manager vacancy indicators, log applications received, and number of services handled. Controls and time effects are partialled out prior to KSS estimation. Office Share $= \widehat{\text{Var}}(\psi)/[\widehat{\text{Var}}(\alpha) + \widehat{\text{Var}}(\psi)]$.

G Assessing Assortative Matching

G.1 AKM Correlations with Differing Outcomes

In addition to the usual AKM estimate of the correlation between manager and office effects we examine matching using two complementary approaches. First, we estimate manager fixed effects (using the AKM model) with manager time taken as the dependent variable and office fixed effects (using the AKM model) with staff time taken as the dependent variable. Because these fixed effects come from different regressions, any correlation between them is not mechanical—unlike the baseline AKM correlation where both effects are identified from the same regression.

We correlate these cross-outcome fixed effects for manager-office pairs in our data. The pooled cross-stage correlation is $\widehat{\text{Corr}}(\alpha^M, \psi^S) = 0.17$ (SE $= 0.07$), as reported in Table 2. This is weakly positive and substantially smaller in magnitude than the within-column correlations ($\rho = -0.59$ plug-in, $\rho = -0.07$ after shrinkage for total time), suggesting that the negative within-regression correlations are largely mechanical artifacts of limited mobility bias.

G.2 Office Productivity by Changes in Manager Productivity

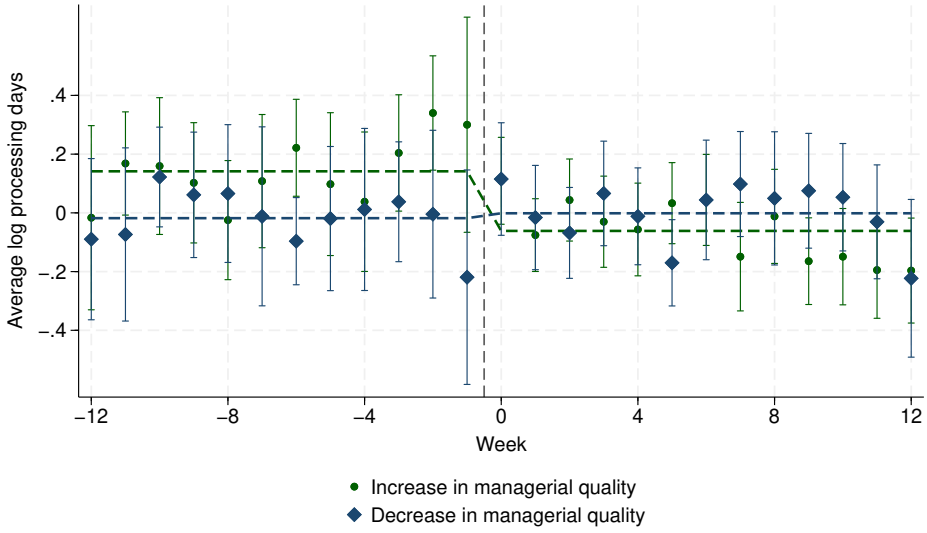
Second, we examine changes in overall office productivity when an incoming manager has a substantially different productivity level from the outgoing manager in a transfer. We calculate each office's change in manager quality (as measured by manager fixed effects derived from a regression of (log) processing days on manager,

week, service and office indicators). To simplify the exposition, we then group these office-level changes in manager quality into quartiles and plot the office-level productivity for the bottom and top quartiles. Figure OA-8 shows changes in overall office-level productivity when the office experiences either an increase (green dots) or decrease (blue diamonds) in manager quality.[37] Offices with an increase in incoming manager quality (i.e. when a higher productivity manager transfers in) show slight improvements in both processing times (a decrease of 0.2 log points) in the weeks following the transfer. Conversely, offices that receive a lower quality incoming manager exhibit marginally worse performance. The results are somewhat noisy, however, and any conclusions are tentative.

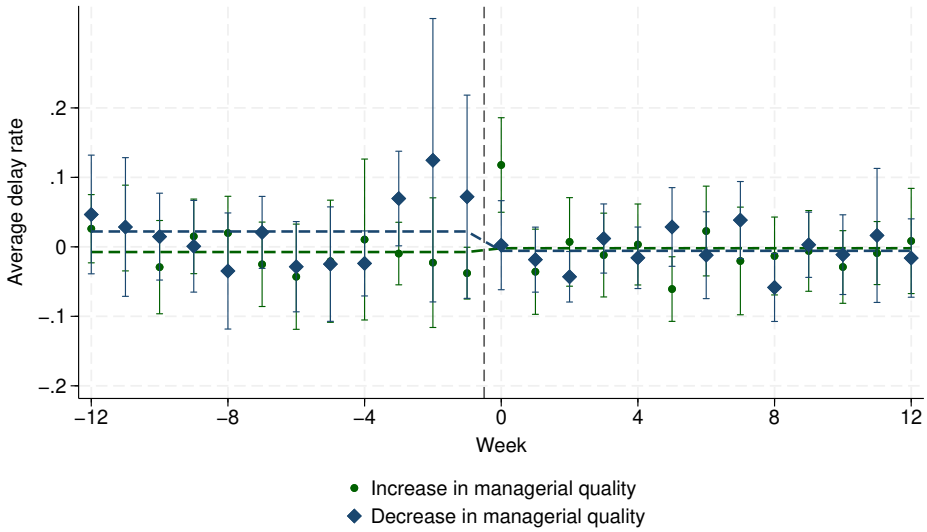
[37] Since the plotted residuals retain the manager fixed effect while the treatment groups are defined by changes in the same fixed effect, part of the observed jump is mechanical. This figure is therefore best viewed as descriptive rather than causal. The formal event study (Figure 2) avoids this circularity: the outcome is raw log productivity (not a residual), the office productivity gap δ_i is computed using leave-one-out means that exclude the focal manager's own observations, and identification comes from the interaction $\delta_i \times \mathbf{1}[r \geq 0]$ rather than from grouping by estimated fixed effects.

Figure OA-8: Office productivity by change in manager quality

(a) Log Process Days



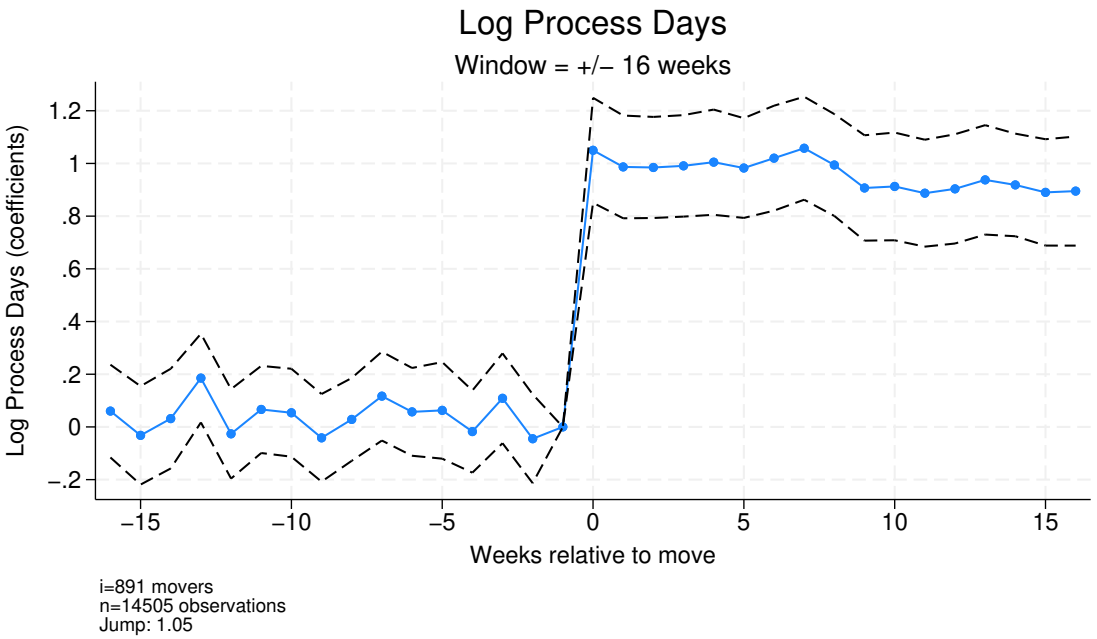
(b) Delay Rate



Note: This figure shows what happens to office-level productivity when it experiences either an improvement or a decline in the quality of the manager in charge. We first de-trend productivity by removing time and office fixed effects. We then calculate each office's change in managerial quality using manager fixed effects of the incoming and outgoing managers. We group these office-level changes in managerial quality into quartiles and plot the bottom and top quartiles.

Figure OA-9: Event study, 16 week window

(a) Log Process Days



(b) Log Delay Rate

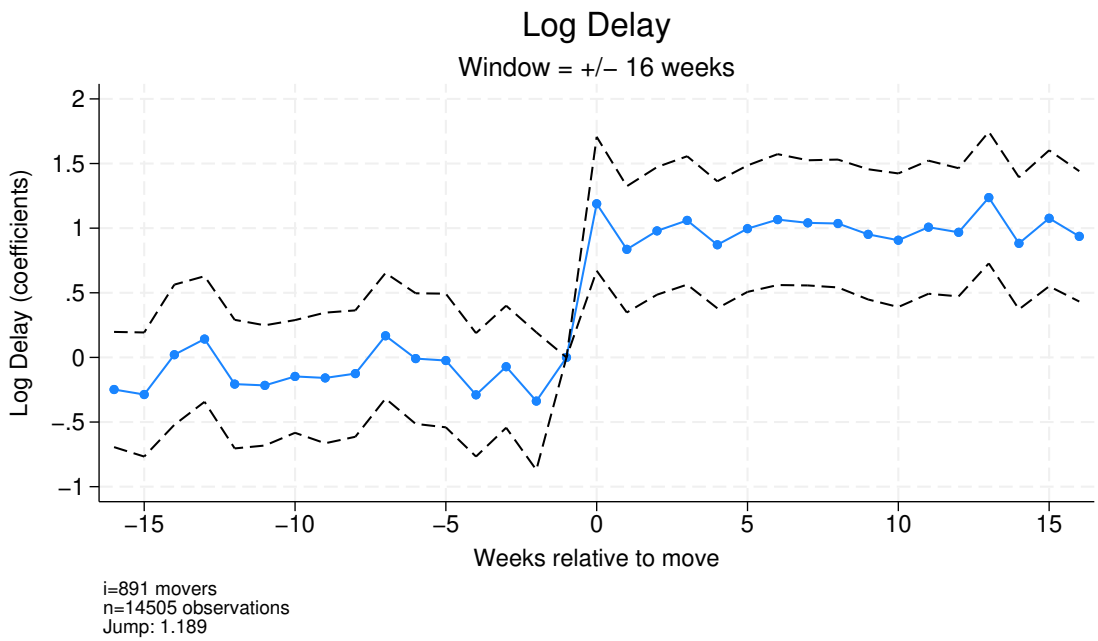
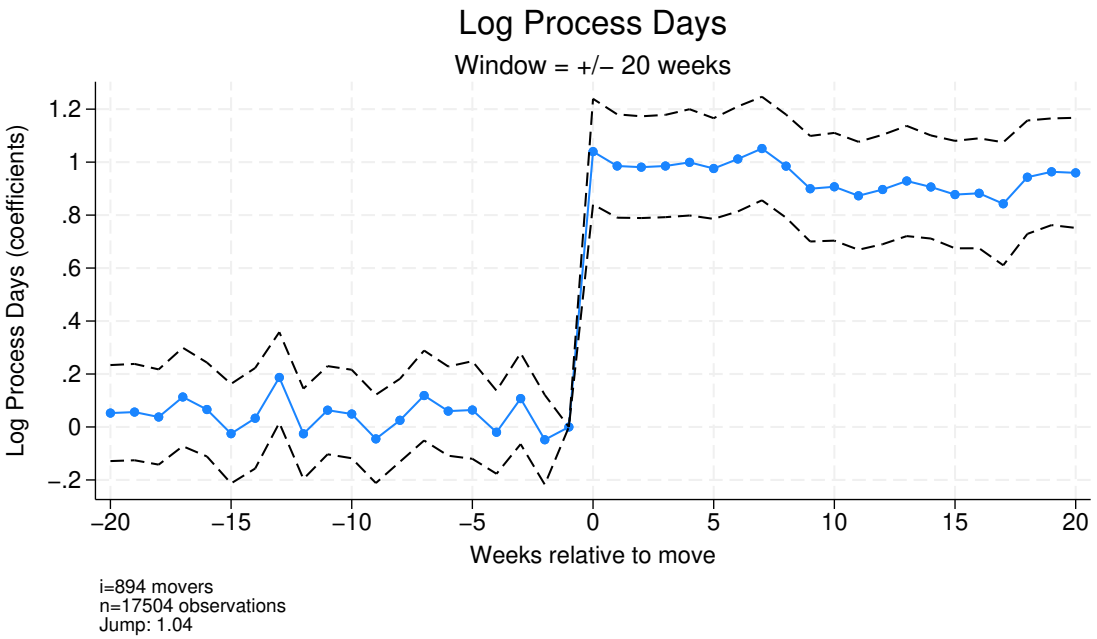


Figure OA-10: Event study, 20 week window

(a) Log Process Days



(b) Log Delay Rate

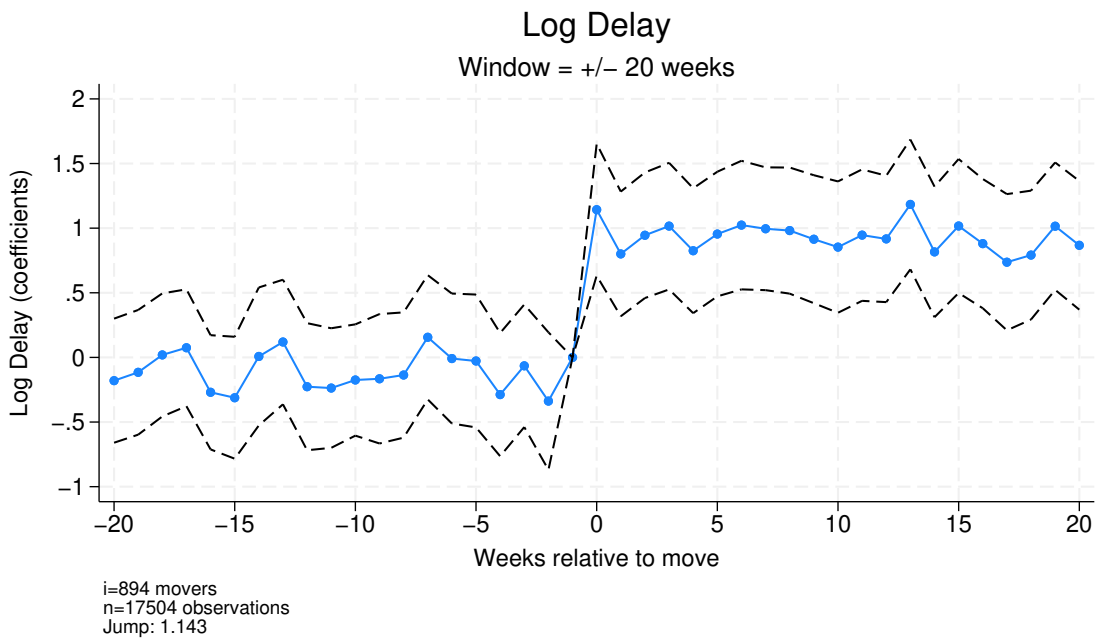


Figure OA-11: Event study, 24 week window

(a) Log Process Days



(b) Log Delay Rate

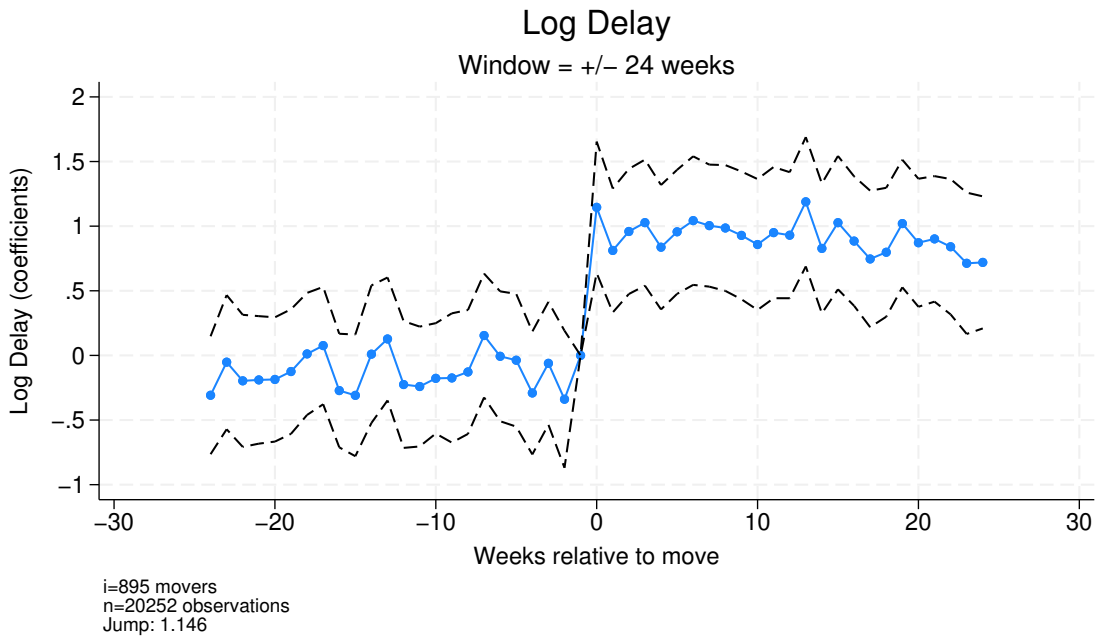
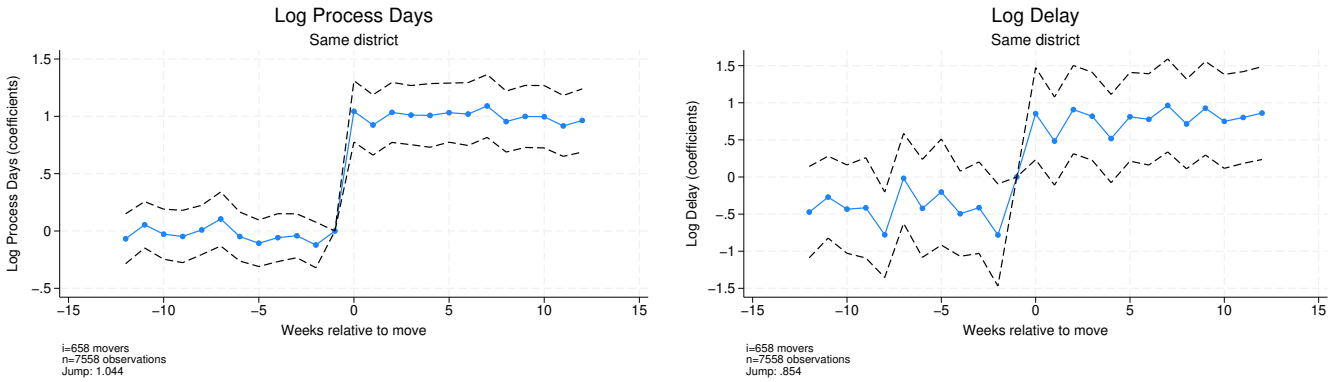
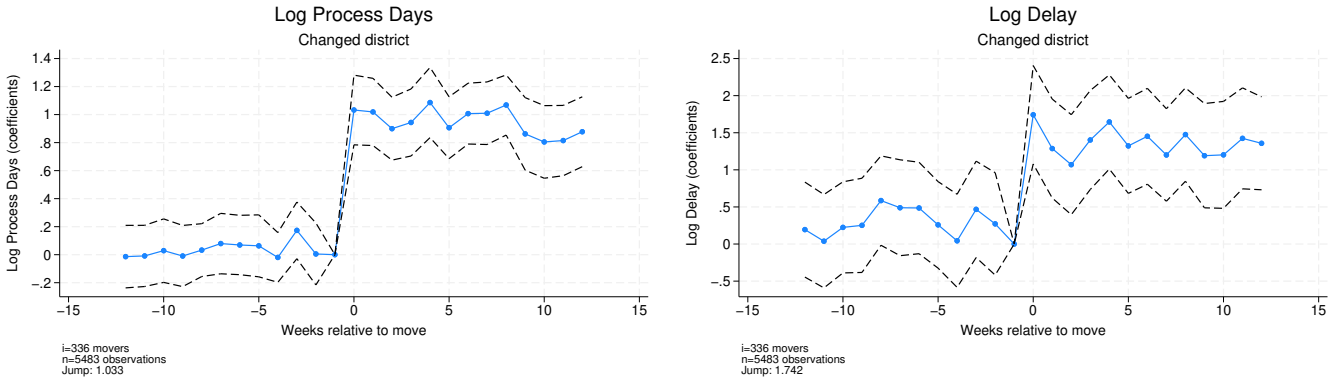


Figure OA-12: Location-based share of overall application processing time, by type of transfer

(a) Do not change districts



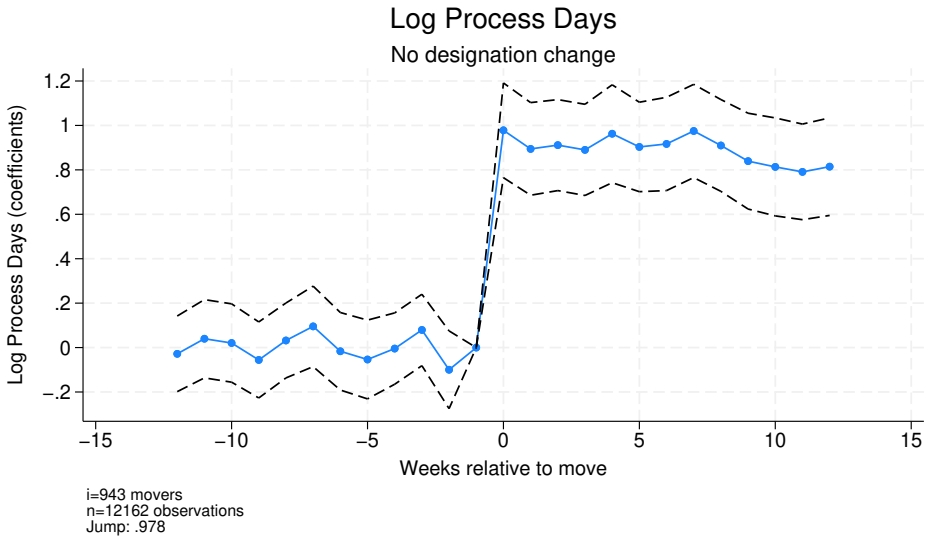
(b) Change districts



Note: Graphs plot the event study results for separate estimation on movers that (a) do not change districts and (b) change districts.

Figure OA-13: Overall location-based share, those that do not change designations

(a) Log Process Days

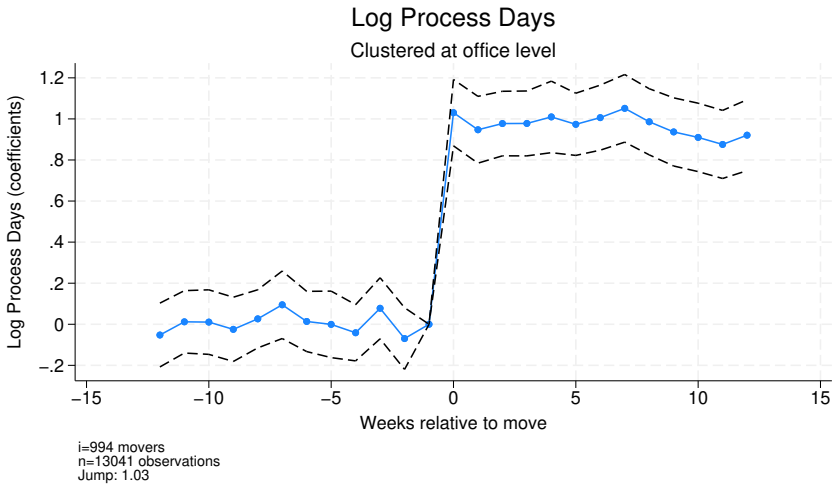


(b) Log Delay Rate



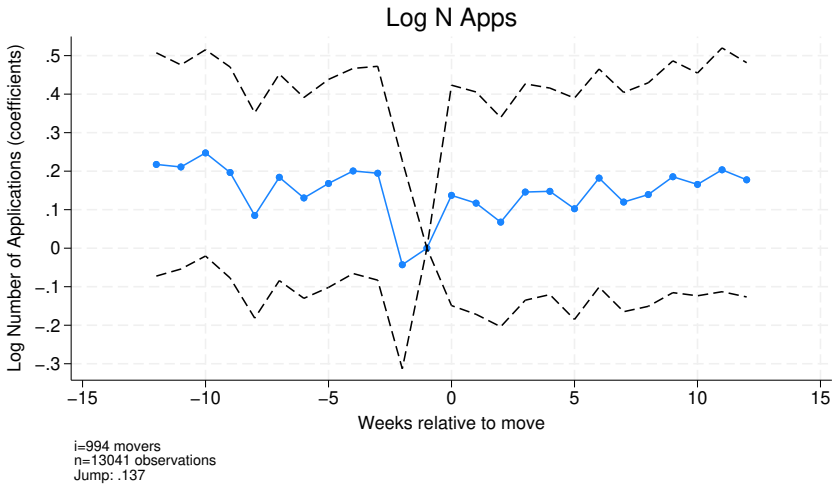
Note: Graphs plot the event study results for the subsample of movers that do not change designations during a transfer.

Figure OA-14: Clustering errors at the office level



Note: Graph shows the event study results with errors clustered at the office level. There are 994 movers in our sample (2021–2025).

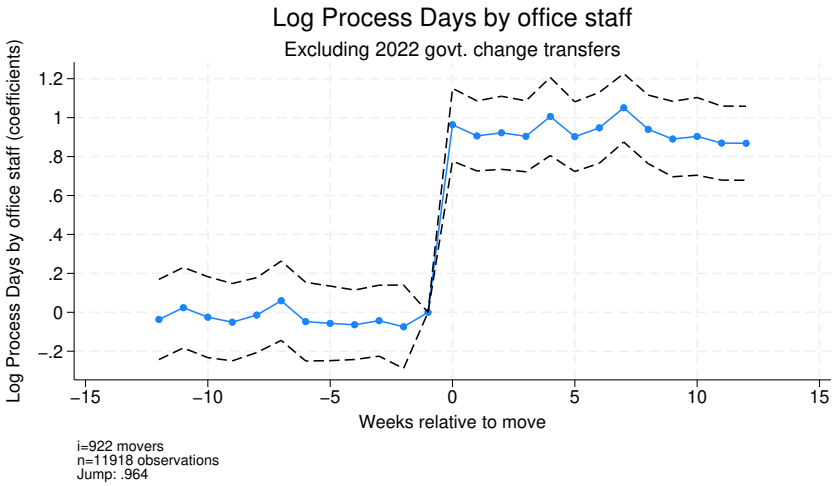
Figure OA-15: Testing manager effect on the number of applications

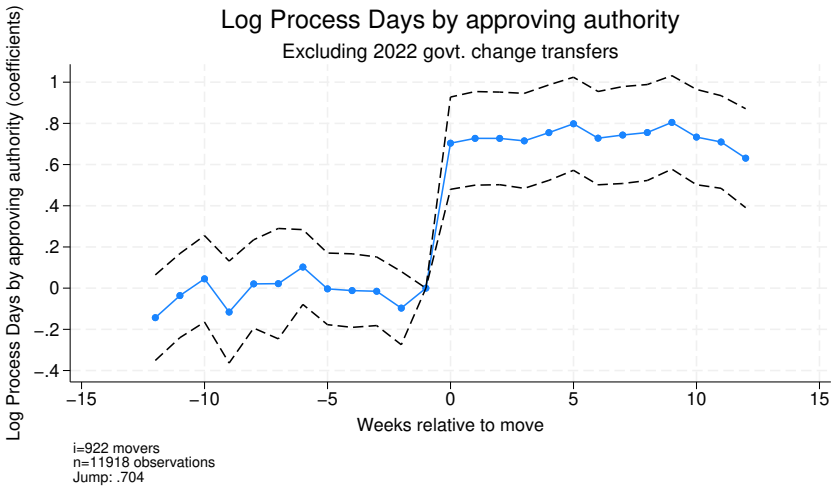


Note: Graph shows the event study results using Log Number of Applications handled by that office as the outcome. There are 994 movers in our sample (2021–2025).

Figure OA-16: Dropping spike of transfers in summer 2022

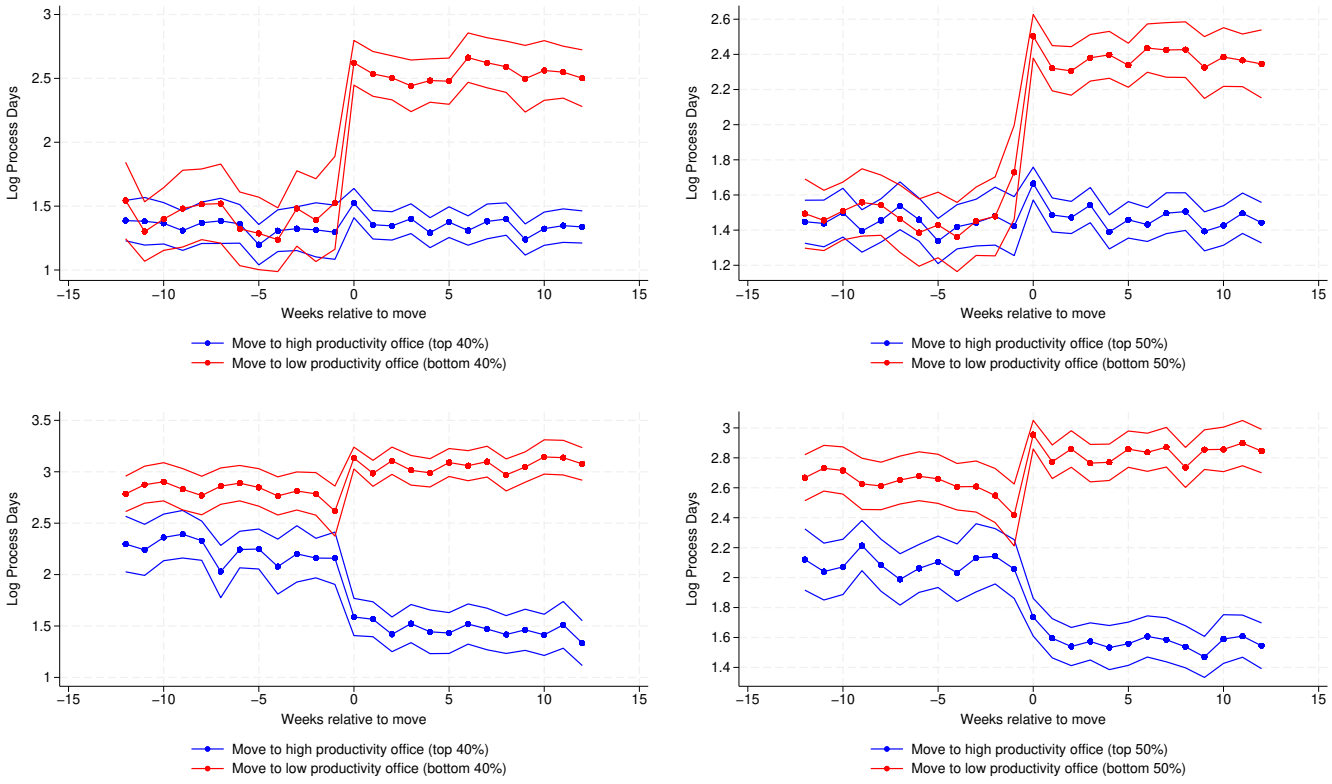






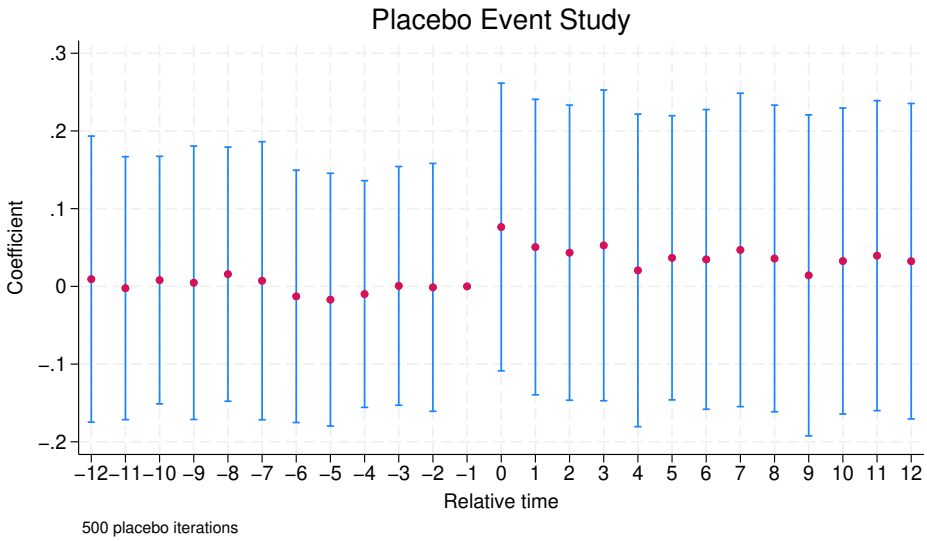
Note: Graphs show the event study results on the sample excluding transfers that occurred during the spikes seen in summer 2022. Sample period: 2021–2025.

Figure OA-17: Changes in manager productivity by origin and destination types



Note: These graphs plot weekly average manager productivity over time for managers that move across offices of different productivity levels. We test two additional definitions of high and low productive offices (top/bottom 40% and top/bottom 50%) using $log(\bar{y})$ for the entire sample period. Figure shows that when a manager moves from a high productive office to another high productive office (blue lines in top graphs), her average productivity remains unchanged. When she moves from a high productive office to a low productive office, however, (red lines in top graphs), her productivity immediately worsens. We show the parallel case for those managers that move from low productivity offices to low/high productive offices in the bottom graphs.

Figure OA-18: Placebo Test



Note: As a placebo test, we examine the event study results from randomly assigning delta values to our sample of movers. We first calculate all possible delta values for within-department moves using $\bar{y}_{o(i)}$ and $\bar{y}_{d(i)}$ observed in the data. Then we randomly select deltas (without replacement) from this distribution and assign them to the 994 movers in our existing sample. Running the event study estimation on this placebo delta group, we obtain and store the coefficients. We repeat this exercise 500 times. Then, we calculate average coefficients and their standard deviations for each relative time bin. Lastly, we calculate 95% confidence intervals using the means and standard deviations. The plot shows the resulting event study plot with coefficients of zero in both pre and post transfer periods.

H Counterfactual Analysis: Framework and Assumptions

In this section we evaluate the potential gains from alternative manager allocation schemes (Fenizia, 2022). We begin by outlining the framework and assumptions required for these counterfactual exercises to be interpretable.

H.1 Framework

Consider a static allocation problem in which a social planner seeks to minimize aggregate processing time across offices. Let P_{ij} denote average processing time at office j when staffed by manager i. We assume:

$$\ln P_{ij} = \alpha_i + \psi_j + \epsilon_{ij} \tag{12}$$

where α_i is the manager fixed effect and ψ_j is the office fixed effect.

Cross-Outcome Identification. A key distinction from previous counterfactual analyses is that *our manager and office fixed effects come from different regressions*. In Fenizia (2022) and related work, both α_i and ψ_j are estimated from a single AKM regression with the same dependent variable (e.g., total processing time or productivity). This creates potential for mechanical correlation: any measurement error or misspecification that inflates manager effects will mechanically deflate office effects (and vice versa), since both are identified from variation in the same outcome.

We instead estimate manager fixed effects from a regression with *manager time* as the dependent variable,

and office fixed effects from a separate regression with *staff time* as the dependent variable. This "cross-outcome" approach provides conceptually cleaner measures: the manager FE are more likely to capture the manager's inherent processing speed (how quickly they personally handle cases), while the office FE are more likely to capture the office environment (how the office affects staff processing time, reflecting culture, resources, or norms). Because these fixed effects are estimated from independent regressions on different outcomes, we side-step the issue of mechanical correlation that arises in single outcome approaches. We compare aggregate processing time under the current assignment versus random assignment (averaged over 1,000 simulations).

The counterfactual exercises require several assumptions that we discuss below. In particular, we need to assume the **portability of fixed effects**: a manager's fixed effect α_i—estimated from their historical performance—would apply unchanged if they were reassigned to a different office. This assumption would be violated for instance if manager productivity depends on tenure at a specific office, managers invest in office-specific human capital that is lost upon transfer or if motivation or effort depends on assignment preferences. We also assume **no general equilibrium effects** so that reassignments do not change office fixed effects ψ_j (office "culture" is exogenous to staffing), other managers' productivity (no peer effects or spillovers) and there are no strategic responses to reassignment. Finally, the counterfactual uses fixed effects estimated from historical data to predict outcomes under alternative assignments. This assumes manager and office productivity are stable characteristics rather than time-varying. Given these assumptions, we provide a simple counterfactual exercise that compares the empirical assignments with random assignment where we randomly reassign managers to offices (averaged over 1,000 simulations).

The comparison between random assignment and the current system is informative about the efficiency of existing allocation mechanisms and depends crucially on the correlation between office and manager effects (denoted by $\rho_{\alpha\psi}$). Table OA-9 reports the cross-outcome correlations and counterfactual separately for the Rural Development and Revenue departments. Both departments exhibit negative cross-outcome correlations, indicating that faster managers (lower α) tend to be assigned to slower offices (higher ψ)—a *compensating* assignment pattern.

Rural Development ($\rho_{\alpha\psi} = -0.13$). The modest negative correlation suggests that better managers are placed at somewhat slower offices. Under random assignment, processing time decreases by 5%, indicating that the current compensating pattern is not large enough to be optimal: random reallocation still yields small gains by better matching manager speeds to office needs.

Revenue ($\rho_{\alpha\psi} = -0.14$). The negative correlation is similar in magnitude to Rural Development. Random assignment reduces processing time by 22%. The larger gain in Revenue likely reflects a wider dispersion of manager and office effects in this department, so that even modest negative sorting leaves substantial room for improvement through reallocation.

Table OA-9: Random vs. Current Assignment: Cross-Outcome Analysis

	Rural Development	Revenue
Sample		
Managers	372	789
Offices	151	322
Cross-Outcome Correlation		
$\rho(\alpha, \psi)$	-0.13	-0.14
Random vs. Current Assignment		
Processing Time Change (%)	-5	-22

Notes: This table compares random assignment to the current assignment using cross-outcome fixed effects (manager FE from manager time, office FE from staff time). The cross-outcome correlation $\rho(\alpha, \psi)$ measures sorting without mechanical bias. Rural Development exhibits negative sorting ($\rho < 0$: faster managers at slower offices), so random assignment changes outcomes by -5%. Revenue exhibits negative sorting ($\rho < 0$), so random assignment changes outcomes by -22%. Random assignment averaged over 1,000 simulations. Extended sample (2021–2025).

Interpretation. Both departments exhibit weak negative sorting, and random assignment reduces processing time in each case (by 5–22%). The gains from reallocation are modest, consistent with our main finding that office effects dominate and manager reassignment has limited scope to improve outcomes. The larger magnitude in Revenue reflects greater dispersion of fixed effects rather than a qualitatively different sorting pattern.

Given the stringency of the assumptions required, the counterfactual results are best interpreted as indicative of sorting patterns rather than precise predictions of policy impacts.

I RCT: Experimental Evidence on Deadline Reductions

I.1 RCT Design and Sample

As mentioned in the main text, the Punjab government randomized service delivery deadlines across 30 services in December 2022. Treatment services (15) received new deadlines set at the 60th–75th percentile of pre-treatment processing time; control services (15) maintained original deadlines. The average deadline reduction was 2.9 days (approximately 16%), with old deadlines ranging from 9–32 days (mean: 17.5) and new deadlines ranging from 6–31 days (mean: 14.6).

Table OA-10 describes the sample. Services were stratified by online/offline status and department prior to randomization. Table OA-11 presents balance tests confirming no systematic pre-treatment differences between treatment and control services.

Table OA-10: RCT Sample Description

	Full Sample
Services (Treatment)	15
Services (Control)	15
Managers	1,213
Offices	832
Observations	154,189

Table OA-11: Pre-Treatment Balance

	(1) Proc. Days	(2) Manager Days	(3) Staff Days	(4) Delay 1	(5) Delay 2	(6) N Apps
Treatment	0.017	0.309	−0.304	0.010	0.043	−10.01
	(1.528)	(0.548)	(0.992)	(0.034)	(0.036)	(11.00)
Observations	73,666	73,666	73,666	73,666	73,666	73,666
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.026	0.009	0.022	0.023	0.044	0.021

Note: Each column reports the coefficient on treatment from a regression of the outcome on a treatment indicator, controlling for week fixed effects. Sample restricted to observations before December 15, 2022. Standard errors clustered at the service level in parentheses.

I.2 Main Treatment Effect Event Study

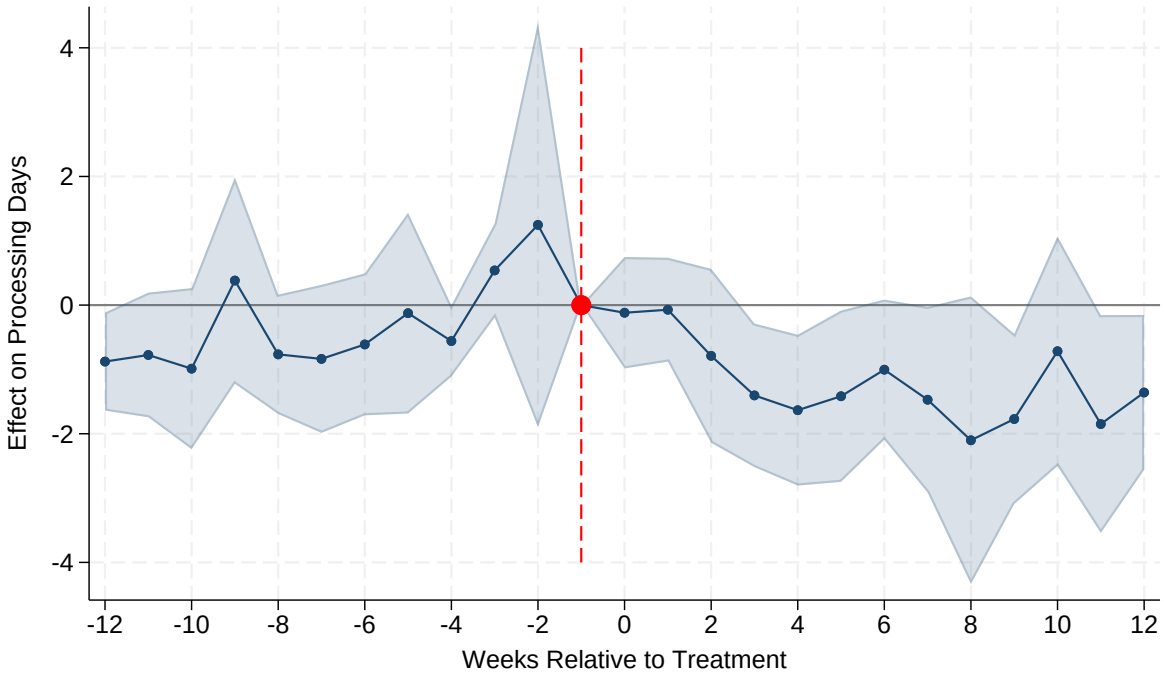
We complement the difference-in-differences specification in Equation (3) with an event study that traces out treatment dynamics:

$$y_{isjt} = \psi_{sj} + \alpha_{si} + \tau_t + \sum_{r \neq -1} \beta_r \cdot \mathbf{1}\{t - t^* = r\} \cdot \text{Treat}_s + \epsilon_{isjt} \tag{13}$$

where t^* denotes the treatment date (December 15, 2022) and the coefficients $\{\beta_r\}$ trace out the treatment effect by week relative to implementation. We normalize $\beta_{-1} = 0$.

Figure OA-19 displays the estimates. The pre-treatment coefficients are centered around zero, consistent with parallel trends. After treatment, effects emerge and grow over time. The pooled DiD coefficient for Manager Time is not individually significant ($p = 0.139$) as it averages across all post-treatment weeks, including the initial weeks when effects are still ramping up; the event study reveals coefficients are statistically significant in 11 of 13 post-treatment weeks ($p < 0.05$ for weeks 1–12 except weeks 0 and 3), with effects growing from -0.46 days immediately after treatment to -0.91 days by week 12.

Figure OA-19: RCT Event Study: Effect of Deadline Reductions on Processing Time



Note: Figure shows the coefficients $\{\hat{\beta}_r\}$ from Equation (13) by week relative to treatment (December 15, 2022). Treatment services received new deadlines set at the 60th–75th percentile of pre-treatment processing time (average reduction of 2.9 days). The outcome is total processing days. Reference period is $t = -1$ (week before treatment). Shaded area shows 95% pointwise confidence intervals. Specification includes service-by-office, service-by-manager and week fixed effects. Standard errors two-way clustered by service and week.

I.3 Spillover Analysis

A potential concern with our RCT design is spillovers across treatment and control services. Since randomization occurred at the service level, managers may process applications for both treatment and control services.

In our sample, 74% of managers (897 out of 1,213) process applications for both treatment and control services during the sample period. The remaining managers work exclusively on treatment services (289 managers, 24%) or exclusively on control services (27 managers, 2%).

Table OA-12 compares treatment effects across manager subsamples. Columns (1)–(2) include managers who work on both service types; columns (3)–(4) restrict to managers who work exclusively on one service type. The single-type manager estimates are substantially larger in magnitude, consistent with spillover attenuation in the main estimates.

Table OA-12: Treatment Effects by Manager Type: Evidence for Spillovers

	With Manager FE		Without Manager FE	
	(1) Full Sample	(2) Both Types	(3) Single Type (Base)	(4) Single Type (+Manager FE)
Panel A: Manager Processing Time				
Treatment × Post	−0.176*	−0.176*	−1.683	−1.668
	(0.092)	(0.098)	(1.208)	(1.360)
Observations	153,569	144,833	8,744	8,733
Panel B: Total Processing Time				
Treatment × Post	−0.640*	−0.684*	−2.310	−4.326
	(0.350)	(0.370)	(1.768)	(2.651)
Observations	153,569	144,833	8,744	8,733

Note: The dependent variable is a measure of processing time by manager i handling service s in office j in week t. "Both Types" restricts to managers who process applications for both treatment and control services. "Single Type" restricts to managers who work exclusively on one service type. Columns (1)–(2) include service×office, service×manager, and week fixed effects. Columns (3)–(4) include service×office and week fixed effects; column (4) adds manager fixed effects. Standard errors two-way clustered by service and week in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I.4 Robustness

Build-up of Fixed Effects. Table OA-13 shows how treatment effect estimates change as we progressively add fixed effects. Column (1) includes only week and service fixed effects. Column (2) adds office fixed effects. Column (3) adds manager fixed effects. Column (4) uses our preferred specification with service×office and service×manager interaction fixed effects, which absorbs heterogeneous baseline processing times by service within each office and manager.

We do not include manager×week or office×week fixed effects because these would limit identification to unit-week cells containing both treatment and control observations in the same week. While 74% of managers process both treatment and control services during the sample period, approximately one-third of manager-week cells contain only treatment or only control applications. The service×unit specification exploits all cross-service variation within managers and offices, not just within-week variation.

Table OA-13: RCT Results: Build-up of Fixed Effects

	(1)	(2)	(3)	(4)	(5)
	(1)	(2)	(3)	(4)	(5)
Treatment × Post	-0.640	-0.607	-0.636	-0.625	-0.632
	(0.903)	(0.946)	(0.592)	(0.619)	(0.841)
Observations	414598	414583	395522	395273	394735
Week FE	Yes	Yes	Yes	Yes	Yes
Service FE	Yes	Yes	Yes	No	No
Office FE	No	Yes	Yes	No	No
Manager FE	No	No	Yes	Yes	No
Service × Office FE	No	No	No	Yes	No
Service × Manager FE	No	No	No	No	Yes
R-squared	0.130	0.224	0.252	0.306	0.280

Note: Dependent variable is total processing days. Treatment × Post is the difference-in-differences estimate. Column (3) is the preferred specification (matching Table 3). Column (4) uses service×office interaction fixed effects with additive manager fixed effects; Column (5) uses service×manager interaction fixed effects. Interaction fixed effects subsume the corresponding additive components. Sample sizes decline slightly across columns as `reghdfe` drops singleton observations. Standard errors two-way clustered by service and week in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: `fe_buildup.do`.

Alternative Clustering. Table OA-14 shows robustness of the main result to different standard error clustering choices.

Table OA-14: Robustness to Alternative Clustering

	Total Time	Manager Time	Staff Time
Treatment × Post	−0.609	−0.192	−0.399
SE: Service + Week (baseline)	(0.818)	(0.126)	(0.732)
SE: Service only	(0.822)	(0.127)	(0.736)
SE: Office only	(0.180)***	(0.043)***	(0.167)**
Observations	394,626	394,626	394,626

Note: All specifications use the preferred regression from Table 3 with week, service×office, and service×manager fixed effects. Point estimates are identical across rows; only standard errors differ. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: `rct_alt_clustering.do`.

Log Transformations. Following Chen and Roth (2024), Table OA-15 examines robustness to different transformations of the outcome variable.

Table OA-15: Robustness to Log Transformations

	(1) log(y+1)	(2) log(y+0.1)	(3) log(y)	(4) asinh	(5) Levels
treat_post	-0.0450	-0.0566	-0.0585	-0.0541	-0.637
	(0.0905)	(0.110)	(0.112)	(0.108)	(0.905)
Observations	413984	413510	413143	414657	414657
Transformation	log(y+1)	log(y+0.1)	log(y), drop zeros	asinh(y)	Levels
R-squared	0.215	0.207	0.207	0.209	0.130

Dependent variable: Processing days (total).

All specifications include service and week fixed effects.

Standard errors clustered at service level.

Column 3 drops observations with zero processing days.

Note: Column (1) uses levels. Column (2) uses $\log(1+y)$. Column (3) uses $\text{asinh}(y)$. Column (4) uses $\log(y)$ (dropping zeros). All specifications include week, service×office, and service×manager fixed effects. Standard errors two-way clustered by service and week in parentheses.

Other Outcomes. Table OA-16 presents treatment effects on delay rates (whether processing exceeded the mandated deadline) and number of applications processed.

Table OA-16: Treatment Effects on Other Outcomes

	(1) Delay 1	(2) Delay 2	(3) N Apps	(4) Log Proc. Days
Treatment × Post	0.027	-0.016	2.54	-0.061^{*}
	(0.017)	(0.017)	(4.64)	(0.034)
Observations	157,824	157,824	157,824	157,824
Dep. Var. Mean	0.12	0.13	26.14	1.83
R^2	0.347	0.362	0.720	0.603

Note: Delay 1 = exceeded original deadline. Delay 2 = exceeded new deadline (treatment) or original deadline (control). N Apps = applications processed. Log Proc. Days = $\log(1 + \text{processing days})$. All specifications include week, service×office, and service×manager fixed effects. Standard errors two-way clustered by service and week in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I.5 Constructing Manager and Office Productivity Measures

This section describes our procedure for estimating manager and office productivity measures used in the heterogeneity analysis of Table 3. We follow the methodology of Best et al. (2023b), which extends standard AKM fixed effects estimation to address two challenges common in administrative data: (i) multiple disconnected mobility networks and (ii) estimation error bias from sparse mobility.

Stage-Specific Productivity Measures. We construct separate productivity measures for managers and offices, exploiting the decomposition of processing time into manager and staff components:

- **Manager productivity**: AKM manager fixed effects estimated using *manager processing time* (time attributable to the approving authority) as the outcome. This captures how quickly managers make approval decisions.
- **Office productivity**: AKM office fixed effects estimated using *staff processing time* as the outcome. This captures the efficiency of office-level administrative processes.

This stage-specific approach ensures that manager productivity reflects managerial decision-making speed while office productivity reflects the productivity of subordinate staff—the key distinction motivating our analysis.

Sample and Estimation Period. To avoid contamination from treatment effects, we estimate productivity measures using only **pre-RCT data** from 2021 through December 14, 2022 (before the intervention). This estimation period increases manager mobility and yields a substantially larger connected set compared to using only 2022 data.

Multiple Connected Components (MCC). Standard AKM estimation restricts the sample to the *largest* connected component of the worker-firm mobility network, discarding observations in smaller components (Card et al., 2013). This is because fixed effects are identified only up to a component-specific constant: a manager with FE $= 0.5$ in component A cannot be directly compared to a manager with FE $= 0.5$ in component B without additional assumptions. In labor market applications, the largest component typically contains 90%+ of observations, so this restriction has minimal cost.

Our setting differs: even with this estimation period, the largest connected component is a relatively small fraction of the full sample. Following Best et al. (2023b), we therefore implement the MCC approach, retaining **all connected components** meeting minimum size thresholds (at least 6 managers and 3 offices). We estimate fixed effects separately within each component, including component-level intercepts to absorb level differences across disconnected networks. For the heterogeneity analysis, we then standardize (mean zero, SD one) within each connected component.

A concern with this approach is that cross-component FE comparisons may not be valid if components have different baseline productivity levels. To address this, we verified that within-component standardization (comparing each manager to others in the same component) yields nearly identical results, with correlations of 0.98 (managers) and 1.00 (offices) between within-component and pooled measures.

Covariance Shrinkage. With sparse mobility, plug-in AKM estimates suffer from substantial estimation error bias: both manager and office variance shares are inflated, and spurious negative correlation between effects emerges. We implement the two-dimensional covariance shrinkage estimator of Best et al. (2023b), which (a) uses a residual bootstrap (100 iterations) to estimate the covariance matrix of estimation errors for both manager and office effects and (b) jointly shrinks both sets of effects toward zero, accounting for correlated estimation errors. This correction is substantial: in the largest connected component, plug-in estimates imply a correlation of -0.59 between manager and office effects for total time, which shrinkage reduces to -0.06—indicating that the apparent negative correlation was largely spurious.

Normalization. For interpretability in the heterogeneity analysis, we normalize productivity measures within each connected component to have mean zero and standard deviation one. Higher values indicate *higher* productivity (faster processing). Thus, a manager with productivity $+1$ processes applications one standard deviation faster than the average manager in the same connected component.

Coverage in RCT Sample. Matching these pre-period productivity measures to the RCT sample yields observations with *both* manager and office measures available comprising approximately 54% of the RCT sample (221,718 out of 414,598). Importantly, the matched sample maintains balance across treatment and control conditions and spans the full range of productivity variation, supporting the validity of the heterogeneity analysis.

Variance Decomposition for the MCC Sample. Table OA-6 reports the full variance decomposition for the MCC sample used in the heterogeneity analysis. The MCC sample retains more managers and offices than the largest connected component (1,755 vs. 1,419 managers; 736 vs. 559 offices). The key finding from the main analysis is robust: office effects dominate, with Office Share of 80–82% for total and staff processing time and 61% for manager time—closely matching the largest connected component estimates. The cross-stage correlation between manager effects (from the Manager Time regression) and office effects (from the Staff Time regression) is 0.12 (SE = 0.05), weakly positive and consistent with the modest positive assortative matching found in the main analysis.

I.6 AKM vs Non-AKM Sample Comparison

The heterogeneity analysis in Panel C of Table 3 requires estimating manager and office productivity using AKM fixed effects. This methodology requires connected components: we can only identify productivity for managers and offices that are linked through manager transfers. Offices without such links—isolated offices—cannot be included in the heterogeneity analysis.

Table OA-17 compares the AKM sample (offices in connected components) with the non-AKM sample (isolated offices). Using the pre-RCT period (2021 through December 2022) and multiple connected components, the AKM sample comprises 221,987 observations (54% of 414,598). After including the $M_i \times O_j$ control, the Panel C regression sample is 221,718. The treatment effects are qualitatively similar for both samples (although e.g. total processing time is somewhat larger in the AKM sample than in the full sample).

Table OA-17: AKM vs Non-AKM Sample Comparison

	Full Sample	AKM Sample	Non-AKM Sample
Panel A: Sample Characteristics			
Observations	414,598	221,987	192,611
Unique managers	1,999	855	1,347
Unique offices	859	466	812
Panel B: Pre-Treatment Processing Time (days)			
Total	11.15	10.44	12.56
Manager	2.09	1.97	2.32
Staff	8.51	7.90	9.70
Panel C: Treatment Effects (Treatment × Post)			
Total	−0.613.	−0.891.	0.172.
	(0.829)	(0.727)	(0.710)
Manager	−0.192.	−0.199.	−0.126.
	(0.126)	(0.129)	(0.156)
Staff	−0.399.	−0.673.	0.332.
	(0.732)	(0.634)	(0.654)

Notes: This table compares characteristics and treatment effects between the AKM sample (offices in connected components with productivity measures) and the non-AKM sample (isolated offices without productivity measures), using pre-RCT AKM fixed effects (2021–Dec 2022). Panel A shows sample sizes. Panel B shows pre-treatment mean processing times. Panel C shows treatment effects from the main specification with week, service×office, and service×manager fixed effects; standard errors two-way clustered by service and week in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I.7 Pre-Treatment Slack by Office Productivity

A key finding in Section 5.3 is that treatment effects are concentrated in low-productivity offices—high-productivity offices show no significant response to deadline reductions. We interpret this as reflecting differential slack: high-productivity offices were already processing applications quickly and had little room to improve.

Table OA-18 provides direct evidence for this interpretation. We compare pre-treatment (before December 15, 2022) staff processing times to the *new* treatment deadlines, separately by office productivity quartile (using within-component normalized AKM fixed effects from the pre-RCT period). The results are striking. High-productivity offices (those with normalized AKM fixed effects > 1 SD above the mean) processed applications in 3.16 days on average, well below the mean treatment deadline. Consequently, 93.1% of observations in high-productivity offices already met the new deadline before the intervention. In contrast, low-productivity offices (< -1 SD) averaged 13.93 days and only 45.9% met the new deadline pre-treatment.

This pattern confirms that high-productivity offices had substantial slack relative to the new deadlines (mean slack of 9.82 days), while low-productivity offices operated beyond the constraint (mean slack of -1.20 days). Deadline pressure can only accelerate processing where there is room to improve.

Table OA-18: Pre-Treatment Processing Times by Office Productivity

	Low Prod (< -1 SD)	Mid-Low (-1 to 0)	Mid-High (0 to 1)	High Prod (> 1 SD)
Panel A: Pre-Treatment Staff Processing Time				
Mean (days)	13.93	10.06	5.45	3.16
SD (days)	18.66	13.40	9.23	6.37
Panel B: Comparison to New (Treatment) Deadlines				
Mean new deadline		12.78 days		
% Meeting new deadline	45.9%	61.6%	80.3%	93.1%
Mean slack (days)	-1.20	2.98	6.85	9.82
Observations	8,481	22,238	14,945	10,366
Number of offices	84	169	135	75

Notes: Sample restricted to treatment services in the pre-RCT period (before December 15, 2022) for offices with AKM productivity estimates. Office productivity is the within-component normalized AKM fixed effect estimated from staff processing time (higher values = faster processing). "Meeting new deadline" means total processing time $\leq$ the treatment deadline for that service. Slack = new deadline $-$ staff processing time. High-productivity offices were already processing well within the new deadlines before the intervention, leaving little room for improvement from deadline pressure.

I.8 Heterogeneity Regression Coefficients

Table 3 Panel C presents implied treatment effects at specific productivity levels (± 1 SD). For completeness, Table OA-19 reports the underlying regression coefficients from the quadruple interaction model. The implied effects are computed as $\text{ATE}(m, o) = \beta_0 + \beta_1 m + \beta_2 o + \beta_3 m \cdot o$, where m and o are standardized manager and office productivity.

The coefficient on Treatment $\times$ Post $\times$ Office ($\beta_2 = 2.31$, $p < 0.01$ for total time) confirms that the treatment effect is significantly more positive in fast offices—equivalently, significantly more negative in slow offices. The interaction term β_3 is small and not statistically significant for total time (-0.03, $p = 0.62$), suggesting no significant interaction between manager and office quality in moderating treatment effects.

I.9 Heterogeneity Event Study

I.10 Robustness of Productivity Measures

This section groups robustness checks for the productivity measures used in the heterogeneity analysis (Table 3 Panel C). Each subsection varies how manager or office productivity is measured and shows that the main findings are robust.

I.10.1 Non-RCT Productivity Measures

A potential concern with using AKM fixed effects in the heterogeneity analysis is that productivity measures estimated from *all* services—including treatment and control—may generate mechanical correlations. We address

Table OA-19: Heterogeneity Regression Coefficients

	(1) Total Time	(2) Manager Time	(3) Staff Time
Treatment × Post (β_0)	-0.89	-0.20	-0.67
	(0.72)	(0.13)	(0.63)
Treatment × Post × Manager (β_1)	0.17	0.30***	-0.14
	(0.18)	(0.04)	(0.16)
Treatment × Post × Office (β_2)	2.34***	0.16**	2.10***
	(0.26)	(0.07)	(0.21)
Treatment × Post × Mgr × Office (β_3)	-0.04	-0.06*	0.01
	(0.07)	(0.03)	(0.05)
Observations	221718	221718	221718

Standard errors in parentheses
This table shows the raw regression coefficients from the quadruple interaction model. AKM FEs estimated from pre-RCT sample (2021 through Dec 14, 2022). FEs normalized within each connected component. The implied treatment effects in Table 3 Panel C are computed as ATE$(m, o) = \beta_0 + \beta_1 m + \beta_2 o + \beta_3 mo$.
All specifications include week, service x office, and service x manager fixed effects. Standard errors two-way clustered by service and week in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

this partially by using only pre-RCT data to define the AKM fixed-effects. In this sub-section we re-estimate AKM fixed effects using only non-RCT services (those not included in the randomization) in the pre-RCT period and replicate the heterogeneity analysis. Restricting to non-RCT services reduces the heterogeneity sample from 210,335 to 186,024 observations (an 11.6% reduction), as some managers and offices only process the experimental services and therefore do not receive non-RCT fixed effect estimates. This sample loss does reduce precision, which should be kept in mind when interpreting the results below.

Panel B: Main Effects (Heterogeneity Sample, Non-RCT FEs)

	Total Time	Manager Time	Staff Time
Treatment × Post	−0.859	−0.215	−0.632
	(0.795)	(0.135)	(0.688)
Observations	190,460	190,460	190,460
Dep. var. mean	10.17	1.86	7.71

Panel C: Implied Treatment Effects by Productivity Type (Non-RCT FEs)

	Total Time	Manager Time	Staff Time
Low Manager, Low Office	-2.85^{***}	-0.84^{***}	-1.92^{**}
	(1.01)	(0.22)	(0.86)
Low Manager, High Office	0.98	-0.10	1.04
	(0.71)	(0.09)	(0.64)
High Manager, Low Office	-2.41^{**}	-0.13	-2.22^{***}
	(0.89)	(0.19)	(0.73)
High Manager, High Office	0.67	0.21^{*}	0.40
	(0.86)	(0.12)	(0.77)
Observations	190,460	190,460	190,460

I.10.2 Heterogeneity Using Pre-Treatment Processing Speed

The main heterogeneity analysis in Table 3 uses AKM fixed effects for both manager and office productivity, estimated from stage-specific outcomes (manager time and staff time, respectively). This approach cleanly separates manager and office contributions but requires connected components through manager mobility.

Table OA-20 presents an alternative specification that is entirely independent of the AKM decomposition. Both productivity measures are based on pre-treatment processing speed for treatment services (prior to December 15, 2022), standardized to mean zero and unit standard deviation. Office productivity uses mean total processing time; manager productivity uses mean AA (manager-stage) processing time. Higher values indicate faster processing. These measures do not require connected components and are available for a larger sample.

The speed measure provides intuitive motivation for heterogeneous treatment effects. Prior to the RCT, average processing times for treatment services were already below the new deadlines in 75% of offices, leaving limited room for improvement from deadline pressure. In the remaining 25% of offices, average processing times exceeded the new deadlines, providing scope for treatment effects.

The results are broadly consistent with the main AKM-based analysis. Treatment effects are concentrated in slow offices. We prefer the AKM-based results in the main text because they more cleanly isolate office-level productivity from manager contributions.

Table OA-20: RCT Heterogeneity Using Pre-Treatment Processing Speed

Panel A: Main Effects (Full Sample)

	Total Time	Manager Time	Staff Time
Treatment × Post	−0.613	−0.192	−0.399
	(0.829)	(0.126)	(0.732)
	[0.465]	[0.139]	[0.590]
Observations	394,626	394,626	394,626
Dep. Var. Mean	10.19	1.89	7.73

Panel B: Main Effects (Heterogeneity Sample)

	Total Time	Manager Time	Staff Time
Treatment × Post	−0.635	−0.198	−0.416
	(0.819)	(0.129)	(0.723)
	[0.444]	[0.136]	[0.570]
Observations	332,068	332,068	332,068
Dep. Var. Mean	10.20	1.91	7.73

Panel C: Implied Treatment Effects by Productivity Type

	Total Time	Manager Time	Staff Time
Low Manager, Slow Office	-9.40^{***}	-3.37^{***}	-5.92^{***}
	(1.27)	(0.57)	(0.97)
	[0.000]	[0.000]	[0.000]
Low Manager, Fast Office	5.02^{**}	-2.12^{**}	7.25^{***}
	(2.19)	(0.77)	(1.85)
	[0.029]	[0.010]	[0.000]
High Manager, Slow Office	-4.33^{**}	2.29^{***}	-6.45^{***}
	(1.73)	(0.47)	(1.55)
	[0.018]	[0.000]	[0.000]
High Manager, Fast Office	3.93^{*}	1.79^{***}	1.92
	(2.13)	(0.62)	(1.64)
	[0.076]	[0.008]	[0.253]
Observations	332,068	332,068	332,068

Notes: Treatment services received new deadlines set at the 60th–75th percentile of pre-treatment processing time. Treatment implemented December 15, 2022. **Panel A**: Full sample. **Panel B**: Sample with both productivity measures. **Panel C**: Heterogeneity by manager and office productivity. Manager productivity is the standardized pre-treatment mean AA (manager) processing speed for treatment services (higher = faster). Office productivity is the standardized pre-treatment mean total processing speed for treatment services (higher = faster). Both measures are independent of the AKM decomposition. "Low"/"High" and "Slow"/"Fast" = ±1 SD from mean. Implied effects computed via $\beta_0 + \beta_1 m + \beta_2 o + \beta_3 m \cdot o$; SEs via delta method. All specifications include week, service×office, and service×manager FEs. SEs two-way clustered by service and week in parentheses; p-values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I.11 Alternative Inference for Table 3 Panel C

Table 3 Panel C uses estimated AKM fixed effects—manager productivity $\hat{\alpha}_i$ and office productivity $\hat{\psi}_j$—as regressors in the heterogeneity specification. Because these are generated regressors estimated from a first-stage

AKM decomposition, conventional standard errors from the second-stage regression understate uncertainty by treating the estimated fixed effects as known (Battaglia et al., 2025; Kline, 2024; Pagan, 1984). Our primary approach to this problem is cross-fitting: Table OA-21 re-estimates the heterogeneity specification using cross-fitted AKM fixed effects following Chernozhukov et al. (2018); Kline (2024), which eliminates the generated-regressor bias entirely and makes conventional standard errors valid without correction (see Appendix I.12 for details). The cross-fitted standard errors are within 15% of the conventional ones for nearly every coefficient, confirming that the generated-regressor correction is negligible. As complementary evidence, Table OA-22 reports p-values from three procedures—pure randomization inference (which also addresses the small number of service clusters), combined bootstrap–randomization inference, and Rubin's combining rules—that corroborate this finding (see Appendix I.12.1).

I.12 Cross-Fitted AKM Fixed Effects

As our primary approach to the generated-regressor problem, we re-estimate the heterogeneity specification using *cross-fitted* AKM fixed effects, following the double/debiased machine learning (DML) framework of Chernozhukov et al. (2018); Kline (2024). Cross-fitting eliminates the generated-regressor problem: by estimating manager and office fixed effects on a training sample and using them as regressors in a held-out sample, the estimation error in the fixed effects becomes independent of the second-stage residuals, making conventional standard errors valid without further correction (see Kline, 2024, Section 5.2, for this argument in the AKM context).

We implement DML K-fold cross-fitting with $K = 5$. Following the recommendation of Kline (2025), we split at the *manager-office match* level—the unit of identifying variation in the AKM model. Each unique manager-office pair (dyad) is randomly assigned to one of $K = 5$ folds. Because most managers serve in multiple offices, removing one match from the training data still leaves the manager's remaining matches to identify their fixed effect in the other $K - 1$ folds.

For each fold k, we estimate the AKM model on the complementary $K - 1$ folds and predict manager and office fixed effects for the held-out observations in fold k. A manager i who serves in $J_i > 1$ offices has J_i matches spread across up to K folds, so they appear in the training sample for up to $K-1$ folds. Each manager's cross-fitted fixed effect is the simple average of these out-of-fold estimates (and analogously for offices). In our pre-RCT sample of 1,000 managers across 548 offices, the average manager appears in 3.7 of the 5 training folds; the average office appears in 3.9. This procedure yields cross-fitted fixed effects for 992 managers (99.2%) and 543 offices (99.1%); the few units lost are those whose matches all fall in a single fold, leaving them without any out-of-fold estimate. The cross-fitted fixed effects correlate 0.90 (managers) and 0.93 (offices) with their full-sample counterparts.[38]

Table OA-21 reports the results. For each cell of Panel C, we show point estimates and standard errors under the conventional and DML specifications. The SE ratios (DML divided by conventional) range from 0.72 to 1.36, with a median near 1, confirming that the generated-regressor correction is negligible. The pattern of significance is unchanged: the key heterogeneity results are robust to cross-fitting.

[38] We also computed exact leave-one-match-out cross-fitted fixed effects on the largest connected component (Kline et al., 2020). Results are qualitatively similar but have a smaller sample (the leave-out connected set comprises roughly two-thirds of the Panel C observations). These results are available upon request.

Table OA-21: Cross-Fitted FE Robustness for Table 3 Panel C

	Total Time		Manager Time		Staff Time	
	Conv.	DML	Conv.	DML	Conv.	DML
Low Mgr, Low Ofc	-3.34^{***}	-3.04^{***}	-0.72^{***}	-0.61^{***}	-2.53^{***}	-2.35^{***}
	(0.89)	(0.85)	(0.19)	(0.16)	(0.75)	(0.75)
Low Mgr, High Ofc	1.13	0.94	-0.29^{***}	-0.12	1.38^{**}	1.03
	(0.72)	(0.80)	(0.08)	(0.09)	(0.66)	(0.78)
High Mgr, Low Ofc	-3.11^{***}	-3.08^{***}	-0.00	-0.19	-3.00^{***}	-2.82^{***}
	(0.83)	(0.94)	(0.20)	(0.27)	(0.68)	(0.75)
High Mgr, High Ofc	1.81^{**}	1.78^{**}	0.20	0.09	1.51^{**}	1.63^{**}
	(0.71)	(0.68)	(0.12)	(0.13)	(0.64)	(0.61)
SE Ratio (DML / Conventional)						
Low Mgr, Low Ofc		0.96		0.84		1.01
Low Mgr, High Ofc		1.11		1.07		1.19
High Mgr, Low Ofc		1.14		1.36		1.10
High Mgr, High Ofc		0.96		1.09		0.95
Observations	221,718	221,256	221,718	221,256	221,718	221,256

Note: Each column shows implied treatment effects (in business days) for the four cells of the heterogeneity analysis, evaluated at ± 1 SD of manager and office quality. "Conv." uses full-sample pre-RCT AKM fixed effects (same as Table 3 Panel C). "DML" uses $K = 5$-fold cross-fitted fixed effects, where entire officer-office matches are assigned to folds. All specifications include week, service×office, and service×manager fixed effects. Standard errors (in parentheses) two-way clustered by service and week. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I.12.1 Resampling Procedures

Table OA-22 reports p-values from three computational inference procedures that complement the cross-fitting exercise. These procedures address two distinct concerns: (i) the small number of service clusters (N=30) over which treatment is randomized, and (ii) first-stage AKM estimation error. Point estimates and conventional standard errors are as reported in Table 3 Panel C.

An important caveat applies to the bootstrap-based procedures (Panels B–C). The pairs bootstrap is generally inconsistent when the number of estimated parameters grows with sample size (Battaglia et al., 2025; Kline, 2024). In our setting, approximately 1,400 manager and 559 office fixed effects are estimated from roughly 50 weeks of data per manager/office. The bootstrap-based results should therefore be viewed as suggestive rather than formally justified.

Panel A: Pure randomization inference. This procedure directly addresses the small-cluster concern by using the exact randomization distribution of treatment across 30 services, holding AKM fixed effects at their point estimates. We permute the treatment indicator 5,000 times across services and re-estimate the heterogeneity regression for each permutation, computing the p-value as the fraction of permutations for which the test statistic exceeds the observed value. This approach is valid regardless of the number of clusters and does not rely on asymptotic approximations.

Panel B: Combined bootstrap–randomization inference. This procedure accounts for both first-stage estimation uncertainty and the discrete randomization distribution simultaneously. For each of $B = 200$ bootstrap AKM replications (resampling managers with replacement from the pre-RCT data and re-estimating the AKM model), we permute the treatment indicator across the 30 services (50 permutations per replication, yielding 10,000 total draws). Each draw produces a new t-statistic under the null, computed with bootstrap-replicated AKM fixed effects *and* permuted treatment. The p-value is the fraction of draws for which $|t_{\text{null}}| \geq |t_{\text{obs}}|$, using the studentized approach of Chung and Romano (2013). This procedure simultaneously accounts for the discrete permutation distribution of treatment across a small number of services and the continuous variation in estimated productivity measures.

Panel C: Rubin's rules. This approach treats each bootstrap AKM replication as a multiple imputation of the unobserved true fixed effects, following the combining rules of Rubin (1987). For each of $B = 200$ bootstrap replications, we re-estimate the heterogeneity regression (Equation (4)) to obtain a point estimate $\hat{\delta}^{(b)}$ and its variance estimate $\hat{V}^{(b)}$. The combined point estimate is $\bar{\delta} = B^{-1} \sum_{b=1}^{B} \hat{\delta}^{(b)}$. The total variance combines within- and between-imputation components:

$$T = \bar{W} + \left(1 + \frac{1}{B}\right) B_v,$$

where $\bar{W} = B^{-1} \sum_b \hat{V}^{(b)}$ is the within-imputation variance (average of the B conventional variance estimates) and $B_v = (B-1)^{-1} \sum_b (\hat{\delta}^{(b)} - \bar{\delta})(\hat{\delta}^{(b)} - \bar{\delta})'$ is the between-imputation variance (reflecting variation across bootstrap AKM realizations). The factor $(1+1/B)$ corrects for the finite number of imputations. Inference uses t-tests with degrees of freedom adjusted for the relative magnitudes of within- and between-imputation variance (Rubin, 1987). When the between-imputation variance is small relative to the within-imputation variance—as in our application—the Rubin's rules standard errors are close to the conventional ones, confirming that first-stage estimation uncertainty contributes little to overall inference uncertainty.

Summary. Across all three panels, the pattern of results is consistent with the cross-fitting evidence in Table OA-21. The Low Manager, Low Office and High Manager, Low Office cells—where treatment reduces processing time most—remain significant or marginally significant under all inference procedures. Panel A, which does not rely on bootstrap resampling, provides the cleanest evidence that the heterogeneity results are robust to the small number of service clusters. The overall heterogeneity is robust to accounting for both finite-sample randomization and first-stage estimation uncertainty.

Table OA-22: Alternative Inference for Table 3 Panel C

	Total Time	Manager Time	Staff Time
Panel A: Randomization inference p-values			
Low Manager, Low Office	0.098*	0.025**	0.155
Low Manager, High Office	0.763	0.049**	0.706
High Manager, Low Office	0.063*	0.994	0.101
High Manager, High Office	0.595	0.640	0.628
Panel B: Combined bootstrap–randomization inference p-values			
Low Manager, Low Office	0.071*	0.045**	0.112
Low Manager, High Office	0.580	0.065*	0.525
High Manager, Low Office	0.049**	0.996	0.062*
High Manager, High Office	0.621	0.555	0.630
Panel C: Rubin's rules p-values			
Low Manager, Low Office	0.006***	0.026**	0.013**
Low Manager, High Office	0.138	0.189	0.068*
High Manager, Low Office	0.006***	0.995	0.003***
High Manager, High Office	0.071*	0.376	0.105
Observations	221,987	221,987	221,987

Notes: Panels A–C report p-values from inference procedures that address both the small number of service clusters (N=30) and first-stage AKM estimation error; see text for details. Panel A: pure randomization inference using 5,000 permutations of treatment assignment across services, with AKM fixed effects held at their point estimates. Panel B: combined bootstrap–randomization inference ($B = 200$ AKM replications $\times$ 50 permutations = 10,000 draws), which simultaneously accounts for AKM estimation error and finite-sample randomization. Panel C: Rubin's rules t-test ($B = 200$ replications). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I.13 Bootstrap Procedure for Table 3 Standard Errors

The bootstrap procedure underlying Panels B–C of Table OA-22 uses a pairs bootstrap at the manager level. In each of $B = 200$ replications, we resample managers with replacement from the pre-RCT estimation sample, re-estimate the two-way fixed effects AKM model to obtain new manager and office productivity estimates, and then re-estimate the heterogeneity regression (Equation (4)) on the RCT sample using the bootstrapped productivity measures. Frequency weights reflect the number of times each manager appears in the bootstrap sample. This pairs bootstrap preserves the correlation structure of the data and is valid under heteroskedasticity.