

Performance pay in the hybrid work economy

Jean-Victor Alipour*

August 4, 2026

[\(Click here for latest version\)](#)

Abstract

The rise of working from home (WFH) has granted workers unprecedented autonomy while weakening employer monitoring. We test the principal-agent prediction that firms respond by introducing performance pay (PP). Exploiting workers' exposure to the 2020 WFH shock through the WFH feasibility of their 2019 jobs, we find that WFH adoption increases PP, particularly among men and workers outside collective bargaining. WFH reduces labor supply when bundled with PP, consistent with incentive substitution: explicit performance-based pay replaces the implicit incentive to signal effort through longer hours. We document selection on gains: WFH is more eagerly adopted when paired with PP.

Keywords: Work from home, incentive pay, principal-agent, monitoring, signaling

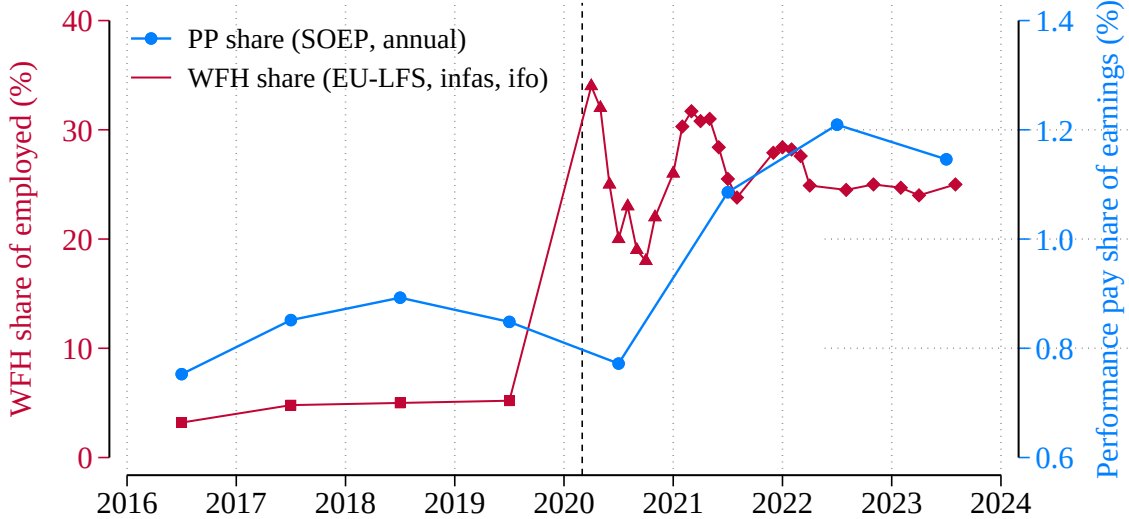
JEL Codes: J33, M52, M54, D82, J22

*alipour@ifo.de. LMU Munich, ifo Institute for Economic Research, and CESifo. I gratefully acknowledge support by the German Science Foundation (Deutsche Forschungsgemeinschaft, DFG) through CRC TRR 190 - project number 280092119.

1 Introduction

Principal-agent theory predicts that rising monitoring costs and increasing worker discretion incentivize employers to adopt output-based pay (Lazear, 2018; Holmstrom, 1979; Grossman and Hart, 1983; Prendergast, 2002; Holmstrom and Milgrom, 1991). Against this backdrop, Holmstrom and Milgrom (1991) speculate that “home office work should be accompanied by a stronger reliance on performance-based pay incentives [...]” (p. 27). The rapid and persistent spread of work from home (WFH, remote work) since 2020 provides a natural experiment to revisit this prediction at scale. Remote work grants workers unprecedented autonomy over their time and effort, while weakening employer monitoring. Indeed, Figure 1 documents a conspicuous comovement between WFH and performance pay (PP) in Germany: between 2019 and 2023, the share of employees regularly working from home rose nearly five-fold, stabilizing at about 25% as the economy settles into a “hybrid work” equilibrium, where the typical arrangement features a mix of home and office-based work (Barrero et al., 2026).¹ Over the same period, the aggregate share of performance pay rose by more than a third to 1.15% of annual earnings.

Figure 1: Trends in working from home and performance pay in Germany



Notes: The figure plots the share of performance pay (profit sharing, bonuses) in annual earnings (data from SOEP) and the share of employees working from home regularly (data from EU-LFS 2016–2019, infas Casa Monitor Survey 2020, and ifo Business Survey 2021–2024).

This paper asks whether there is a causal link between the rise in remote work and the importance of performance pay. Using panel data from the German Socio-Economic

¹In contrast to casual reports of a “return-to-office trend”, WFH adoption worldwide shows no signs of reversal and is widely expected to persist (Aksoy et al., 2025a; Barrero et al., 2025).

Panel (SOEP) through 2023, we study how WFH adoption, forced by the Covid-19 pandemic, affects within-individual changes in compensation structure. The identifying variation comes from worker-level differences in exposure to the WFH shock, captured by whether their 2019 job *could* be done from home (*WFH feasibility*). We estimate difference-in-differences (DiD) models that compare workers with different WFH feasibility over time, and instrumented-DiD (DiD-IV) models that instrument 2023 WFH status by 2019 feasibility to correct for self-selection into WFH arrangements. The results consistently support [Holmstrom and Milgrom](#)’s prediction: WFH adoption significantly raises the importance of performance pay in workers’ compensation.

This is the first study to examine the causal effect of WFH on performance pay. We contribute to understanding the labor-market consequences of the “Big Shift” to remote work after 2020, which bundled lasting changes in work arrangements, worker preferences, technology, and employer perceptions ([Barrero et al., 2026](#); [Davis et al., 2026](#); [Barrero et al., 2023](#); [Chen et al., 2023](#); [Monte et al., 2023](#)). A fast-growing literature studies the effects of WFH on compensation ([Castex and Dechter, 2026](#); [De Fraja et al., 2024](#); [Choi et al., 2026](#)), labor supply ([Schüller, 2026](#); [Harrington and Kahn, 2025](#); [Alipour, 2025](#)), worker selection ([Emanuel and Harrington, 2024](#); [Atkin et al., 2023](#); [Bloom et al., 2015](#)), and productivity. Productivity effects are highly context-dependent, and tend to settle on neutral to moderately positive under hybrid WFH ([Bloom et al., 2024](#); [Choudhury et al., 2024](#); [Emanuel and Harrington, 2024](#); [Angelici and Profeta, 2023](#); [Shen, 2023](#); [Fenizia and Kirchmaier, 2025](#)). Workers overwhelmingly derive non-pecuniary benefits from WFH ([Bagga et al., 2025](#); [Drake et al., 2025](#); [Mas and Pallais, 2017](#); [Nagler et al., 2024](#); [Barrero et al., 2023](#); [Maestas et al., 2023](#)), which allows employers to trade off slower wage growth against the amenity value of remote work ([De Fraja et al., 2024](#); [Barrero et al., 2022](#); [Aksoy et al., 2025b](#)).

Our key contribution is to show that WFH changes compensation *structure* by shifting pay from fixed to variable components, consistent with the principal-agent prediction that weaker monitoring and greater worker discretion increase reliance on performance incentives to overcome moral hazard. This mechanism also offers a new argument for why labor supply may fall under WFH: when working hours, including overtime and nonstandard hours, become less visible to the employer, they lose value as a signal of effort or worker quality ([Frederiksen et al., 2024](#); [Bohnet and Ciasullo, 2026](#); [Landers et al., 1996](#); [Anger, 2008](#)). Consistent with this prediction, the data reveal a drop in total hours and *uncompensated* overtime due to WFH. The decline is larger where WFH is paired with performance pay, which is expected when

implicit incentives (hours-signaling for raises or promotions) are replaced by an explicit output-based contract (bonus for performance) (Baker et al., 1994; Bohnet and Ciasullo, 2026).

The paper thus speaks directly to the empirical principal-agent literature and the design of incentive contracts (Lazear, 2018; Hartzell et al., 2010). While previous work has documented associations between monitoring costs and contract form, causal evidence comes almost entirely from case studies of single firms or narrow industries (Baker and Hubbard, 2004; Nagin et al., 2002; Pierce et al., 2015), lab experiments (Falk and Kosfeld, 2006; Dickinson and Villeval, 2008; Boly, 2011), or informal contracting in developing countries (Houeix, 2026; Kelley et al., 2024).² Broader evidence on pay composition typically focuses on executives (Cuñat and Guadalupe, 2005, 2009; Garen, 1994; Baker and Hall, 2004; Mengistae and Colin Xu, 2004; Gayle and Miller, 2009). Our study distinguishes itself by exploiting WFH as an unanticipated, *economy-wide* shock to monitoring costs and worker autonomy in a formal labor market. We identify its effect on the within-individual (within-firm) composition of pay, an outcome directly implied by the classical principal-agent model yet rarely examined outside the executive suite.³

Finally, we contribute to the literature on structured management practices (SMPs), which identifies formalized monitoring, target-setting, and incentive systems as key drivers of firm performance (Bloom and Van Reenen, 2007; Bloom et al., 2013). Lamorgese et al. (2024) emphasize a complementarity between WFH and SMP by showing that better-managed firms adopted remote work more extensively during the pandemic, due mainly to differences in monitoring and incentive practices. Rather than asking whether firms with better management practices adopt remote work, we show that the WFH shock itself induces firms to strengthen performance-based incentives within existing worker-firm relationships. The observed correlation between management practices and WFH thus partly reflects firms adopting new practices in response to the technology shock, and not only the selection of better-managed firms into WFH arrangements.

²These studies also examine adjacent outcomes rather than within-firm pay composition: asset ownership in trucking (Baker and Hubbard, 2004), worker shirking, effort, and theft (Nagin et al., 2002; Dickinson and Villeval, 2008; Boly, 2011; Pierce et al., 2015), and driver effort and contract terms in informal transit (Houeix, 2026; Kelley et al., 2024).

³Cuñat and Guadalupe (2005) and Cuñat and Guadalupe (2009) come closest to our setting, showing that increases in product-market competition raise the variable share of executive compensation.

Restoring input monitoring technologically is an alternative response to the same shock. [Ko and Baek \(2025\)](#) show that electronic monitoring increased productivity among teleworking patent examiners at the USPTO but also raised worker dissatisfaction. In Germany, however, employee data-protection laws and works-council co-determination rights substantially restrict comparable forms of remote surveillance.⁴ These institutional constraints make restoring input monitoring more difficult and increase the relevance of output-based incentives as an organizational response to WFH.

The core analysis uses a panel of 3,443 individuals employed in 2019 and reinterviewed in 2023 and 2024 in the German SOEP ([Goebel et al., 2019](#)). The key outcomes are an indicator for receiving performance pay and its share in annual earnings. PP includes profit-sharing and bonus pay, but excludes other extras such as 13th-month pay and Christmas or vacation allowances.

OLS estimates show a positive within-individual correlation between WFH at least once per week in 2023 and 2019–2023 changes in performance pay incidence, both at the entry margin (workers transitioning from fixed pay to some PP) and the retention margin (workers with PP in 2019 remaining in PP). A causal interpretation is problematic, of course, because workers self-select into WFH arrangements along characteristics that may correlate with trends in compensation structure.

The solution is to exploit variation across workers in their *ex ante* exposure to the WFH shock through the WFH feasibility of their 2019 job. By instrumenting realized WFH with feasibility, we can disentangle the causal channel from selection. Because the Big Shift to WFH was exogenous, workers did not select jobs based on anticipated WFH opportunities. And because feasibility is predetermined, it is unaffected by pandemic shocks that may also affect WFH adoption.⁵ We show that 2019 WFH feasibility strongly predicts realized WFH in 2023, confirming its relevance as an instrument.

However, WFH feasibility is hardly as-good-as-randomly assigned. It correlates with worker and job characteristics (e.g., income, education), violating the independence assumption of a standard IV. We therefore adopt a DiD-IV design ([de Chaisemartin,](#)

⁴The USPTO scheme would likely be unlawful in Germany. Recording login and logout times is permitted, but continuously metering keyboard and mouse activity to detect idle time is generally considered a disproportionate form of performance monitoring under employee data-protection law (§ 26 *Bundesdatenschutzgesetz*) and the general right of personality. German courts interpret these limits narrowly, holding even covert keystroke loggers unlawful absent concrete suspicion of serious misconduct (BAG, 27 July 2017, 2 AZR 681/16).

⁵Even workers with pre-2020 WFH experience could not have foreseen the shift in intensity, supporting technologies, and perceptions, and should therefore be considered “treated” by the Covid-induced WFH shock.

2010; Miyaji, 2025), which replaces the independence assumption with common-trends restrictions: the outcome and endogenous variable (realized WFH) would have evolved in parallel across WFH-feasibility levels absent the post-2020 WFH expansion. The design identifies a local average treatment effect on the treated (LATET): the effect for workers who work remotely because their 2019 job was already compatible with WFH.

To address concerns that unobserved post-2019 shocks confound the estimates, we condition on fixed effects that restrict identifying variation to individuals within the same industry. Results are robust to additional broad-occupation and state fixed effects plus demographic and job controls, which account for different regional mitigation policies, labor-market, demand, or supply shocks, and differential trends in compensation caused by observable individual characteristics.

Event-study estimates of the effect of 2019 WFH feasibility on performance pay and realized WFH show no meaningful pre-trends. Both PP incidence and its share of annual earnings increase from 2020 onward. By 2023, workers in WFH-feasible 2019 jobs are 7.5 percentage points more likely to receive PP; the PP share of annual earnings is 0.55 percentage points higher. Both the entry and retention margins contribute to this increase: workers without performance pay at baseline are more likely to enter PP contracts, while those with PP are less likely to exit them. Among workers already on PP contracts, however, WFH does not increase the importance of PP in their annual earnings, on average. We present several pieces of evidence that the observed changes indeed reflect changes in compensation structure rather than a higher probability of triggering an existing performance-pay contract.

We estimate the DiD-IV model of realized WFH in 2023 on performance pay using 2SLS. The estimated LATET implies that WFH adoption raises the individual likelihood of receiving PP by 19 percentage points between 2019 and 2023 (equivalent to a 79% increase relative to the 2019 treated mean). The PP share of annual earnings rises by 1.4 percentage points (70%). In line with the DiD estimates, the effect on PP incidence is driven by both entry into PP and retention, whereas the WFH-induced increase in the PP share is confined to the entry margin.

The IV estimates exceed OLS, pointing to meaningful self-selection. WFH and pay structure are likely bargained jointly. In that case, compliers (whose feasibility translates into WFH adoption) accept a bundle combining WFH and performance pay. We find support for this “selection on gains” pattern in our analysis of marginal treatment effects (MTE). The MTE curve slopes downward in unobserved resistance to treatment,

indicating that workers more readily induced into WFH by higher feasibility are also more likely to transition into performance pay. Or from the firm’s perspective, WFH is more readily adopted when it can be paired with management practices that align workers’ incentives with firm objectives.

Individuals with a lower private value of WFH may also be less willing to accept a shift toward performance pay in exchange for remote work, pulling down the OLS estimate. Consistent with this interpretation, never-takers exhibit stronger career orientation than compliers, potentially reflecting promotion concerns related to WFH ([Castex and Dechter, 2026](#)) or greater complementarity between their tasks and office work. Never-takers are also more plausible to receive performance pay through traditional career channels (senior-role bonuses, promotions) even absent the WFH shock ([Prendergast, 1999](#); [Lazear and Rosen, 1981](#)), explaining why the OLS is attenuated.

Heterogeneity analyses further support the selection-on-gains interpretation. WFH-induced increases in performance pay are concentrated among workers not covered by collective wage agreements, where employers have greater discretion to adjust compensation. The effect is also concentrated among men, consistent with evidence that women are more averse to incentive pay, primarily because of differences in risk preferences ([Dohmen and Falk, 2011](#)). Another possibility is that employers anticipate that incentive pay is less needed to ensure women’s effort once WFH weakens monitoring ([Segal, 2012](#)).⁶ Finally, the effects are similar among workers who remain with the same employer, ruling out that the findings are driven by sorting across firms.

Finally, we show that WFH reduces total hours worked and the incidence of uncompensated overtime, which often serves as a signal of effort or worker quality to the principal ([Anger, 2008](#); [Bohnet and Ciasullo, 2026](#)), and can lead to inefficiently high labor supply ([Landers et al., 1996](#)). This is consistent with hours losing their signaling value once they become harder to observe. The overtime reduction is concentrated among groups with stronger WFH-induced shifts toward performance pay, where explicit performance incentives substitute for the implicit career rewards associated with overtime signaling. This incentive-substitution mechanism cannot be cleanly separated from labor-supply responses to the flexibility of WFH (e.g., time savings and greater autonomy), so we interpret the evidence cautiously. Nonetheless, the overtime reduction survives conditioning on 2019 baseline contractual hours and the change in contractual hours, lending confidence to the substitution mechanism.

⁶Conditional on receiving incentive pay, a meta-analysis finds no gender differences in responsiveness ([Bandiera et al., 2021](#)).

Moreover, a formal mediation analysis corroborates that the shift toward performance pay mediates part of the reduction in hours.

We close with a brief outlook on the implications of a structurally larger role for performance pay. First, the finding that WFH is more readily adopted when bundled with performance pay may help reconcile some of the ambiguous results on the performance effects of WFH. If they hinge on complementary investments in incentive schemes or other organizational changes, effects will vary with firms’ capacity or willingness to make such investments. Second, a rise in performance-based compensation implies a lower optimal degree of progressivity than benchmarks that treat earnings risk as exogenous rather than endogenously set by the firm (Doligalski et al., 2023). Because the WFH-induced adoption of PP is concentrated among workers with higher baseline earnings, it also has the potential to widen top-end earnings inequality within jobs exposed to the WFH shock. We present descriptive evidence consistent with this consequence.

The next section describes the data and the construction of the WFH feasibility measure; Section 3 outlines the instrumented DiD strategy and discusses the identifying assumptions; Section 4 presents the results. We conclude and discuss implications in Section 5.

2 Data & WFH feasibility measure

2.1 German Socioeconomic Panel (SOEP)

The SOEP is a nationally representative annual panel survey of households in Germany (Goebel et al., 2019). We use Version 41, which covers waves through 2024.

Analysis sample. The estimation sample consists of individuals employed in 2019 (excluding marginal workers, trainees, self-employed, freelancers, and military personnel), aged 15–60 at baseline, and observed in both 2019 and 2023 ($N = 3,443$ workers). We use 2019 and 2023 as the pre- and post-treatment years, which is the longest within-person window for which the key outcomes are available.

Performance pay (PP). The SOEP wave records, for each respondent, whether and how much “profit-sharing, gratification, and bonus pay” (*Gewinnbeteiligung, Gratifikation, Prämie*) was received in the *prior* calendar year, together with annual labor earnings for that same year. We label this composite “performance pay” (PP) because, under standard principal-agent theory, both tie remuneration to a signal correlated with effort or output and shift income risk onto the worker in exchange for stronger

incentives ([Holmstrom, 1979](#); [Prendergast, 1999](#)).⁷ We construct outcomes for each reference calendar year from the retrospective questions: a binary indicator for receiving any performance pay, and its share of annual labor earnings.

Realized WFH. The 2023 wave records how often workers use WFH in the current year. We code a binary indicator equal to one for workers reporting WFH at least once per week. No WFH status is recorded in 2019. We explain in [Section 3](#) that, given the low take-up pre-2020 and the substantial changes to the organization, scale, supporting technologies, and perceptions of WFH, assuming no WFH in 2019 does not present a concern for identification. Robustness checks that exclude likely pre-2020 adopters yield the same results. In addition, we use administrative repeated cross-sectional data from the micro census (*Mikrozensus*), covering 947,453 dependent employees with information on realized WFH from 2017 to 2023.⁸ The data are useful to test the absence of pre-trends in WFH adoption by future exposure to the Covid shock and to confirm that the subsequent rise in WFH is driven by those classified as having a teleworkable job.

2.2 Measuring exposure to the WFH shock

We capture worker-level exposure to the WFH shock by the WFH feasibility of each individual’s 2019 job. Importantly, the SOEP does not elicit WFH feasibility from workers before 2020. For lack of job-level measures, the literature has typically relied on occupation-level feasibility indices (e.g., by [Dingel and Neiman, 2020](#)). Instead, we follow [Alipour \(2025\)](#), who adopts a two-step procedure to recover job-level feasibility. The first step is to train a simple classifier model to predict job-level WFH feasibility from pre-pandemic job and worker characteristics in a separate dataset. The 2018 BIBB-BAuA Employment Survey provides the relevant information. We code jobs as WFH-feasible if the respondent neither reports that “WFH is impossible” in their job nor *frequent* exposure to work conditions incompatible with remote work (heavy lifting, outdoor work, machine operation, exposure to dust or chemicals). We train and validate a simple LASSO-logit classifier model on the BIBB that predicts binary WFH feasibility as a function of job and demographic characteristics (3-digit occupation, 2-digit industry, skill-requirement level, age, gender, and migration background). In the second step, the trained model is applied to the 2019 cross section of the SOEP

⁷Routine end-of-year supplements such as 13th-month pay and Christmas or vacation allowances are not part of the PP measure.

⁸The Mikrozensus includes the German part of the EU labor force survey and is comparable to the U.S. CPS, and is supplied by the Federal Statistical Office (Destatis).

sample. Using corresponding characteristics, it predicts a continuous WFH feasibility score $\in [0, 1]$ and a binary indicator at the optimal classification cutoff for each SOEP worker.

Two features make this two-step approach attractive. First, the imputation attenuates idiosyncratic reporting variation in the self-reported feasibility item that arises from subjective constraints unrelated to the job content, such as employer policy, home-office space, or local infrastructure. The predicted measure thus approximately captures WFH feasibility explained by jobs' task profiles. Second, the measure varies *within* occupation, providing useful identifying variation. Importantly, because feasibility is determined by 2019 characteristics, the measure itself and any sorting are plausibly unrelated to the subsequent Covid-induced WFH shock. Construction details and validation diagnostics are reported in [Appendix C](#).

3 Empirical Strategy

Our goal is to estimate the causal impact of post-2020 WFH adoption on workers' compensation structure. Because workers self-select into WFH along characteristics that plausibly correlate with the outcome, OLS conflates the effect of *who* works from home with the causal effect of WFH. We address this with an IV design that instruments 2023 WFH status with the WFH feasibility of the worker's 2019 job. This isolates the variation in WFH adoption driven by exposure to the exogenous WFH shock caused by Covid.

A standard IV would identify the local average treatment effect (LATE) under the assumption that the instrument is as-good-as randomly assigned (instrument independence). But this fails in our setting because WFH feasibility correlates with worker and job characteristics, including income and education ([Alipour et al., 2023](#)), which may independently affect the outcome. The solution is to adopt an *instrumented difference-in-differences* (DiD-IV) design ([de Chaisemartin, 2010](#); [Miyaji, 2025](#)), which replaces the independence assumption with two common-trends restrictions: that the outcome (performance pay) *and* the endogenous variable (realized WFH) would have evolved in parallel across WFH-feasibility levels absent the post-2020 WFH expansion. Instead of a LATE (the average effect of WFH for the full complier population), the design

identifies a local average treatment effect *on the treated* (LATET):⁹ the effect for compliers in the exposed group, i.e., workers who work remotely in 2023 *because* their 2019 job was WFH-feasible.¹⁰

We estimate the DiD-IV model by two-stage least squares (2SLS) on within-individual first differences (2019–2023). The first stage estimates WFH status in 2023 as a function of 2019 feasibility:

$$WFH_i^{2023} = \pi_0 + \pi_1 WFH_FEAS_i^{2019} + \pi_2' \mathbf{x}_i^{2019} + \vartheta_i, \quad (1)$$

and the second stage estimates 2019–2023 outcome changes as a function of predicted WFH:

$$\Delta y_i = \beta_0 + \beta_1 \widehat{WFH}_i^{2023} + \beta_2' \mathbf{x}_i^{2019} + \epsilon_i, \quad (2)$$

where $\Delta y_i \equiv y_i^{2023} - y_i^{2019}$ is the within-individual change in outcome y , WFH_i^{2023} is a binary indicator equal to one if individual i reports working from home at least once per week in 2023, $WFH_FEAS_i^{2019}$ is the (predicted) WFH feasibility of i 's 2019 job (binary at the Youden cutoff, or continuous on $[0, 1]$), and $\mathbf{x}_i^{2019}$ is a vector of pre-treatment controls measured in 2019. In the baseline, we include industry fixed effects (motivated below). The reduced-form intent-to-treat counterpart replaces $\widehat{WFH}_i^{2023}$ in Equation 2 with $WFH_FEAS_i^{2019}$ directly. OLS uses observed WFH_i^{2023} in Equation 2. Standard errors are clustered by individual in the baseline.¹¹

We do not observe WFH status in 2019, so the design identifies the effect of post-2020 WFH adoption regardless of pre-pandemic status. Assuming zero WFH in 2019 is sensible for two reasons. First, pre-2020 WFH was rare: about 5% of workers in Germany regularly worked from home in 2019 (Figure 1). But more importantly, the forced experiment in 2020 permanently changed the organization, technology, and perceptions of WFH arrangements. The collaboration tools that make modern WFH viable (video-conferencing and cloud sharing) were adopted at scale only after 2020 (Barrero et al., 2023), and the pre-pandemic stigma against WFH dissolved as workers

⁹An alternative formalization is the Fuzzy DiD framework of de Chaisemartin and D'Haultfœuille (2018), which targets a switcher LATE on the treated (SLATET) rather than the LATET and relies on stronger, asymmetric assumptions on treatment adoption than the symmetric monotonicity of the DiD-IV design (Miyaji, 2025). Both frameworks use the same Wald-DiD estimand (the ratio of the DiD estimand of the outcome to that of the treatment), which is estimated by 2SLS and differ only in the parameter it identifies and the required assumptions. With negligible WFH adoption before 2020 and the common trends in WFH we maintain (see below), the two parameters coincide.

¹⁰In this sense, the LATET stands to the LATE as the ATT stands to the ATE in a standard DiD.

¹¹By first-differencing the outcome, the model is equivalent to a two-period DiD absorbing time-invariant individual heterogeneity, with $WFH_FEAS_i^{2019}$ as the treatment variable in the reduced form.

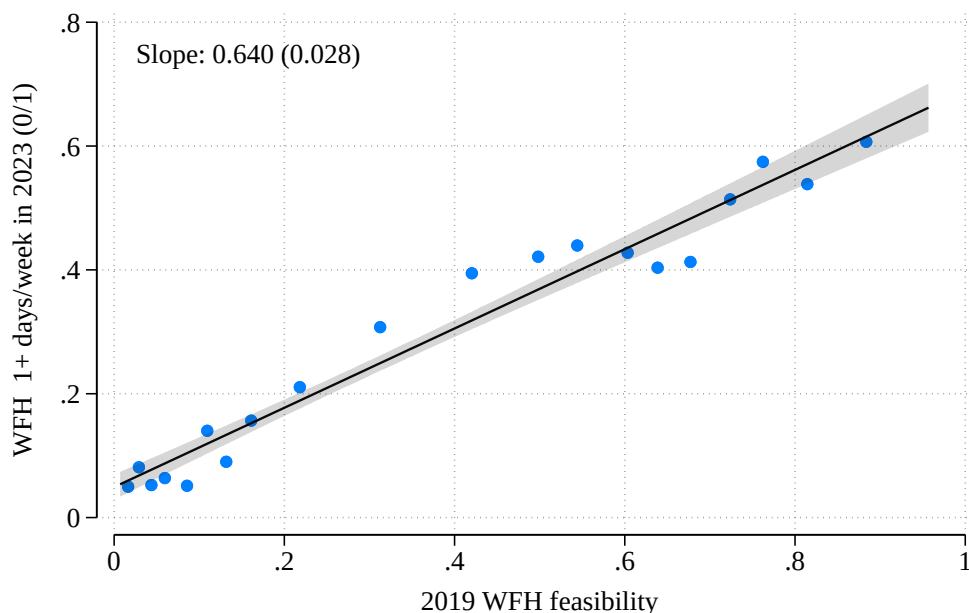
and employers gained experience (Barrero et al., 2026). Consequently, the amenity premium for WFH roughly quadrupled between 2019 and 2023 (Bagga et al., 2025). Thus, even workers with pre-2020 WFH experience should be considered treated by the Covid-induced WFH shock. Nonetheless, we show below that the results are robust to excluding (likely) pre-2020 adopters.

We estimate the model both pooled and separately by baseline (2019) PP status (0/1), since performance pay can respond to WFH on two different margins: workers without PP at baseline can transition onto performance pay (*entry* margin), whereas workers already on PP can experience changes in its share or composition (*retention* margin).

The validity of the DiD-IV design hinges on several assumptions.

Relevance. First, we require that the instrument have a non-trivial impact on realized WFH adoption: $\pi_1 > 0$. This is confirmed in Figure 2, which shows a binscatter regression of 2023 WFH status on 2019 WFH feasibility, conditional on industry fixed effects. Moving WFH feasibility from 0 to 1 raises the probability of WFH at least once per week in 2023 by about 63 percentage points. The Kleibergen–Paap first-stage F -statistic exceeds 400, refuting weak-instrument concerns.

Figure 2: First-stage effect of 2019 WFH feasibility on WFH status in 2023



Notes: The figure shows a binscatter regression of WFH in 2023 (0/1) on WFH feasibility 2019, conditional on industry fixed effects (Equation 1). 95% confidence bands are drawn based on standard errors clustered by individual. The positive and significant first-stage effect is consistent with the relevance condition required for LATET interpretation of the DiD-IV estimand.

Monotonicity. Second, monotonicity requires that higher 2019 WFH-feasibility does not reduce the likelihood of working from home in 2023 for any worker, ruling out “defiers”. The assumption cannot be tested directly; an indirect check is whether the first-stage relationship between feasibility and realized WFH is positive across subgroups. [Table B.1](#) confirms that subgroup first-stage estimates are positive across sector, occupation, education, age, firm-size, and baseline PP status cells, supporting the monotonicity assumption and showing that the instrument is relevant even within occupations with low average WFH feasibility such as land/forestry and construction.

No anticipation and parallel trends (PT). Third, we require two DiD-related conditions, (i) *No anticipation*: workers and firms did not react in advance to the 2020 WFH shock, which is plausible given that the pandemic and its WFH consequences were unforeseen, and (ii) *Parallel trends*: in the counterfactual absent the WFH shock, post-2020 outcome and WFH-adoption trajectories would have evolved similarly across workers with different WFH feasibility. PT could fail due to selection on anticipated trends, i.e., workers sorting into jobs with different feasibility based on expectations about how WFH or performance pay would evolve. Again, this is arguably of minor relevance given the absence of meaningful WFH expansion before 2020 and the exogenous nature of the pandemic. A more serious concern is that correlates of WFH feasibility, such as educational attainment or industry, may co-determine trends in performance pay. In this case, the DiD-IV estimates of WFH could be confounded by pre-existing trends. PT are not directly testable. We assess the assumption indirectly by estimating event-studies and inspecting pre-2020 trends, which can indicate violations of PT. The results, presented in the following section, show no evidence of relevant pre-trends. But the absence of pre-trends cannot rule out that *post-2020 pandemic shocks* are correlated with WFH feasibility and also cause breaks in outcome trajectories; for instance, demand or supply shocks caused by shutdowns at the industry or the local labor market level. We address this by controlling for industry fixed effects in all specifications, ensuring that workers are compared within the same industry, and we test robustness to more demanding specifications that restrict the identifying variation within industry, occupation, and federal state, and control for job-specific and demographic trends.

Exclusion restriction. Fourth, we require that $WFH_FEAS_i^{2019}$ affects Δy_i only through realized WFH. Importantly, this means that any direct or non-WFH effects of WFH feasibility are not problematic so long as they are time-invariant. Besides this, 2019 WFH-feasibility is a job-task characteristic with no obvious channel through

which it could influence compensation structure except by enabling actual WFH adoption. An exclusion violation operating through some non-WFH channel would manifest as differential pre-2020 trends across feasibility groups, at least to the extent that such a channel was active in the pre-treatment period. Checking pre-trends in the event study thus provides a useful diagnostic for the exclusion restriction as well.

4 Results

4.1 Reduced-form event study results

Performance pay. [Figure 3](#) reports event-study DiD estimates of WFH feasibility on performance pay using the [Callaway and Sant’Anna \(2021\)](#) estimator, conditional on industry fixed effects. We estimate the model on the pooled sample (Panel I) and separately for the *entry margin* (workers without PP in 2019, Panel II) and the *retention margin* (workers with PP in 2019, Panel III). The estimates recover the intent-to-treat (ITT) effect of WFH on performance pay and serve as a diagnostic for the parallel-trends assumption underlying the DiD-IV design.

WFH feasibility predicts a sharp rise in performance pay after 2020. In the pooled sample (Panel I), a 0 to 1 jump in WFH feasibility raises the individual likelihood of receiving PP by 8.3 percentage points on average across 2020–2023 (Panel I.A) and the PP share of annual earnings by 0.3 percentage points (Panel I.B). Pre-trends are insignificant for both outcomes.

Splitting the sample by 2019 PP status shows that the pooled rise comes mostly from entry into PP schemes. Panel II shows that, among workers without PP at baseline, WFH feasibility raises the likelihood of transitioning into PP by 5.4 percentage points and raises the PP share of annual earnings by 0.5 percentage points, compared to workers less exposed to WFH. Panel III shows that, among workers already on PP in 2019, WFH feasibility raises the probability of remaining on a PP contract by 18.8 percentage points but leaves PP intensity unchanged, on average.

Pre-2020 coefficients are nearly all insignificant (except in Panel II.B), but carry some imprecision, which may mask underlying violations of the PT assumption ([Roth, 2022](#)). We assess how sensitive treatment effects of WFH exposure are to PT violations using the approach of [Rambachan and Roth \(2023\)](#). Assuming that post-2020 deviations from PT are “not too different” from pre-2020 deviations, the method calculates confidence intervals that account for both deviations from PT and statistical uncertainty of pre-trends estimates. Results in [Figure A.1](#) show that all average treatment effects

are preserved under linear violations of PT. More conservatively, allowing for (compounding) deviations equal to half the pre-2020 curvature continues to preserve the entry effects and the pooled effect on PP receipt; confidence intervals for the retention margin widen substantially and include zero.¹² Taken together, the event-study results give us confidence in the DiD-IV design and indicate that the WFH shock caused a permanent rise in performance pay, consistent with [Holmstrom and Milgrom](#)’s prediction.

Realized WFH. The DiD-IV design also requires parallel trends in the endogenous variable (realized WFH). We assess this assumption using repeated cross-sectional data spanning 2017–2023 from the German *Mikrozensus*. Panel A of [Figure 4](#) plots the share WFH at least half of working days, by WFH feasibility. Among WFH-infeasible jobs, the rate stayed near 1% before 2020 and rose only modestly afterward, to 3–4%. Among WFH-feasible jobs, it was higher before Covid (around 6%), then jumped sharply in 2020 and remained above 22% through 2023. Panel B reports the corresponding event-study estimates. Higher feasibility leads to a sharp and persistent increase in realized WFH across 2020–2023, while the pre-2020 coefficients are negligible.

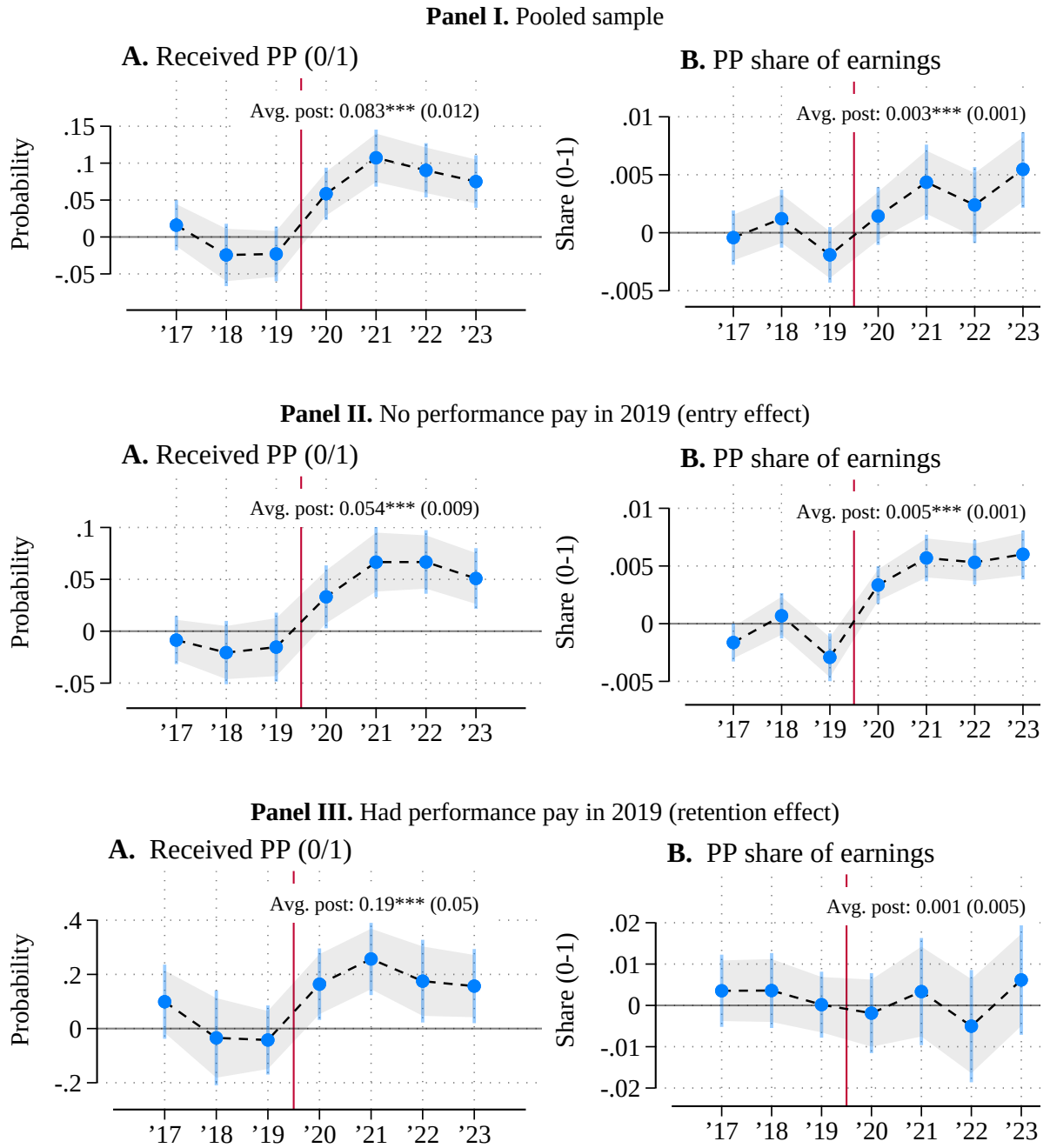
4.2 The causal effect of WFH on performance pay

The intent-to-treat effects recovered by the DiD estimator are attenuated compared to the effect of *realized* WFH, because not all WFH opportunities are taken up. The DiD-IV estimates recover the causal effect of realized WFH for the subpopulation of workers who work remotely because their 2019 job was amenable to it.

[Figure 5](#) reports OLS relationships between WFH at least once per week in 2023 (0/1) and within-individual (2019–2023) change in performance pay, followed by 2SLS conditional on industry fixed effects, and 2SLS with additional occupation and federal-state fixed effects (Equations 1 and 2). We show results separately for the pooled sample (Panel I), workers without PP in 2019 (Panel II, entry margin), and workers with PP in 2019 (Panel III, retention margin). [Table B.2](#) reports the corresponding regression table.

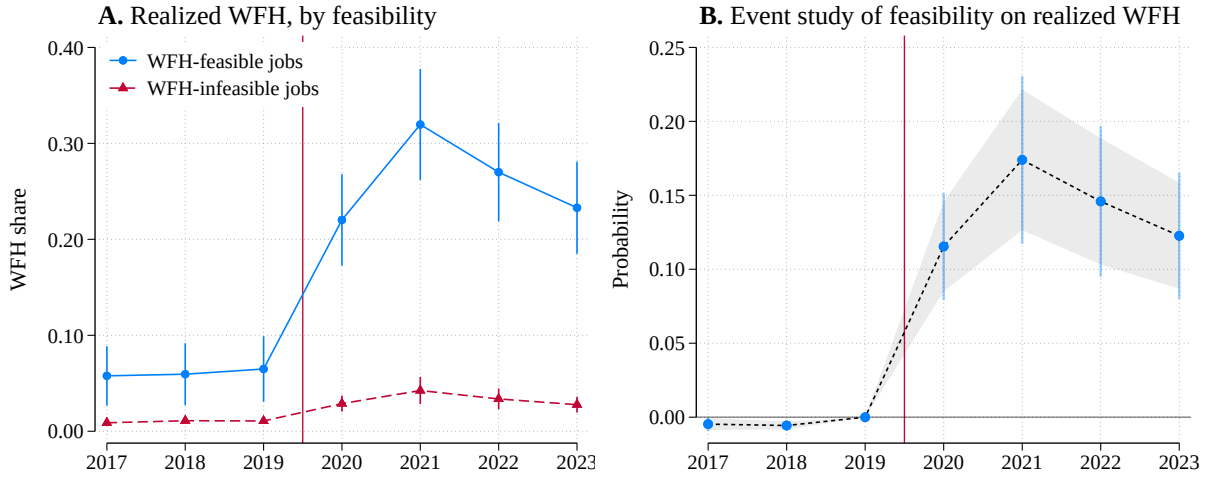
¹²Formally, we impose the smoothness restriction $\Delta^{SD}(M)$, which bounds how much the slope of the difference in trends may change between consecutive periods. $M = 0$ imposes linearity. We also use $M > 0$, at 0.5 of $\hat{M}$, which is the average second difference among the 2016–2018 coefficients. The sets are uniformly valid and reflect both statistical uncertainty and uncertainty about how the difference in trends extrapolates.

Figure 3: DiD estimates of 2019 WFH feasibility on individual performance pay



Notes: The figure shows event-study estimates using the [Callaway and Sant'Anna \(2021\)](#) estimator, conditional on industry fixed effects. Outcomes are receipt of performance pay (0/1) (Panel A) and the PP share of annual earnings (Panel B). Panel I is the pooled estimation sample; Panel II restricts to workers without performance pay in 2019 (entry margin); Panel III to workers with performance pay in 2019 (retention margin). Shaded areas and spikes show 90% and 95% confidence intervals, respectively, using standard errors clustered by individual.

Figure 4: WFH feasibility on realized WFH (repeated cross section)



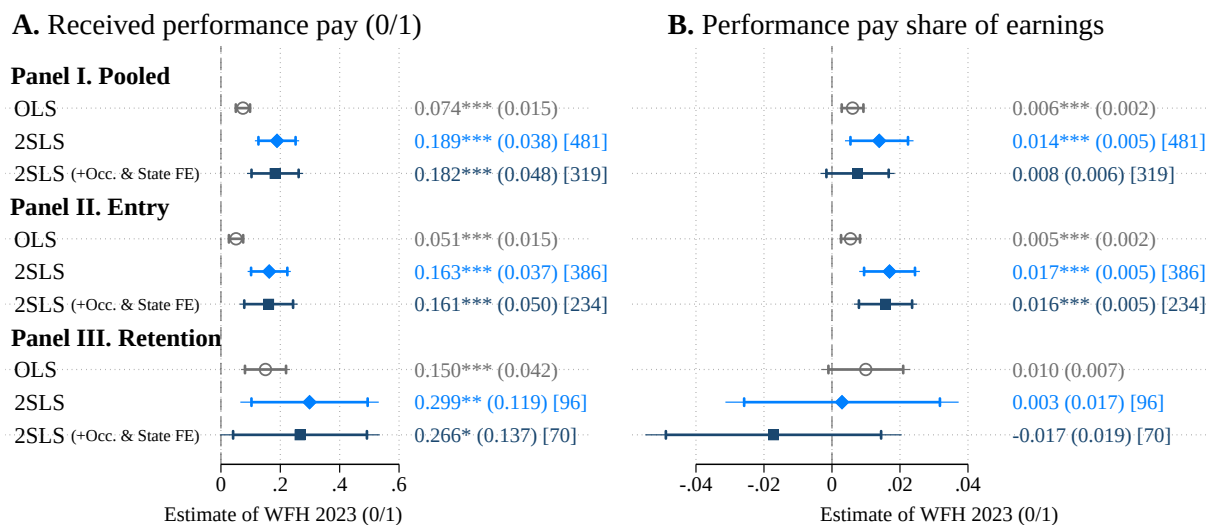
Notes: Panel A shows WFH adoption rates (at least half of working days) by WFH feasibility, conditional on industry fixed effects. The sample comprises 947,453 dependent employees aged 15–64 in the *Mikrozensus* (the German component of the EU Labour Force Survey) across 2017–2023. Panel B shows event-study estimates of WFH feasibility (0/1) on realized WFH (0/1) using the [Callaway and Sant’Anna \(2021\)](#) estimator for repeated cross-sectional data, conditional on industry fixed effects. Shaded areas and spikes show 90% and 95% confidence intervals, respectively, using standard errors clustered at the occupation (3-digit KldB) level.

Effects on performance pay. OLS and IV estimates consistently show that WFH adoption increases the incidence of performance pay. 2SLS implies that WFH raises the individual likelihood of receiving PP by 19 percentage points in the baseline pooled sample (Panel I), which corresponds to a 79% increase relative to the 2019 treated mean of 24%. The corresponding extensive-margin effect is 16 percentage points on the entry margin (Panel II) and 30 percentage points on the retention margin (Panel III). WFH increases the PP share of annual earnings by 1.4 percentage points pooled (70% of the 2.0% treated mean) and 1.7 percentage points on the entry margin. By contrast, the effect is null on the retention margin, suggesting that WFH retains workers on PP contracts but does not significantly raise its intensity among those already receiving it. Adding occupation-segment and federal-state fixed effects slightly attenuates the estimates: while the positive effects on PP incidence and the entry-margin PP share remain significant, the pooled PP-share estimate loses statistical significance, which reinforces our conclusion that the shift in compensation structure operates predominantly through entry into new performance-pay arrangements.

Contract change versus target-hitting. A key question is whether these effects reflect changes in compensation contracts or instead a higher chance of meeting performance targets (under existing incentive contracts) when working remotely. Because the survey records only the *receipt* of performance pay, we cannot separate the two a

priori. In [Appendix D](#), we discuss several pieces of evidence corroborating contract change as the underlying mechanism: *i*) the entry effect persists among workers with no performance pay in 2017–2019, i.e., where an incentive contract is unlikely to exist; *ii*) incumbents, who are on PP in both 2019 and 2023, see no rise in their PP share, contrary to better target-hitting; and *iii*) WFH exposure leaves the PP exit hazard unchanged, suggesting no post-shock improvement in target-hitting.

Figure 5: The effect of WFH 2023 (0/1) on 2019–2023 individual changes in performance pay



Notes: The figure shows OLS and 2SLS estimates of the effect of WFH at least once per week in 2023 (0/1) on the within-individual 2019–2023 change in receipt of performance pay (Panel A) and in the PP share of annual earnings (Panel B); see Equations 1 and 2. 2SLS specifications instrument 2023 WFH status with 2019 WFH feasibility. Panel I uses the pooled sample and controls for baseline PP status ($N = 3,443$); Panel II restricts to workers without PP in 2019 (entry margin, $N = 2,813$); Panel III restricts to workers with PP in 2019 (retention margin, $N = 630$). All specifications are conditional on industry fixed effects. The third row of each panel adds occupation-segment and federal-state fixed effects. Standard errors clustered by individual in parentheses. Brackets report the Kleibergen–Paap first-stage F -statistic. Bars show 90% and 95% confidence intervals. [Table B.2](#) reports the corresponding regression table, including baseline means of outcome variables. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Robustness. We probe the estimates against a battery of robustness checks in [Appendix F](#). The results are stable across additional baseline demographic and job controls, within-occupation and within-state variation, trimming workers who were likely already remote before 2020, and alternative clustering. All results also hold under wild-cluster-bootstrap inference (to account for potential finite-sample distortions of 2SLS) and a doubly-robust estimator that does not rely on the linear covariate adjustment of 2SLS ([Śloczyński et al., 2022](#)). We also show that no single occupation, industry or state drives the effect.

4.3 Selection on gains and the OLS–IV gap

The 2SLS estimates of the effect of WFH on performance pay in [Figure 5](#) exceed OLS by a factor of about 2.5. What explains this gap? One possibility is selection into WFH. If workers who adopt WFH would have experienced slower growth in performance pay absent the WFH shock, OLS understates the causal effect of WFH on performance pay. A second possibility is treatment-effect heterogeneity. Whereas OLS averages over all treated workers, 2SLS averages over compliers with WFH-feasible 2019 jobs. If compliers experience larger gains than the average treated worker, the LATET can exceed OLS.

We probe these channels with three exercises: a complier analysis, an estimation of the marginal treatment effect (MTE) curve, and heterogeneity analyses. Together, they paint a coherent picture. Never-takers, who dominate the OLS control group, are more career-oriented and would likely have experienced greater growth in performance pay through traditional career channels even absent the WFH shock, attenuating the OLS estimate. Compliers, by contrast, lie at the high-gain end of the treatment-effect distribution, with gains declining sharply in unobserved resistance to adopting WFH. Finally, the heterogeneity by subsamples suggests that effects are concentrated in groups that plausibly have greater willingness or capability to adopt WFH-for-PP packages.

Complier analysis. We profile compliers (workers induced into WFH by feasibility) relative to never-takers (who do not WFH despite high feasibility) and always-takers (who WFH despite low feasibility) using the approach of [Marbach and Hangartner \(2020\)](#); [Hangartner et al. \(2021\)](#). The detailed analysis in [Appendix E](#) shows that never-takers have higher baseline earnings, are less likely to be covered by collective wage agreements, have longer firm tenure, more often hold supervisory positions, and work more overtime. These characteristics are consistent with stronger career orientation, making them less likely to adopt WFH, possibly because of promotion concerns or greater complementarity between their tasks and office work ([Castex and Dechter, 2026](#)). It is plausible that these workers systematically experience steeper PP trajectories even absent WFH through traditional promotions ([Prendergast, 1999](#); [Lazear and Rosen, 1981](#)). Because never-takers dominate the OLS control group, this attenuates the OLS estimate. Always-takers, who account for only about 8% of the sample, have characteristics associated with a high private value of WFH: younger, more likely to have children at home, and have longer commutes. But they are also plausibly less likely to sort into performance-pay contracts. Their inclusion in the OLS treatment group thus tends to pull the coefficient down as well.

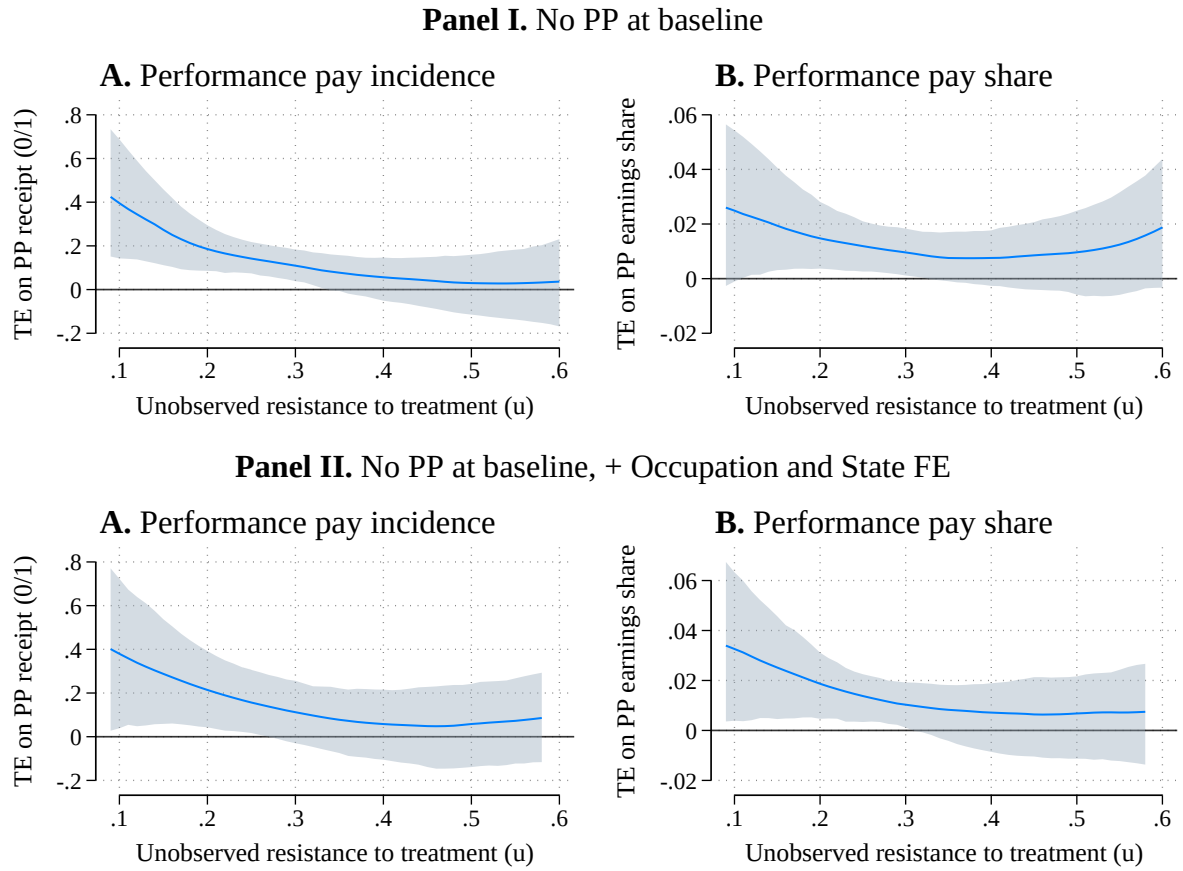
Marginal treatment effects (MTE). The complier analysis characterizes differences along *observables*, whereas the MTE framework characterizes heterogeneity in unobservables (Cornelissen et al., 2016). Workers are ranked by their unobserved resistance to WFH adoption (u): those with low resistance are easily induced into WFH by higher job feasibility, whereas those with high resistance remain unwilling to adopt even when feasibility is high. The MTE at a given value of u is the average treatment effect for workers whose feasibility level just induces WFH adoption.

We estimate the MTE semi-parametrically following Andresen (2018): the propensity score is estimated by logit and the MTE is fitted by a local polynomial of degree two, with confidence bands from 500 bootstrap replications. To ensure adequate overlap and common support, we trim the extreme 5% tails of the estimated propensity-score distribution. Consequently, the MTE is empirically identified for workers with unobserved resistance to WFH (u) between approximately 0.1 and 0.6.

Figure 6 plots the MTE curve specifically for workers without performance pay at baseline (the entry margin). The curve for performance-pay incidence (Panel I.A) slopes steeply downward, indicating positive selection on gains: workers with the lowest resistance experience the largest increases in PP receipt, whereas the effect declines markedly for workers with higher resistance. The same downward slope appears for the PP share of annual earnings (Panel I.B), and both patterns are robust to adding occupation-segment and federal-state fixed effects (Panel II). This suggests that the WFH-induced shift to performance pay is concentrated in employment relationships where WFH-for-PP packages are more easily implemented or accepted. Or from the firm’s perspective: WFH is more readily adopted where firms can complement it with the supporting incentive scheme. Importantly, this downward trajectory implies that incrementally extending WFH to more resistant workers—at least within our identified support—would yield substantially smaller average effects on compensation structure than those recovered by the LATET. Figure A.2 reports the corresponding curves for the pooled sample. Incidence (Panel A) shows the same downward slope. The MTE on the PP share of earnings (Panel B) shows no clear pattern and is imprecisely estimated, consistent with the 2SLS estimates showing a null effect on PP intensity among workers already on PP in 2019, and again suggesting that the shift operates mostly through entry into PP.

Effect heterogeneity by group. 2SLS estimates of the WFH effect on performance pay by subgroup further support the selection-on-gains interpretation (Panels A and B of Figure 7). The first row reproduces the baseline estimates for the pooled sample. Re-

Figure 6: Marginal treatment effect of WFH in 2023 (0/1) on within-individual changes in performance pay, entry margin



Notes: The figure presents semi-parametric estimates of the marginal treatment effect (MTE) of WFH at least once per week in 2023 (0/1) on the receipt of performance pay (Panel A) and on the PP share of annual earnings (Panel B), as a function of the unobserved component of resistance to treatment u . The sample is restricted to workers without performance pay in 2019 (entry margin). Panel I is conditional on industry fixed effects; Panel II adds occupation-segment and federal-state fixed effects. Lower values of u correspond to workers more easily induced into WFH by higher 2019 WFH feasibility. The propensity score is estimated by logit and the MTE is fitted by a local polynomial of degree two. Shaded areas show 95% confidence bands from 500 bootstrap replications. The horizontal axis spans the trimmed region of common propensity-score support, $u \in [0.09, 0.58]$.

stricting the sample to workers who do not change employer leaves the estimates essentially unchanged, indicating that the effects arise primarily through contract renegotiation within existing employment relationships rather than job switching.

The remaining rows show that the effects are concentrated among men and workers whose compensation is not covered by collective bargaining agreements (CBAs), both for the probability of receiving performance pay (Panel A) and its share in earnings (Panel B). These patterns are consistent with evidence that women are less willing to accept incentive pay, partly because of greater risk aversion ([Dohmen and Falk, 2011](#)), and may require weaker explicit incentives because they exert more effort when un-

observed (Segal, 2012). They are also consistent with employers outside CBAs having greater discretion to redesign compensation contracts. The magnitudes are economically meaningful. Among men and in non-CBA-covered jobs, WFH nearly doubles the probability of receiving performance pay. The results also suggest that an expansion of WFH to women or CBA-covered workers would likely cause changes far smaller than those recovered by the LATET.

4.4 Signaling and incentive substitution

A longstanding literature suggests that long or nonstandard working hours, including (uncompensated) overtime, are often used as signals of worker quality and effort to employers, which can induce an inefficiently high supply of hours (Anger, 2008; Frederiksen et al., 2024; Bohnet and Ciasullo, 2026; Landers et al., 1996; Bratti and Staffolani, 2007). Under such implicit contracts, workers are rewarded later through raises and promotions. The signaling value depends, of course, on the principal being able to observe the signal (Frederiksen et al., 2024).

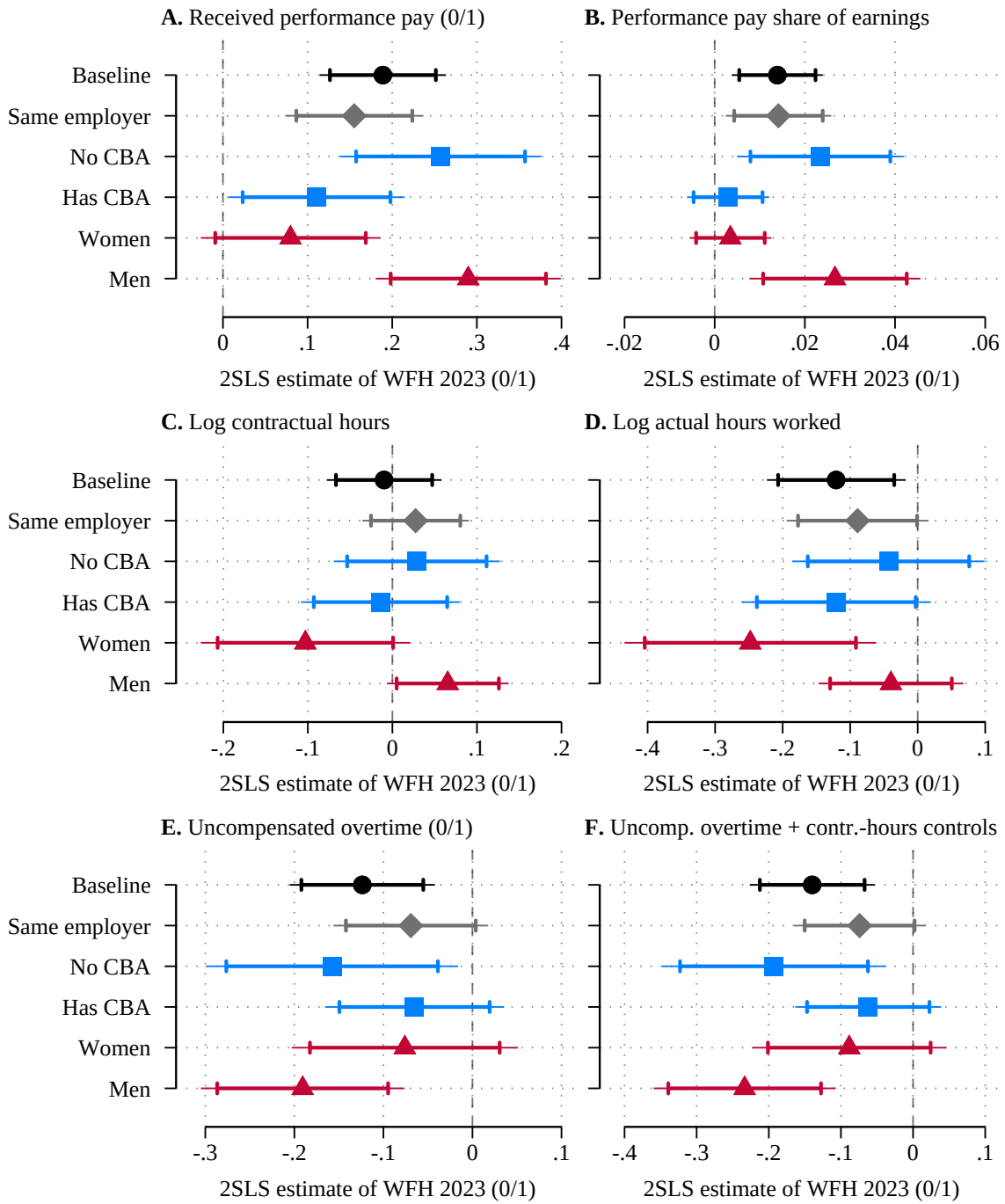
WFH is a shock to observability, given that the principal no longer observes presence or who stays late. The logic is well captured in a recent survey: one in five hybrid workers in Germany report coming to the office mainly “to make it visible that they are really working” (Kunze and Hampel, 2026). Thus, while the signaling motive of observable hours is not entirely eliminated in the hybrid work economy, its value has depreciated considerably. We consequently expect falling labor supply under WFH, all else equal.

Performance pay amplifies this mechanism. When WFH is packaged with output-based pay, effort is rewarded directly through bonuses. PP effectively replaces the implicit “hours-for-raises” contract by an explicit “bonus-for-output” one (Baker et al., 1994). This incentive substitution implies that the hours reduction should be larger where WFH is accompanied by a shift into performance pay.

Isolating the signaling motive from other reasons for supplying long hours, such as human-capital accumulation, is notoriously hard (Bohnet and Ciasullo, 2026). In addition, WFH bundles less observability with commute-time savings, autonomy, and flexibility, so that no single factor can be cleanly identified as the driver of changes in labor supply.

We provide suggestive evidence for the signaling and incentive substitution channel. First, besides the effect of WFH on changes in contractual and actual hours, we focus on the supply of *uncompensated* overtime, which is a margin closely tied to the signal-

Figure 7: DiD-IV estimates of WFH on performance pay, hours, and overtime by subgroup



Notes: The figure shows 2SLS estimates of the effect of WFH at least once per week in 2023 on within-individual 2019–2023 changes (Equations 1 and 2). Outcomes are receipt of performance pay (0/1) (Panel A), the PP share of annual earnings (Panel B), log contractual hours (Panel C), log actual hours worked (Panel D), and an indicator for working uncompensated overtime, unconditionally (Panel E) and controlling for 2019 baseline contractual hours and the 2019–2023 change in contractual hours (Panel F). Each row is a separate regression on the indicated subgroup, conditional on industry fixed effects and baseline PP status. *Baseline* uses the baseline estimation sample. *Same employer* restricts to workers who do not change employer between 2019 and 2023; *No CBA / Has CBA* split the sample by 2019 collective-bargaining-agreement coverage. *Women / Men* split by gender. Bars show 90% and 95% confidence intervals from standard errors clustered by individual. The corresponding regression table, including first-stage diagnostics and baseline outcome means, is reported in [Table B.3](#).

ing motive in the office era. Second, we condition on 2019 contractual hours and their 2019–2023 change, to account for alternative labor-supply channels in the overtime specification.

Panels C–F of [Figure 7](#) report the 2SLS estimates of WFH on hours and overtime by subgroup. Contractual hours do not change on average (Panel C); but the null hides a gender asymmetry: WFH cuts women’s contractual hours by 10.3 log points while raising men’s slightly, by 6.5 log points. Actual hours fall on average, and by more than contractual hours across all groups (Panel D). To the extent that the labor-supply response to saved commuting and greater flexibility is captured in contractual hours, this larger drop in *actual* hours points to adjustments beyond flexibility and time-savings alone.

The effects on uncompensated overtime corroborate this interpretation (Panel E). Unpaid overtime falls on average and for all subgroups. Remarkably, the decline mirrors the shift into performance pay. WFH reduces overtime most strongly for men and workers outside collective bargaining, which are the groups with the largest WFH-induced moves into PP (Panels A and B). Women and CBA-covered workers show smaller changes in PP and overtime. Conditioning on 2019 contractual hours and their 2019–2023 change leaves the pattern unchanged (Panel F). This suggestively shows that the shift to PP channels some of the reduction in uncompensated overtime, consistent with incentive substitution.

We also probe whether performance pay mediates the effect of WFH on labor supply with a more formal decomposition, using the single-instrument mediation estimator of [Dippel et al. \(2020, 2017\)](#).¹³ The approach decomposes the effect of an endogenous treatment (WFH) on an outcome (uncompensated overtime/hours worked) into an indirect component operating through an endogenous mediator (performance pay) and a residual direct component. Identification adds one restriction to the standard IV assumptions: the unobserved determinants of moving into performance pay must be independent of those of labor supply. This restriction may fail if a common factor, such as career ambition or employer type, affects changes in both. Because a single instrument cannot test this assumption, we interpret the exercise as suggestive. The indirect effect through performance pay is negative and significant for both uncompensated overtime (−0.26, SE 0.11) and log actual hours worked (−0.23, SE 0.11), consistent with incentive substitution. The instrument strongly predicts WFH adoption (first-stage $F \approx 447$) and predicts the shift into performance pay ($F \approx 11$). The

¹³We use Stata’s `ivmediate` package for the decomposition ([Dippel et al., 2020](#)).

direct effect is positive for both outcomes, but its magnitude cannot be meaningfully interpreted; specifically, [Dippel et al. \(2020\)](#) show in simulations that direct effects are recovered reliably only with a substantially stronger mediator first stage. As discussed above, a positive sign is not implausible, as time saved from commuting may be reallocated to work, including overtime ([Arntz et al., 2022](#)). Overall, the decomposition corroborates the interpretation that the shift from implicit (input-based) toward explicit (output-based) incentives accounts for part of the decline in labor supply following WFH adoption.

5 Conclusion and outlook

Working from home weakens principals’ monitoring abilities and grants employees autonomy over their effort and time allocation across professional (and private) tasks. In their seminal article, [Holmstrom and Milgrom \(1991\)](#) predict that such arrangements should be accompanied by a stronger reliance on performance-based incentives (p. 27). The post-2020 shift to remote work provides a large-scale natural experiment to revisit this prediction. Exploiting differences in exposure to the WFH shock through individual jobs’ compatibility with WFH, we find broad support for [Holmstrom and Milgrom’s](#) hypothesis. WFH adoption raises the incidence of performance pay and increases its importance in total earnings.

The shift in compensation structure comes with a reduction in labor supply, including uncompensated overtime, consistent with the reduced signal value of long hours when they are not observed. The reduction is partly mediated by the WFH-induced shift to performance pay, in line with explicit output-based incentives substituting for the implicit reward of overtime signaling for promotions. We document that WFH adoption follows a selection-on-gains pattern, where WFH is more readily adopted when paired with performance pay.

WFH is here to stay, and so is the greater role of performance-based pay. Selection into WFH by those with the largest gains in performance pay may help reconcile the mixed evidence on its performance effects, which largely comes from case studies and field experiments. If WFH depends on complementary investments in incentive schemes, organizational design, management practices, human capital, or technology, its ef-

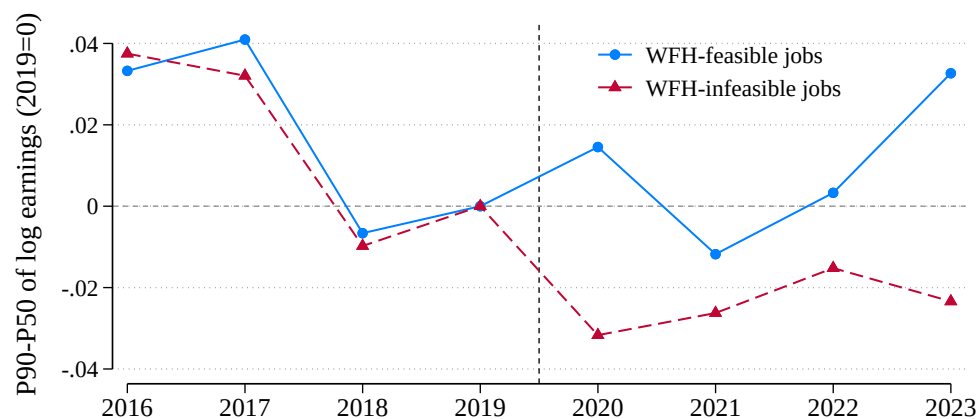
fects will vary with firms' ability (or willingness) to make such investments.¹⁴ Recent work does point to several such complements, including coordinated face-to-face time between seniors and juniors, periodic in-person anchor days, or remote monitoring (Emanuel et al., 2026; Aksoy et al., 2026; Ko and Baek, 2025). However, the evidence remains too limited to draw general conclusions about *when* WFH actually works, especially given the scale of WFH in the economy. This paper provides a starting point.

The shift also raises several questions about distributional consequences and policy implications. One concerns the optimal design of the income tax: when earnings risk is generated endogenously by firms' incentive contracts rather than imposed by exogenous shocks, progressive taxation can lose its insurance function as firms re-optimize their contracts to re-align incentives (Doligalski et al., 2023). The welfare cost of ignoring this channel scales with the share of workers paid for performance. Thus, the structural shift to PP caused by the Big Shift to remote work presents a case for revisiting the progressivity of the income tax.

In addition, WFH feasibility and PP exposure are generally more prevalent among workers with higher baseline earnings. Whether WFH-induced PP translates into greater overall earnings inequality is an empirical question, and the available cross-country evidence is mixed. For the United States, Lemieux et al. (2009) attribute essentially all of the post-1970s growth in upper-tail wage inequality to the spread of PP, while the lower-tail gap grew almost independently of it. In Germany, Sommerfeld (2013) finds that PP has done little to widen the overall wage distribution, but has increased wage dispersion among PP recipients. Figure 8 previews our own descriptive evidence: the upper-tail (P90–P50) earnings gap inside WFH-feasible jobs (solid line) has diverged sharply from that inside infeasible jobs after 2019. This aggregate divergence is consistent with the hypothesis that the WFH-induced shift to PP may be exacerbating upper-tail inequality. We take this observation as a starting point for a discussion of the distributional consequences of WFH, and leave a rigorous causal decomposition to future work.

¹⁴Productivity effects of WFH appear mixed and highly context-dependent (Mang and Anwar, 2026). Estimates range from strongly negative in fully remote settings (Emanuel and Harrington, 2024; Atkin et al., 2023; Gibbs et al., 2023; Künn et al., 2022), to roughly neutral under hybrid schedules (Bloom et al., 2024; Choudhury et al., 2024; Angelici and Profeta, 2023), to positive, particularly where work is routine, as in call centers (Bloom et al., 2015; Aksoy et al., 2025b; Fenizia and Kirchmaier, 2025).

Figure 8: Upper-tail earnings dispersion (P90–P50 gap) by WFH feasibility



Notes: The figure shows the year-by-year evolution of the P90–P50 gap (the difference between the 90th and the 50th percentile, in log points) of log annual labor earnings, plotted separately for workers in WFH-feasible (solid blue) and WFH-infeasible (dashed red) jobs and expressed as the deviation from each series' 2019 level. The sample comprises employed individuals aged 16–64. Data are from the SOEP.

References

- Ahrens, A., Hansen, C. B., and Schaffer, M. E. (2020). LassoPack: Model selection and prediction with regularized regression in Stata. *The Stata Journal*, 20(1):176–235.
- Aksoy, C. G., Barrero, J. M., Bloom, N., Davis, S. J., Dolls, M., and Zarate, P. (2025a). The global persistence of work from home. *Proceedings of the National Academy of Sciences*, 122(27):e2509892122.
- Aksoy, C. G., Bloom, N., Davis, S. J., Marino, V., and Ozguzel, C. (2025b). Remote Work, Employee Mix, and Performance. NBER Working Paper 33851, National Bureau of Economic Research.
- Aksoy, C. G., Bloom, N., Davis, S. J., Marino, V., and Ozguzel, C. (2026). The Value of One Office Day a Month. Working Paper 35331, National Bureau of Economic Research.
- Alipour, J.-V. (2025). Does Remote Work Reinforce Gender Gaps in (Un)Paid Labor? CESifo Working Paper 12052.
- Alipour, J.-V., Falck, O., Krause, S., Krolage, C., and Wichert, S. (2025). Working from Home and Consumption in Cities. CESifo Working Paper 10000.
- Alipour, J.-V., Falck, O., and Schüller, S. (2023). Germany’s capacity to work from home. *European Economic Review*, 151:104354.
- Andresen, M. E. (2018). Exploring Marginal Treatment Effects: Flexible Estimation Using Stata. *The Stata Journal: Promoting communications on statistics and Stata*, 18(1):118–158.
- Angelici, M. and Profeta, P. (2023). Smart Working: Work Flexibility Without Constraints. *Management Science*.
- Anger, S. (2008). Overtime Work as a Signaling Device. *Scottish Journal of Political Economy*, 55(2):167–189.
- Arntz, M., Ben Yahmed, S., and Berlingieri, F. (2022). Working from home, hours worked and wages: Heterogeneity by gender and parenthood. *Labour Economics*, 76:102169.
- Atkin, D., Schoar, A., and Shinde, S. (2023). Working from Home, Worker Sorting and Development. Technical Report 31515.
- Bagga, S., Mann, L. F., Şahin, A., and Violante, G. L. (2025). Job Amenity Shocks and Labor Reallocation. Working Paper 33787, National Bureau of Economic Research.
- Baker, G., Gibbons, R., and Murphy, K. J. (1994). Subjective Performance Measures in Optimal Incentive Contracts. *The Quarterly Journal of Economics*, 109(4):1125–1156.
- Baker, G. P. and Hall, B. J. (2004). CEO Incentives and Firm Size. *Journal of Labor Economics*, 22(4):767–798.

- Baker, G. P. and Hubbard, T. N. (2004). Contractibility and Asset Ownership: On-Board Computers and Governance in U. S. Trucking*. *The Quarterly Journal of Economics*, 119(4):1443–1479.
- Bandiera, O., Fischer, G., Prat, A., and Ytsma, E. (2021). Do Women Respond Less to Performance Pay? Building Evidence from Multiple Experiments. *American Economic Review: Insights*, 3(4):435–454.
- Barrero, J. M., Bloom, N., Davis, S., Foster, K., Meyer, B., and Mihaylov, E. (2025). US Executives Predict Work from Home is Here to Stay. SIEPR Policy Brief 3-2025, Stanford.
- Barrero, J. M., Bloom, N., and Davis, S. J. (2023). The Evolution of Work from Home. *Journal of Economic Perspectives*, 37(4):23–49.
- Barrero, J. M., Bloom, N., and Davis, S. J. (2026). Why Working From Home Will Stick. Technical Report 174.
- Barrero, J. M., Bloom, N., Davis, S. J., Meyer, B. H., and Mihaylov, E. (2022). The Shift to Remote Work Lessens Wage-Growth Pressures. Working Paper 30197, National Bureau of Economic Research.
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does Management Matter? Evidence from India. *The Quarterly Journal of Economics*, 128(1):1–51.
- Bloom, N., Han, R., and Liang, J. (2024). Hybrid working from home improves retention without damaging performance. *Nature*, pages 1–6.
- Bloom, N., Liang, J., Roberts, J., and Ying, Z. J. (2015). Does Working from Home Work? Evidence from a Chinese Experiment. *The Quarterly Journal of Economics*, 130(1):165–218.
- Bloom, N. and Van Reenen, J. (2007). Measuring and Explaining Management Practices Across Firms and Countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.
- Bohnet, L. and Ciasullo, L. (2026). Signaling within the firm. Technical Report.
- Boly, A. (2011). On the incentive effects of monitoring: Evidence from the lab and the field. *Experimental Economics*, 14(2):241–253.
- Bratti, M. and Staffolani, S. (2007). Effort-based career opportunities and working time. *International Journal of Manpower*, 28(6):489–512.
- Callaway, B. and Sant’Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.
- Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90(3):414–427.
- Castex, G. and Dechter, E. (2026). Working From Home and Wage Dynamics. *Journal of Human Resources*.

- Chen, Y., Cortes, P., Kosar, G., Pan, J., and Zafar, B. (2023). The Impact of COVID-19 on Workers' Expectations and Preferences for Remote Work. *AEA Papers and Proceedings*, 113:556–561.
- Choi, S., Davis, S. J., Jaramillo, O., and Villena-Roldan, B. (2026). Remote Work, Competition, and Wages: Disentangling Demand and Supply Influences. Working Paper.
- Choudhury, P., Khanna, T., Makridis, C. A., and Schirmann, K. (2024). Is Hybrid Work the Best of Both Worlds? Evidence from a Field Experiment. *The Review of Economics and Statistics*, pages 1–24.
- Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2016). From LATE to MTE: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41:47–60.
- Cuñat, V. and Guadalupe, M. (2005). How Does Product Market Competition Shape Incentive Contracts? *Journal of the European Economic Association*, 3(5):1058–1082.
- Cuñat, V. and Guadalupe, M. (2009). Globalization and the Provision of Incentives inside the Firm: The Effect of Foreign Competition. *Journal of Labor Economics*, 27(2):179–212.
- Davis, S. J., Bloom, N., and Codreanu, M. A. (2026). Demand-Driven Technical Change: Evidence from WFH Technologies. NBER Working Paper 35083, National Bureau of Economic Research.
- de Chaisemartin, C. (2010). A note on instrumented difference in differences. Unpublished.
- de Chaisemartin, C. and D'Haultfœuille, X. (2018). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, 85(2):999–1028.
- De Fraja, G., Matheson, J., Mizen, P., Rockey, J., Taneja, S., and Thwaites, G. (2024). Remote Work and Compensation Inequality. SSRN Scholarly Paper 4962603, Social Science Research Network, Rochester, NY.
- Dickinson, D. and Villego, M.-C. (2008). Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic Behavior*, 63(1):56–76.
- Dingel, J. I. and Neiman, B. (2020). How Many Jobs can be Done at Home? *Journal of Public Economics*, 189:104235.
- Dippel, C., Ferrara, A., and Heblich, S. (2020). Causal mediation analysis in instrumental-variables regressions. *The Stata Journal*, 20(3):613–626.
- Dippel, C., Gold, R., Heblich, S., and Pinto, R. (2017). Instrumental Variables and Causal Mechanisms: Unpacking the Effect of Trade on Workers and Voters. CESifo Working Paper 6816.
- Dohmen, T. and Falk, A. (2011). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *The American Economic Review*, 101(2):556–590.

- Doligalski, P., Ndiaye, A., and Werquin, N. (2023). Redistribution with Performance Pay. *Journal of Political Economy Macroeconomics*, 1(2).
- Drake, M., Thakral, N., and Tô, L. T. (2025). Wage Differentials and the Price of Workplace Flexibility.
- Emanuel, N. and Harrington, E. (2024). Working Remotely? Selection, Treatment, and the Market for Remote Work. *American Economic Journal: Applied Economics*, 16(4):528–59.
- Emanuel, N., Harrington, E., and Pallais, A. (2026). The Power of Proximity to Coworkers. *The Quarterly Journal of Economics*, 141(3):1825–1870.
- Falk, A. and Kosfeld, M. (2006). The Hidden Costs of Control. *American Economic Review*, 96(5):1611–1630.
- Fenizia, A. and Kirchmaier, T. (2025). Not incentivized yet efficient: Working from home in the public sector. Technical Report 2036, LSE.
- Frederiksen, A., Kato, T., and Smith, N. (2024). Working Hours, Top Management Appointments, and Gender: Evidence from Linked Employer-Employee Data. *Journal of Labor Economics*, page 733490.
- Garen, J. E. (1994). Executive Compensation and Principal-Agent Theory. *Journal of Political Economy*, 102(6):1175–1199.
- Gayle, G.-L. and Miller, R. A. (2009). Has Moral Hazard Become a More Important Factor in Managerial Compensation? *The American Economic Review*, 99(5):1740–1769.
- Gibbs, M., Mengel, F., and Siemroth, C. (2023). Work from Home and Productivity: Evidence from Personnel and Analytics Data on Information Technology Professionals. *Journal of Political Economy Microeconomics*, 1(1):7–41.
- Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik*, 239(2):345–360.
- Grossman, S. J. and Hart, O. D. (1983). An Analysis of the Principal-Agent Problem. *Econometrica*, 51(1):7–45.
- Hall, A., Hünefeld, L., and Rohrbach-Schmidt, D. (2020). BIBB/BAuA-Erwerbstätigenbefragung 2018 - Arbeit und Beruf im Wandel.
- Hangartner, D., Marbach, M., Henckel, L., Maathuis, M. H., Kelz, R. R., and Keele, L. (2021). Profiling Compliers in Instrumental Variables Designs. Technical Report arXiv:2103.06328, arXiv.
- Harrington, E. and Kahn, M. E. (2025). Has the Rise of Work from Home Reduced the Motherhood Penalty in the Labor Market? Working Paper 34147, National Bureau of Economic Research.

- Hartzell, J. C., Parsons, C. A., and Yermack, D. L. (2010). Is a Higher Calling Enough? Incentive Compensation in the Church. *Journal of Labor Economics*, 28(3):509–539.
- Holmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, 10(1):74.
- Holmstrom, B. and Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, 7:24–52.
- Houeix, D. (2026). Asymmetric Information and Digital Technology Adoption: Evidence from Senegal. Technical report.
- Kelley, E. M., Lane, G., and Schönholzer, D. (2024). Monitoring in Small Firms: Experimental Evidence from Kenyan Public Transit. *American Economic Review*, 114(10):3119–3160.
- Ko, Y. H. and Baek, I. G. (2025). The Effect of Computer Monitoring on Employees’ Productivity in Telecommuting Arrangements. *Management Science*, 71(1):103–122.
- Künn, S., Seel, C., and Zegners, D. (2022). Cognitive Performance in Remote Work: Evidence from Professional Chess. *The Economic Journal*, 132(643):1218–1232.
- Kunze, F. and Hampel, K. (2026). Homeoffice in der Energiekrise: Pendeln, Effizienz und hybrides Arbeiten in Deutschland 2026. Konstanzer Homeoffice-Studie Ergebnisreport April 2026, University of Konstanz.
- Lamorgese, A., Patnaik, M., Linarello, A., and Schivardi, F. (2024). Management Practices and Resilience to Shocks: Evidence from COVID-19. *Management Science (INFORMS)*, 70(12):9058–9072.
- Landers, R. M., Rebitzer, J. B., and Taylor, L. J. (1996). Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms. *The American Economic Review*, 86(3):329–348.
- Lazear, E. P. (2018). Compensation and Incentives in the Workplace. *Journal of Economic Perspectives*, 32(3):195–214.
- Lazear, E. P. and Rosen, S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5):841–864.
- Lemieux, T., MacLeod, W. B., and Parent, D. (2009). Performance Pay and Wage Inequality. *The Quarterly Journal of Economics*, 124(1):1–49.
- Luchman, J. N. (2021). Determining relative importance in Stata using dominance analysis: Domin and domme. *The Stata Journal*, 21(2):510–538.
- Maestas, N., Mullen, K. J., Powell, D., von Wachter, T., and Wenger, J. B. (2023). The Value of Working Conditions in the United States and the Implications for the Structure of Wages. *American Economic Review*, 113(7):2007–2047.
- Mang, C. F. and Anwar, A. (2026). Does Working from Home Improve Employees’ Productivity? Empirical Evidence from a Meta-Analysis. *ILR Review*, page 00197939261456260.

- Marbach, M. and Hangartner, D. (2020). Profiling Compliers and Noncompliers for Instrumental-Variable Analysis. *Political Analysis*, 28(3):435–444.
- Mas, A. and Pallais, A. (2017). Valuing Alternative Work Arrangements. *American Economic Review*, 107(12):3722–3759.
- Matthes, B., Meinken, H., and Neuhauser, P. (2015). Berufssektoren und Berufssegmente auf Grundlage der KldB 2010. Methodenbericht Der Statistik Der BA April 2015, Bundesagentur für Arbeit, Nuremberg.
- Mengistae, T. and Colin Xu, L. (2004). Agency Theory and Executive Compensation: The Case of Chinese State-Owned Enterprises. *Journal of Labor Economics*, 22(3):615–637.
- Miyaji, S. (2025). Instrumented Difference-in-Differences with Heterogeneous Treatment Effects. Technical Report arXiv:2405.12083, arXiv.
- Monte, F., Porcher, C., and Rossi-Hansberg, E. (2023). Remote Work and City Structure. Technical report.
- Nagin, D. S., Rebitzer, J. B., Sanders, S., and Taylor, L. J. (2002). Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment. *American Economic Review*, 92(4):850–873.
- Nagler, M., Rincke, J., and Winkler, E. (2024). Working from home, commuting, and gender. *Journal of Population Economics*, 37(3):58.
- Pierce, L., Snow, D. C., and McAfee, A. (2015). Cleaning House: The Impact of Information Technology Monitoring on Employee Theft and Productivity. *Management Science*, 61(10):2299–2319.
- Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature*, 37(1):7–63.
- Prendergast, C. (2002). The Tenuous Trade-off between Risk and Incentives. *Journal of Political Economy*, 110(5):1071–1102.
- Rambachan, A. and Roth, J. (2023). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*, page rdad018.
- Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights*, 4(3):305–322.
- Schüller, S. (2026). Estimating the effect of working from home on parents’ division of child-care and housework: A new panel IV approach. *Labour Economics*, 99:102866.
- Segal, C. (2012). Working When No One Is Watching: Motivation, Test Scores, and Economic Success. *Management Science*, 58(8):1438–1457.
- Shen, L. (2023). Does working from home work? A natural experiment from lockdowns. *European Economic Review*, 151:104323.

- Słoczyński, T., Uysal, S. D., and Wooldridge, J. M. (2022). Doubly Robust Estimation of Local Average Treatment Effects Using Inverse Probability Weighted Regression Adjustment.
- Sommerfeld, K. (2013). Higher and higher? Performance pay and wage inequality in Germany. *Applied Economics*, 45(30):4236–4247.
- Young, A. (2022). Consistency without Inference: Instrumental Variables in Practical Application. *European Economic Review*, 147:104112.

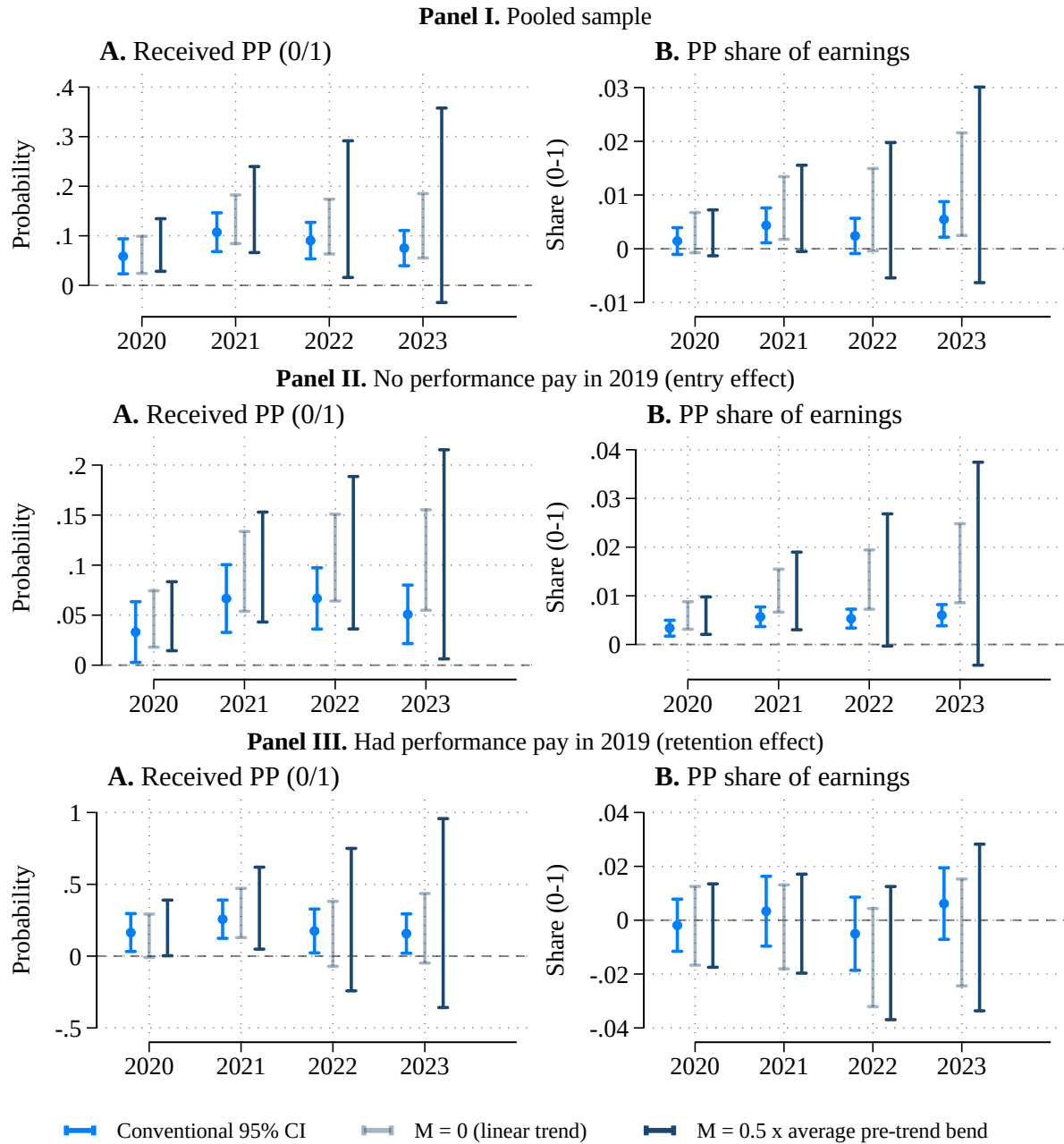
Appendix

Performance pay in the hybrid work economy

by *Jean-Victor Alipour*

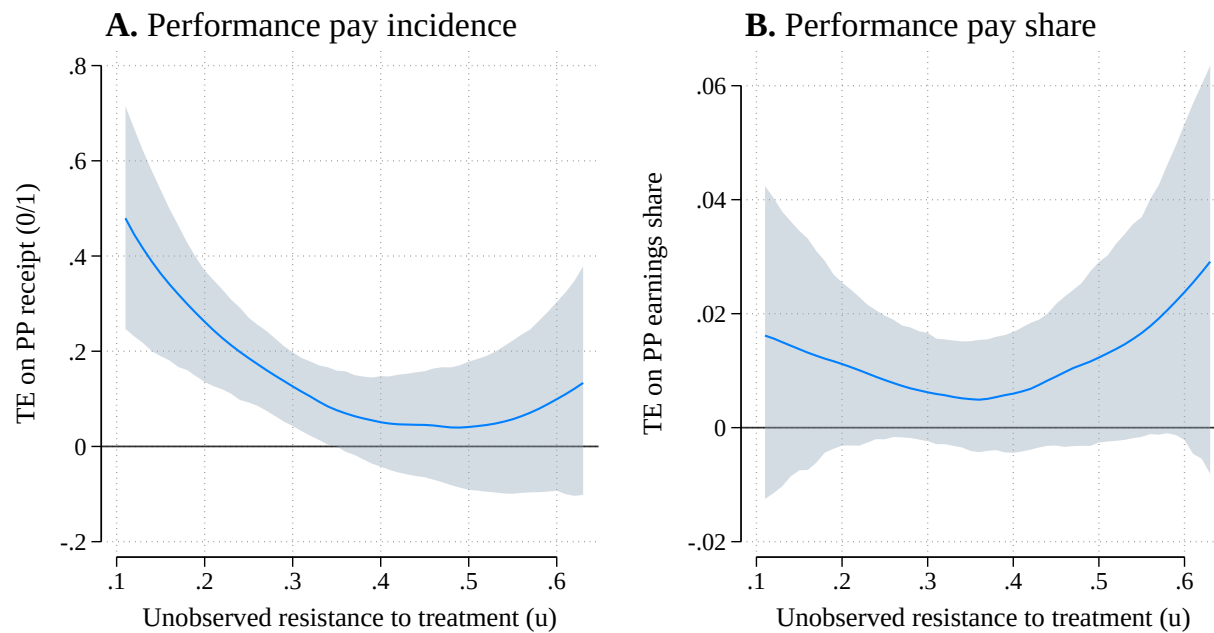
Appendix A Figures

Figure A.1: Honest-DiD confidence sets for event-study estimates under relaxations of parallel trends



Notes: The figure shows 95% robust confidence sets for the effect of 2019 WFH feasibility on performance pay in each post-treatment year, following [Rambachan and Roth \(2023\)](#). Event-study coefficients come from the [Callaway and Sant’Anna \(2021\)](#) estimator, conditional on industry fixed effects (see [Figure 3](#)). Outcomes are receipt of performance pay (0/1) (Panel A) and the PP share of annual earnings (Panel B). Panel I is the pooled estimation sample; Panel II restricts to workers without performance pay in 2019 (entry margin); Panel III to workers with performance pay in 2019 (retention margin). Instead of requiring parallel trends to hold exactly, the method bounds by M how much the slope of the difference in trends may change between consecutive periods: $M = 0$ permits linear violations. $M > 0$ permits deviations from linearity. The average pre-trend bend is the average year-to-year change in slope over 2016–2018. Because a deviation in one period shifts the counterfactual path in all subsequent periods, the confidence sets widen from 2020 to 2023. The conventional 95% confidence interval, using standard errors clustered by individual, is shown for comparison.

Figure A.2: Marginal treatment effect of WFH in 2023 (0/1) on within-individual changes in performance pay, pooled sample



Notes: The figure presents semi-parametric estimates of the marginal treatment effect (MTE) of WFH at least once per week in 2023 (0/1) on the receipt of performance pay (Panel A) and on the PP share of annual earnings (Panel B), as a function of the unobserved component of resistance to treatment u . The sample is the pooled estimation sample; both panels are conditional on industry fixed effects and baseline PP status. Lower values of u correspond to workers more easily induced into WFH by higher 2019 WFH feasibility. The propensity score is estimated by logit and the MTE is fitted by a local polynomial of degree two. Shaded areas show 95% confidence bands from 500 bootstrap replications. The horizontal axis spans the trimmed region of common propensity-score support, $u \in [0.11, 0.63]$. The corresponding estimates for the entry margin are reported in [Figure 6](#).

Appendix B Tables

Table B.1: First-stage estimates of 2019 WFH feasibility on 2023 WFH adoption

Sample	WFH at least 1x/week (0/1)	First-stage F	Observations (#Individuals)
Sector: Agriculture & natural resources (A, B)	0.505 (0.303)	2.8	46
Sector: Food, beverages, tobacco (C10–12)	0.344 (0.181)*	3.6	87
Sector: Machinery & transport equipment (C28–30)	0.758 (0.076)***	99.7	324
Sector: Other manufacturing (C13–27, C31–33)	0.683 (0.062)***	121.7	521
Sector: Other industry incl. construction (D, E, F)	0.645 (0.081)***	63.0	278
Sector: Trade, transport, tourism (G, H, I)	0.616 (0.061)***	102.7	904
Sector: Knowledge-intensive services (J, K, L, M)	0.444 (0.091)***	23.7	714
Sector: Other business & personal services (N, R, S)	0.645 (0.062)***	108.9	379
Sector: Public admin & education (O, P)	0.672 (0.052)***	165.6	959
Sector: Health, social & other non-market services (Q, T, U)	0.514 (0.066)***	61.1	878
Occupation: S11 Land/forestry/horticulture	1.420 (0.640)**	4.9	68
Occupation: S12 Manufacturing/crafts	0.923 (0.174)***	28.0	156
Occupation: S13 Engineering & manuf. tech	0.769 (0.088)***	76.9	448
Occupation: S14 Construction & building services	0.879 (0.146)***	36.5	186
Occupation: S21 Food & hospitality	0.763 (0.199)***	14.7	191
Occupation: S22 Health (medical & non-medical)	0.289 (0.136)**	4.5	493
Occupation: S23 Social & cultural services	0.961 (0.058)***	270.1	823
Occupation: S31 Trade & sales	0.813 (0.072)***	129.0	453
Occupation: S32 Management & organisation	0.914 (0.140)***	42.9	665
Occupation: S33 Business services (finance, law, media)	0.820 (0.127)***	41.4	746
Occupation: S41 IT & natural sciences	0.950 (0.104)***	82.6	297
Occupation: S51 Security	1.684 (0.594)***	8.0	110
Occupation: S52 Transport & logistics	1.304 (0.169)***	59.7	336
Occupation: S53 Cleaning	2.759 (2.136)	1.7	118
Educational attainment: Low (Hauptschule/none)	0.420 (0.074)***	32.2	699
Educational attainment: Mid (Realschule)	0.551 (0.036)***	234.3	1,607
Educational attainment: High (FH/Abitur)	0.643 (0.033)***	379.6	2,216
Age: 16-25	0.622 (0.115)***	29.5	265
Age: 26-30	0.730 (0.067)***	117.9	404
Age: 31-35	0.785 (0.062)***	162.1	415
Age: 36-40	0.763 (0.050)***	230.8	628
Age: 41-45	0.804 (0.048)***	276.0	765
Age: 46-50	0.746 (0.044)***	293.0	933
Age: 51-55	0.755 (0.041)***	335.3	915
Age: 56-60	0.506 (0.045)***	124.4	765
Firm size: <20	0.426 (0.046)***	87.4	923
Firm size: 20-199	0.572 (0.037)***	235.2	1,220
Firm size: 200-1,999	0.769 (0.037)***	423.9	1,166
Firm size: 2,000+	0.873 (0.031)***	787.9	1,710
Gender: Men	0.802 (0.027)***	895.1	2,287
Gender: Women	0.647 (0.026)***	643.0	2,803
Migration background: Native	0.702 (0.021)***	1151.9	4,080
Migration background: First/second generation	0.823 (0.046)***	319.3	1,010
Region: West Germany	0.741 (0.021)***	1206.5	3,853
Region: East Germany	0.519 (0.042)***	150.5	984
Performance pay 2019: No PP	0.690 (0.021)***	1034.4	4,036
Performance pay 2019: Had PP	0.739 (0.047)***	244.7	839
Collective bargaining 2019: No	0.720 (0.027)***	714.8	2,182
Collective bargaining 2019: Yes	0.736 (0.028)***	675.2	2,465

Notes: The table reports first-stage estimates of 2019 WFH feasibility on working from home at least once per week in 2023 (0/1) ([Equation 1](#)), separately for the indicated subsample (as of 2019). Standard errors clustered by individual in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table B.2: Estimates of WFH in 2023 on within-individual changes in performance pay

	Pooled			No PP 2019			Had PP 2019		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Panel A: Received performance pay (0/1)									
WFH 2023 (0/1)	0.074*** (0.015)	0.189*** (0.038)	0.182*** (0.048)	0.051*** (0.015)	0.163*** (0.037)	0.161*** (0.050)	0.150*** (0.042)	0.299** (0.119)	0.266* (0.137)
<i>2019 DV means:</i>									
Pooled sample	0.18	0.18	0.18	0.00	0.00	0.00	1.00	1.00	1.00
Treated (WFH_FEAS=1)	0.24	0.24	0.24	0.00	0.00	0.00	1.00	1.00	1.00
Control (WFH_FEAS=0)	0.12	0.12	0.12	0.00	0.00	0.00	1.00	1.00	1.00
Panel B: Performance pay share of earnings									
WFH 2023 (0/1)	0.006*** (0.002)	0.014*** (0.005)	0.008 (0.006)	0.005*** (0.002)	0.017*** (0.005)	0.016*** (0.005)	0.010 (0.007)	0.003 (0.017)	-0.017 (0.019)
<i>2019 DV means:</i>									
Pooled sample	0.01	0.01	0.01	0.00	0.00	0.00	0.06	0.06	0.06
Treated (WFH_FEAS=1)	0.02	0.02	0.02	0.00	0.00	0.00	0.07	0.07	0.07
Control (WFH_FEAS=0)	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.04	0.04
First-stage coef.		0.628	0.800		0.624	0.782		0.656	0.862
First-stage <i>F</i>		[481]	[319]		[386]	[234]		[96]	[70]
Individuals	3,443	3,443	3,443	2,813	2,813	2,813	630	630	630
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Occupation FE			✓			✓			✓
State FE			✓			✓			✓
Baseline PP	✓	✓	✓						

Notes: The table reports the estimates plotted in [Figure 5](#): OLS and 2SLS estimates of the effect of WFH at least once per week in 2023 (0/1) on within-individual 2019–2023 changes in receipt of performance pay (Panel A) and the PP share of annual earnings (Panel B), based on Equations 1 and 2. 2SLS specifications instrument 2023 WFH status with 2019 WFH feasibility. Columns 1–3 use the pooled sample and control for baseline PP status. Columns 4–6 restrict to workers without PP in 2019 (entry margin). Columns 7–9 restrict to workers with PP in 2019 (retention margin). Kleibergen–Paap first-stage *F*-statistics are shown in square brackets. Standard errors clustered by individual in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.3: DiD-IV estimates of WFH by subgroup

	(1) Baseline	(2) Same employer	(3) No CBA	(4) Has CBA	(5) Women	(6) Men
Panel A: Received performance pay (0/1)						
WFH 2023 (0/1)	0.189*** (0.038)	0.155*** (0.042)	0.257*** (0.061)	0.111** (0.053)	0.080 (0.054)	0.290*** (0.056)
2019 DV mean:						
Pooled sample	0.18	0.19	0.22	0.16	0.14	0.23
Treated (WFH_FEAS=1)	0.24	0.24	0.30	0.18	0.18	0.31
Control (WFH_FEAS=0)	0.12	0.13	0.11	0.14	0.09	0.15
Panel B: Performance pay share of earnings						
WFH 2023 (0/1)	0.014*** (0.005)	0.014** (0.006)	0.023** (0.009)	0.003 (0.005)	0.003 (0.005)	0.027*** (0.010)
2019 DV mean:						
Pooled sample	0.01	0.01	0.02	0.01	0.01	0.02
Treated (WFH_FEAS=1)	0.02	0.02	0.03	0.01	0.01	0.03
Control (WFH_FEAS=0)	0.00	0.00	0.01	0.00	0.00	0.01
Panel C: Log contractual hours						
WFH 2023 (0/1)	-0.010 (0.035)	0.027 (0.032)	0.029 (0.050)	-0.014 (0.048)	-0.103 (0.063)	0.065* (0.037)
2019 DV mean (level, hours):						
Pooled sample	33.6	33.8	33.9	33.7	30.1	37.8
Treated (WFH_FEAS=1)	34.3	34.4	34.9	34.1	31.7	37.9
Control (WFH_FEAS=0)	32.7	33.2	32.5	33.3	28.2	37.7
Panel D: Log actual hours worked						
WFH 2023 (0/1)	-0.121** (0.052)	-0.089* (0.054)	-0.043 (0.073)	-0.121* (0.072)	-0.248*** (0.095)	-0.040 (0.055)
2019 DV mean (level, hours):						
Pooled sample	36.8	37.1	37.5	36.9	32.7	41.6
Treated (WFH_FEAS=1)	38.3	38.5	39.3	37.8	34.9	42.7
Control (WFH_FEAS=0)	35.1	35.5	34.9	35.9	30.1	40.5
Panel E: Uncompensated overtime (0/1)						
WFH 2023 (0/1)	-0.124*** (0.042)	-0.069 (0.044)	-0.158** (0.072)	-0.065 (0.051)	-0.076 (0.065)	-0.191*** (0.058)
2019 DV mean:						
Pooled sample	0.11	0.11	0.17	0.07	0.08	0.14
Treated (WFH_FEAS=1)	0.17	0.16	0.23	0.11	0.12	0.23
Control (WFH_FEAS=0)	0.05	0.05	0.09	0.03	0.04	0.06
Panel F: Uncomp. overtime, incl. contractual-hours controls						
WFH 2023 (0/1)	-0.140*** (0.044)	-0.074 (0.046)	-0.193** (0.079)	-0.062 (0.052)	-0.088 (0.069)	-0.233*** (0.064)
2019 DV mean:						
Pooled sample	0.11	0.11	0.17	0.07	0.08	0.14
Treated (WFH_FEAS=1)	0.17	0.16	0.23	0.11	0.12	0.23
Control (WFH_FEAS=0)	0.05	0.05	0.09	0.03	0.04	0.06
First-stage coef.	0.628	0.658	0.620	0.644	0.549	0.709
First-stage F	[481]	[406]	[206]	[230]	[193]	[268]
Individuals	3,443	2,704	1,482	1,716	1,896	1,547
Industry FE	✓	✓	✓	✓	✓	✓
Baseline PP	✓	✓	✓	✓	✓	✓

Notes: The table reports 2SLS estimates of the effect of WFH at least once per week in 2023 (0/1) on the within-individual 2019–2023 changes (Equations 1 and 2). Outcomes are the receipt of performance pay (Panel A), the PP share of annual earnings (Panel B), log contractual hours (Panel C), log actual hours worked (Panel D), and an indicator for working uncompensated overtime, unconditionally (Panel E) and controlling for 2019 baseline contractual hours and the 2019–2023 change in contractual hours (Panel F). All regressions are run on the indicated sample and are conditional on industry fixed effects and baseline PP status. *Baseline* is the baseline estimation sample. *Same employer* restricts to workers who do not change employer between 2019 and 2023. *No CBA* / *Has CBA* split the sample by 2019 collective-bargaining-agreement coverage; *Women* / *Men* split by gender. Standard errors clustered by individual in parentheses. The corresponding coefficient plot is [Figure 7](#). *p<0.1; **p<0.05; ***p<0.01.

Appendix C Construction of the WFH feasibility measure

We capture exposure to the WFH shock of individuals in the SOEP dataset by imputing the WFH feasibility of their 2019 job, following the approach of [Alipour \(2025\)](#). The first step is to train a simple classifier model to predict job-level WFH feasibility from pre-pandemic job and worker characteristics in a separate dataset, the 2018 wave of the BIBB-BAuA Employment Survey ([Hall et al., 2020](#)).¹⁵ The survey elicits working conditions and job features from a representative sample of the German workforce in 2018. We exclude marginal workers (“Minijobs”), military personnel, and freelancers, as well as observations with missing occupation or industry codes, leaving a training sample of 16,790 individuals. In the second step, the trained model is applied to the 2019 cross section of the SOEP sample. Using corresponding characteristics, it predicts a continuous WFH feasibility score $\in [0, 1]$ and a binary indicator at the optimal classification cutoff for each SOEP worker.

This imputation reduces idiosyncratic measurement error in the self-reported BIBB feasibility item that arises from subjective constraints unrelated to a job’s inherent task content—employer WFH policy, local infrastructure, or living conditions such as available home-office space. Under the assumption that these subjective factors are orthogonal to the attributes used in the prediction model, the imputation effectively filters out the noise component and retains only the systematic, task-related variation. A further advantage is that the resulting measure varies at the individual rather than the occupational level, allowing more precise identification.

Measuring WFH feasibility. When respondents in the BIBB-BAuA survey indicate that “WFH is impossible in my job” when asked whether they would accept a WFH offer, their job is coded as not WFH-feasible. In addition, a job is coded as not WFH-feasible if the respondent reports *frequent* exposure to any of the following conditions:

1. Lifting or carrying loads of more than 10kg (women) / 20kg (men)
2. Exposure to smoke, dust, gases, or vapors
3. Exposure to cold, heat, moisture, humidity, or draughts
4. Working with oil, grease, or dirt
5. Handling microorganisms such as pathogens, bacteria, moulds, or viruses
6. Working the majority of working hours outdoors
7. Repairing or renovating

¹⁵The survey is conducted by the Federal Institute for Vocational Education and Training (BIBB) and the Federal Institute for Occupational Safety and Health (BAuA).

8. Protecting, guarding, monitoring, or regulating traffic
9. Cleaning, waste disposal, or recycling
10. Monitoring or controlling machines, plants, or technological processes

These additional restrictions ensure that the measure captures jobs in which a substantial fraction of tasks can realistically be performed from home. Under this definition, 39% of jobs in the BIBB sample are classified as WFH-feasible. This is broadly consistent with evidence that at the peak of the pandemic about one-third of employed persons in Germany regularly worked from home ([Alipour et al., 2025](#), Fig. 1).

Imputing WFH feasibility in the SOEP. The variable is constructed in two steps. First, we train a supervised classifier on the BIBB-BAuA training sample to predict (binary) WFH feasibility as a function of job characteristics. Second, we apply the trained model to impute 2019 WFH feasibility for each SOEP individual based on their 2019 occupation, industry, skill-requirement level, and demographic features.

Prediction model. The classifier for binary WFH feasibility is a logistic model with predictors comprising occupation (KldB 3-digit, 139 categories), industry (NACE 2-digit, 84 categories), skill-requirement level (KldB 5th digit, 4 categories), age (linear and squared), gender, and migration background (0/1). Predictors are selected by LASSO logit with the penalty parameter λ chosen by 5-fold cross-validation to minimize binomial deviance.¹⁶ At the optimal λ , the model retains 208 of the 231 features, dropping some sparsely populated levels of the categorical regressors and the linear age term. Dominance analysis using McFadden’s pseudo- R^2 as the fit statistic shows that occupation accounts for about two-thirds of the model’s fit (65% of the pseudo- R^2), with industry (19%) and skill-requirement level (15%) providing the remaining meaningful shares. The demographic controls jointly contribute about 1%; the overall pseudo- R^2 is 0.33.¹⁷

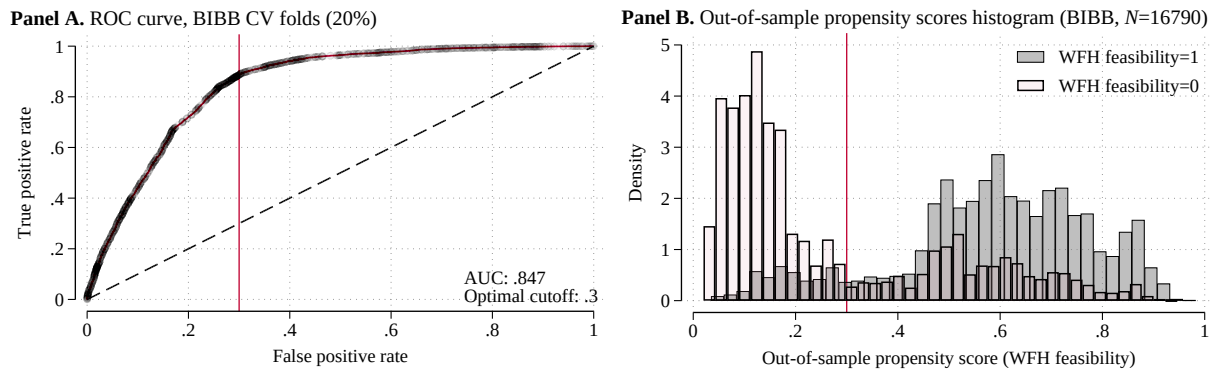
Model performance is summarized in [Figure C.1](#), based on out-of-sample predictions pooled across the five validation folds ($N = 16,790$). Panel A depicts the receiver-operating-characteristic (ROC) curve. The area under the curve (AUC) of 0.85 indicates an 85% probability that a randomly selected WFH-feasible job receives a higher

¹⁶We use Stata’s `cvlassologit` package ([Ahrens et al., 2020](#)).

¹⁷Dominance analysis for the LASSO-selected model averages each predictor’s marginal improvement in fit over all possible submodels, yielding a Shapley-value decomposition of explanatory power. We use Stata’s `domin` package ([Luchman, 2021](#)).

propensity score than a randomly selected non-feasible one.¹⁸ By conventional standards, an AUC above 0.8 indicates strong predictive performance. The vertical line at 0.30 marks the classification cutoff that maximizes Youden’s J (the difference between true-positive and false-positive rates), weighting the two error rates equally. Panel B plots the distribution of propensity scores by realized WFH-feasibility status; at the Youden cutoff the classifier correctly classifies 77% of cases (sensitivity 0.89; specificity 0.70).

Figure C.1: WFH feasibility classifier: out-of-sample validation



Notes: The figure summarizes the out-of-sample performance of the WFH-feasibility classifier. The classifier is a logistic model with predictors selected by 5-fold cross-validated LASSO, trained on the 2018 BIBB-BAuA Employment Survey. *Panel A* plots the receiver-operating-characteristic (ROC) curve of out-of-sample predictions pooled across the five validation folds. The area under the curve (AUC) of 0.85 summarizes the model’s ability to rank WFH-feasible and non-feasible jobs. The vertical line at 0.30 marks the classification cutoff that maximizes Youden’s J (true-positive minus false-positive rate). *Panel B* shows the distribution of out-of-sample propensity scores by realized WFH-feasibility status; the vertical line again marks the Youden cutoff. At this cutoff the classifier correctly classifies 77% of cases (sensitivity 0.89; specificity 0.70).

Imputing WFH feasibility in the SOEP. In the final step, the trained model is applied to each SOEP individual based on their 2019 occupation, industry, skill-requirement level, and demographic features. This yields a continuous propensity score and a binary indicator (evaluated at the Youden cutoff) of WFH feasibility. The average predicted feasibility score in the SOEP estimation sample is 0.38, which matches average WFH feasibility in the BIBB data.

¹⁸A random model produces an AUC of 0.5; a perfect model produces an AUC of 1. AUC summarizes model performance independently of any specific classification cutoff.

Appendix D Contract change versus target-hitting

The rise in performance pay incidence documented in [Section 4](#) is consistent with the principal-agent story of a *contract-change* to realign incentives under remote work. However, as the survey elicits performance pay *receipt*, it may also reflect a higher *target-hitting* probability under *existing* PP-contracts, caused by performance benefits under WFH, even among those without reported PP in 2019 (see Panel II, [Figure 3](#)).

The target-hitting story cannot be dismissed a priori, because the productivity effects of WFH are mixed and strongly context-dependent ([Mang and Anwar, 2026](#)). Estimates range from strongly negative in fully remote settings ([Emanuel and Harrington, 2024](#); [Atkin et al., 2023](#); [Gibbs et al., 2023](#); [Künn et al., 2022](#)), to roughly neutral under hybrid schedules ([Bloom et al., 2024](#); [Choudhury et al., 2024](#); [Angelici and Profeta, 2023](#)), to positive, particularly where the work is routine, as in call centers ([Bloom et al., 2015](#); [Aksoy et al., 2025b](#); [Fenizia and Kirchmaier, 2025](#)).

In the following, we present several pieces of corroborating evidence for the *contract-change* mechanism.

Performance-pay entry among workers with no prior contracts. First, the target-hitting mechanism can operate only where an incentive contract already exists, so entry among workers who never held performance pay isolates contract restructuring. Our entry-margin results already show rising PP incidence among those without reported PP in 2019. To ensure that a 2019 zero reflects the absence of a contract rather than a missed payout, we further exclude every worker who reported performance pay at any point between 2017 and 2019. [Table D.1](#) reports, for the baseline entry sample and this stricter “never-PP” sample, both the reduced form of 2019 WFH feasibility on the 2019–2023 change in performance-pay receipt (columns 1–4) and the 2SLS effect of WFH in 2023 instrumented by feasibility (columns 5–8, reproducing [Table B.2](#)). In both the reduced form and IV, the effects on the stricter sample are slightly smaller but remain positive and significant, though only at the 10% level in the most demanding specification with occupation and state fixed effects (columns 4 and 8). The never-PP effect is about two-thirds of the baseline, so excluding workers with any pre-2020 performance pay leaves the effect largely intact, suggesting that the estimates do mainly pick up genuine contract changes.

Intensification of incumbent contracts. Second, if WFH raised performance pay by improving target-hitting, workers already on incentive contracts should receive *larger* payouts. [Table D.2](#) tests this for workers who receive performance pay in both 2019

and 2023. Across three IV specifications (no fixed effects, industry fixed effects, and industry + state + occupation FE), WFH has no effect on the performance-pay share of annual earnings. This mirrors the null retention-margin share effect in [Table B.2](#). Note that conditioning on receipt in 2023 selects on a post-treatment outcome. And under better target-hitting that selection would push the share effect upward. Thus, if anything, the test is biased toward finding intensification, yet we find none.

Target-miss margins over time. Third, we examine whether WFH exposure predicts a different evolution of the annual exit hazard from performance pay, i.e., the probability of receiving none this year conditional on receiving PP last year. The idea is to test whether workers in WFH-feasible jobs and on PP are less likely to miss their targets after 2019. [Figure D.1](#) plots the exit hazard by WFH feasibility, conditional on staying with the same employer. Panel A shows that the target-miss rates are consistently lower for WFH-feasible workers. However, the lines trend in parallel throughout the observation period. Panel B shows this formally in an event study of WFH feasibility on the hazard rate: all estimates are close to zero and statistically insignificant at the 95% level, suggesting no break in target-hitting probabilities after the WFH shock.

Taken together, the evidence points to a change in contract structure rather than better target-hitting as the main mechanism. Performance pay rises sharply among workers who had no contract to begin with; it does not intensify among those who already held one, and higher WFH exposure does not yield different PP exit hazards.

Table D.1: Reduced-form and 2SLS estimates of WFH on performance-pay entry, by pre-2020 PP history

	Reduced form				2SLS			
	No PP 2019		Never PP		No PP 2019		Never PP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2019 WFH feasibility	0.101*** (0.023)	0.126*** (0.039)	0.068*** (0.023)	0.075* (0.039)				
WFH 2023 (0/1)					0.163*** (0.037)	0.161*** (0.050)	0.111*** (0.037)	0.098* (0.051)
First-stage F					[386]	[234]	[330]	[191]
Individuals	2,813	2,813	2,514	2,514	2,813	2,813	2,514	2,514
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓
Occupation FE		✓		✓		✓		✓
State FE		✓		✓		✓		✓

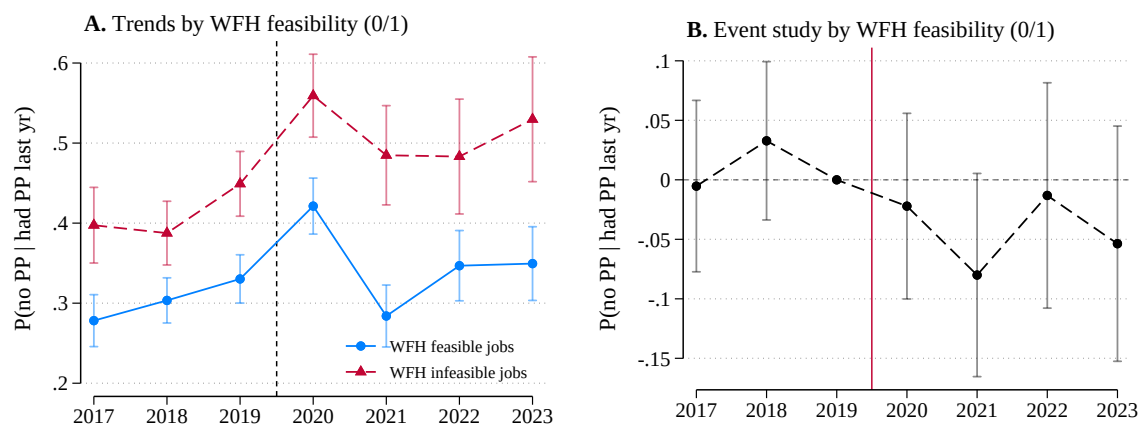
Notes: Columns (1)–(4) report reduced-form estimates of 2019 WFH feasibility on the within-individual 2019–2023 change in receipt of performance pay (0/1). Columns (5)–(8) report 2SLS estimates of WFH at least once per week in 2023 (Equations 1 and 2). The “No PP 2019” sample restricts to workers without performance pay in 2019 (entry margin, baseline). The “Never PP” sample restricts to workers without performance pay in any observed year 2017–2019. Kleibergen–Paap first-stage F -statistics are shown in square brackets. Standard errors clustered by individual in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table D.2: 2SLS estimates of WFH on the performance-pay share of earnings among continuing recipients

	PP earnings-share change (stable PP-holders)		
	(1)	(2)	(3)
WFH 2023 (0/1)	0.022 (0.015)	0.026 (0.024)	-0.005 (0.021)
First-stage F	[140]	[63]	[53]
Individuals	321	321	320
Industry FE		✓	✓
Occupation FE			✓
State FE			✓
<i>2019 DV mean:</i>			
Treated (WFH_FEAS=1)	0.086	0.086	0.086
Control (WFH_FEAS=0)	0.041	0.041	0.041

Notes: The table reports 2SLS estimates of the effect of WFH at least once per week in 2023 (0/1) on the within-individual 2019–2023 change in the PP share of annual earnings (Equations 1 and 2). The sample is restricted to workers who receive performance pay in both 2019 and 2023. 2023 WFH status is instrumented with 2019 WFH feasibility. Column (1) includes no fixed effects; column (2) adds industry fixed effects; column (3) adds occupation-segment and federal-state fixed effects. Kleibergen–Paap first-stage F -statistics are shown in square brackets. Standard errors clustered by individual in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Figure D.1: Target-miss hazard by WFH feasibility among performance-pay recipients



Notes: The figure plots the annual target-miss hazard (the probability of receiving no performance pay in year t among employed workers who received it in $t - 1$) by 2019 WFH feasibility, for workers who remain with the same employer between $t - 1$ and t , conditional on industry fixed effects (Panel A). Panel B plots the yearly DiD coefficient of WFH feasibility on individual hazard rates. Confidence intervals are at the 95% level, using standard errors clustered by individual. The unconditional hazard, which pools all workers regardless of employer continuity, is almost identical.

Appendix E Complier analysis

The IV estimates identify a local average treatment effect on the treated (LATET): the effect among compliers in the exposed group, i.e., workers who work remotely in 2023 because their 2019 job is WFH-feasible. Whether the LATET generalizes depends on how compliers differ from the rest of the workforce. We use the methodology of [Marbach and Hangartner \(2020\)](#); [Hangartner et al. \(2021\)](#) to characterize compliers relative to never-takers (workers who do not adopt WFH despite feasibility) and always-takers (workers who adopt WFH despite WFH-infeasible jobs).¹⁹

[Figure E.1](#) plots covariate means with 95% confidence intervals for the full sample and the three subpopulations; [Table E.1](#) reports the point estimates with bootstrapped standard errors. Compliers account for about 40% of the sample, never-takers for 51%, and always-takers for 8% (Panel A).

The contrast between the complier and never-taker subpopulations is informative. Never-takers earn 25% more at baseline (Panel F), have 2.8 more years of firm tenure (Panel H), are 13 percentage points more likely to hold a supervisor position (Panel J), and work overtime 13 pp more often (Panel I). They are also 13 pp more likely to be female (Panel C) and slightly less likely to have children under 14 in the household (Panel D). The pattern speaks to career-oriented workers, who look for traditional advancement (via personnel responsibility, longer tenure, bonuses and promotions that accompany seniority) and for whom the WFH-plus-PP package offers little incremental value, perhaps due to career concerns over remote work or complementarity between their tasks and office environments. Compliers look like workers who place higher private value on the WFH amenity: younger, more likely to have school-age children, and have shorter firm tenure. Always-takers ($\approx 8\%$ of the sample) are too small to characterize precisely. Their characteristics also suggest high private value of WFH (younger, more likely to have children, and longer commutes).

Two implications follow. First, the compliers' profile is consistent with workers who place a high private value on the WFH amenity, allowing the LATET to be read as the effect for worker-employer matches where WFH-for-PP packages are more easily implemented or accepted. Second, the never-takers' observable characteristics,

¹⁹We apply the decomposition from [Marbach and Hangartner \(2020\)](#) as a descriptive heuristic to characterize differences between observable always-takers (workers assigned to control who take the treatment) and never-takers (assigned treated who refuse treatment). Complier means are then identified by subtracting the weighted covariate means of always-takers and never-takers from the overall sample mean. Because our DiD-IV design relies on common trends rather than unconditional random assignment of WFH feasibility, the exact equalities may hold only approximately.

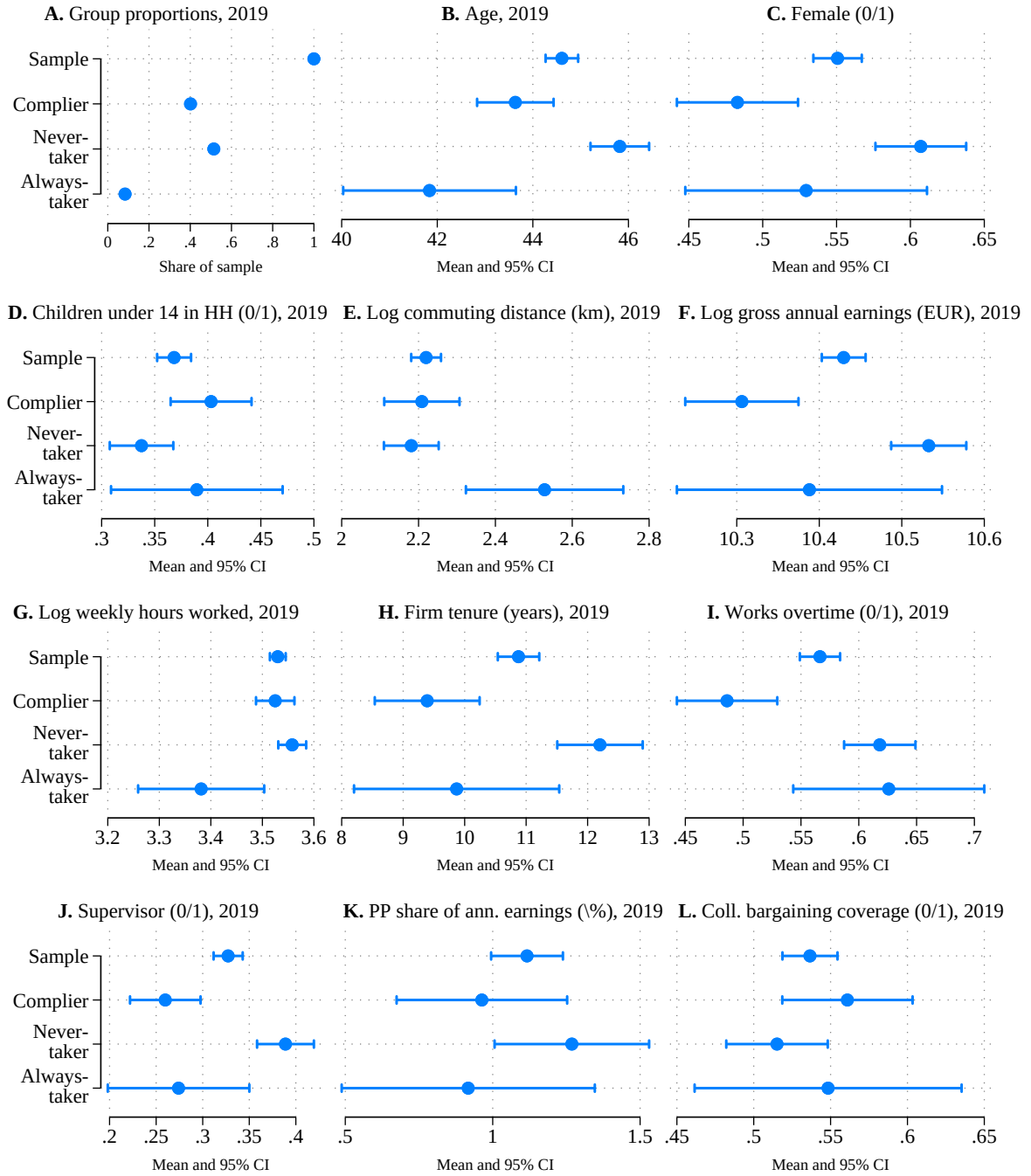
such as higher baseline earnings, longer tenure, and supervisory roles, are consistent with workers on established career tracks. If these workers systematically experience steeper counterfactual PP trajectories absent WFH through traditional promotions ([Prendergast, 1999](#); [Lazear and Rosen, 1981](#)), their inclusion in the OLS control group would bias the untreated trend upward. Under this plausible scenario, the IV-OLS gap reflects a combination of downward selection bias in OLS and the treatment-effect heterogeneity documented in our MTE analysis.

Table E.1: Complier analysis: covariate means by compliance type

	(1) Sample	(2) Complier	(3) Never-taker	(4) Always-taker
<i>Group share, 2019</i>	1.00	0.40 (0.01)	0.51 (0.01)	0.08 (0.01)
<i>Age, 2019</i>	44.61 (0.17)	43.63 (0.41)	45.82 (0.31)	41.84 (0.92)
<i>Female (0/1)</i>	0.55 (0.01)	0.48 (0.02)	0.61 (0.02)	0.53 (0.04)
<i>Children under 14 in HH (0/1), 2019</i>	0.37 (0.01)	0.40 (0.02)	0.34 (0.02)	0.39 (0.04)
<i>Log commuting distance (km), 2019</i>	2.22 (0.02)	2.21 (0.05)	2.18 (0.04)	2.53 (0.10)
<i>Log gross annual earnings (EUR), 2019</i>	10.43 (0.01)	10.31 (0.04)	10.53 (0.02)	10.39 (0.08)
<i>Log weekly hours worked, 2019</i>	3.53 (0.01)	3.52 (0.02)	3.56 (0.01)	3.38 (0.06)
<i>Firm tenure (years), 2019</i>	10.88 (0.17)	9.39 (0.44)	12.20 (0.35)	9.87 (0.85)
<i>Works overtime (0/1), 2019</i>	0.57 (0.01)	0.49 (0.02)	0.62 (0.02)	0.63 (0.04)
<i>Supervisor (0/1), 2019</i>	0.33 (0.01)	0.26 (0.02)	0.39 (0.02)	0.27 (0.04)
<i>PP share of ann. earnings (%), 2019</i>	1.12 (0.06)	0.96 (0.15)	1.27 (0.13)	0.92 (0.22)
<i>Coll. bargaining coverage (0/1), 2019</i>	0.54 (0.01)	0.56 (0.02)	0.52 (0.02)	0.55 (0.04)

Notes: The table reports means with standard errors (bootstrapped, 1,000 replications) in parentheses for the full sample and the three subpopulations identified using the method by [Marbach and Hangartner \(2020\)](#). The instrument is 2019 WFH feasibility (0/1), the endogenous variable is WFH at least once per week in 2023 (0/1). All covariates are measured in 2019. Corresponding plot: [Figure E.1](#).

Figure E.1: Complier analysis: covariate means by compliance type



Notes: The figure reports, in Panel A, the share of the sample classified as compliers, never-takers, and always-takers under the [Marbach and Hangartner \(2020\)](#) method. Panels B–L report covariate means with 95% confidence intervals (based on bootstrapped standard errors, 1,000 replications) for the full sample (*Sample*) and the three compliance subpopulations: *Compliers* adopt WFH iff their 2019 job is feasible; *Never-takers* do not adopt WFH irrespective of feasibility; *Always-takers* adopt WFH irrespective of feasibility. All covariates are measured in 2019 (pre-treatment).

Appendix F Robustness

The 2SLS estimates in [Figure 5](#) are robust to a range of specification and inference checks. [Table F.1](#) reports the baseline estimates alongside five alternative specifications that progressively add observable controls, remove likely pre-2020 remote workers, or modify the inference procedure. We focus on the pooled sample (Panel A) and the entry margin among workers without performance pay in 2019 (Panel B), which together capture most of the WFH-induced reallocation toward performance pay. In addition to conventional cluster-robust standard errors (reported in parentheses), each specification reports wild cluster bootstrap- t p -values (in square brackets), computed under the null using Rademacher weights and 10,000 replications to mitigate finite-sample distortions in 2SLS inference ([Young, 2022](#); [Cameron et al., 2008](#)).

Demographic controls. Column (2) adds indicators for cohabitation with a partner, children under age 14, gender, and migration background, together with age and age squared (all measured in 2019). The IV estimates are virtually unchanged. In the pooled sample, the estimated effect on PP incidence increases slightly from 18.9 to 19.0 percentage points, while the effect on the PP share remains at 1.4 percentage points; both retain bootstrap p -values below 0.01. The entry-margin estimates are equally stable (Panel B). These results suggest that the estimated WFH-induced shift toward performance pay is not driven by differential trends across observable demographic characteristics.

Job-level controls. Column (3) further adds controls for large firms (0/1), tenure, supervisory responsibility (0/1), collective bargaining agreement (CBA) coverage (0/1), and actual weekly hours worked (all measured in 2019). The effect on PP incidence declines modestly, to 16.5 percentage points in the pooled sample and 14.2 percentage points on the entry margin. The effect on the PP share falls to 1.0 (pooled) and 1.3 (entry) percentage points; but all remain statistically significant. Thus, the effects are also not simply driven by selection into jobs with different observable attributes.

Within-occupation and within-state variation. Column (4) presents the most demanding specification by adding fixed effects at the *KldB* occupation segment level (14 groups, [Matthes et al., 2015](#)) and the federal state level, restricting identifying variation to within-occupation, within-state, within-industry comparisons in addition to controlling for job-specific and demographic-specific trends. The PP-incidence effect is stable in the pooled and entry sample, at 15.6 and 14.6 percentage points, and significant under the bootstrap. The pooled PP-share effect is positive but no longer sig-

nificant; the entry-margin share effect attenuates to 1.0 percentage points but remains significant at the 5% level.

Trimming jobs with highest teleworkability. Column (5) drops observations with WFH-feasibility scores above the 90th percentile. The motivation is that workers in the upper tail are likely to be working remotely already before 2020. Because 2019 WFH status is unobserved, individuals who were already working remotely before 2020 risk contaminating the LATET because the estimator misclassifies them as shock-induced compliers, violating the common-trends assumption on the endogenous variable. Dropping these observations attenuates the pooled PP-incidence effect to 15.8 percentage points and the PP-share effect to 0.7 percentage points. The entry-margin estimates remain stable and significant. This pattern suggests that always-takers are more concentrated among workers with PP already in place at baseline, but that they do not meaningfully alter the conclusions.

Alternative inference. Column (6) clusters standard errors at the *KldB* 3-digit occupation level (123 categories) rather than at the individual level, allowing for arbitrary correlation across workers in the same occupation. The clusters are deliberately coarse and correspond to the features that explain most of the variation in individual WFH feasibility (see [Appendix C](#)). The estimates remain significant, confirming that inference is robust to this more demanding clustering. More broadly, the bootstrap *p*-values agree with the asymptotic clustered ones throughout [Table F.1](#), suggesting that conventional clustered inference is reliable in our setting.

Doubly-robust IV estimation. [Słoczyński et al. \(2022\)](#) show that when treatment effects are heterogeneous, adding covariates linearly to the second stage of 2SLS does not necessarily recover the LATE. The authors propose a doubly-robust estimator, which weights separate models for the outcome and treatment adoption by the inverse of the instrument propensity score (inverse-probability-weighted regression adjustment, IPWRA). The estimator is *doubly robust*, i.e., it is consistent as long as *either* the instrument propensity model *or* the outcome and treatment adoption models are correctly specified. Because the procedure requires a binary instrument, we replace the continuous feasibility score with the binary 2019 WFH-feasibility indicator and restrict the sample to the propensity-score overlap region. While this simultaneous change in estimator, instrument definition, and sample means the IPWRA estimates are not exact numerical counterparts to our baseline 2SLS, they qualitatively reinforce our main conclusions ([Table F.1](#)). The positive effects on PP incidence and the entry-margin PP share remain economically large and statistically significant across all specifications.

Notably, the pooled PP-share estimate, which attenuates in the heavily controlled 2SLS specification, is larger and significant under IPWRA (column 4).

Leave-one-out sensitivity. Finally, we test whether the effect of WFH on performance pay rests on a single occupation, sector, or region, using a leave-one-out strategy. The baseline 2SLS is estimated repeatedly, each time omitting one broad occupation (*KldB* segment), one NACE section, or one federal state. [Figure F.1](#) plots the distribution of estimates against the baseline (vertical solid line). Every estimate falls inside the baseline 95% confidence interval (dashed lines), and the first stage remains strong throughout ($F > 330$). Thus, we are confident that the estimated effect is broad-based and not driven by any single occupation, sector, or region.

Table F.1: Robustness of main DiD-IV estimates

	(1) Baseline	(2) +Demog. controls	(3) +Job controls	(4) +Occup. & State FE	(5) p90 WFH feas. trim	(6) Occup. 3-digit cluster
Panel A: Pooled sample						
<i>(i) Received performance pay (0/1)</i>						
WFH 2023 (0/1)	0.189*** (0.038) [0.000]	0.190*** (0.038) [0.000]	0.165*** (0.042) [0.000]	0.156*** (0.053) [0.002]	0.158*** (0.043) [0.000]	0.189*** (0.033) [0.000]
<i>(ii) Performance pay share of earnings</i>						
WFH 2023 (0/1)	0.014*** (0.005) [0.003]	0.014*** (0.005) [0.004]	0.010* (0.005) [0.061]	0.002 (0.006) [0.754]	0.007 (0.006) [0.183]	0.014** (0.007) [0.041]
Panel B: Entry margin (Baseline PP = 0)						
<i>(i) Received performance pay (0/1)</i>						
WFH 2023 (0/1)	0.163*** (0.037) [0.000]	0.163*** (0.036) [0.000]	0.142*** (0.040) [0.000]	0.146*** (0.052) [0.005]	0.166*** (0.042) [0.000]	0.163*** (0.035) [0.000]
<i>(ii) Performance pay share of earnings</i>						
WFH 2023 (0/1)	0.017*** (0.005) [0.000]	0.017*** (0.005) [0.000]	0.013*** (0.005) [0.003]	0.010** (0.005) [0.032]	0.015*** (0.005) [0.001]	0.017** (0.007) [0.003]
First-stage F (Pooled)	[481]	[489]	[396]	[258]	[368]	[109]
First-stage F (Entry)	[386]	[407]	[342]	[196]	[294]	[108]
Individuals (Pooled)	3,443	3,443	3,123	3,123	3,093	3,443
Individuals (Entry)	2,813	2,813	2,524	2,524	2,576	2,813
Industry FE	✓	✓	✓	✓	✓	✓
Demographic controls		✓	✓	✓		
Job controls			✓	✓		
Occupation + State FE				✓		
Cluster	Individual	Individual	Individual	Individual	Individual	Occup. 3-digit

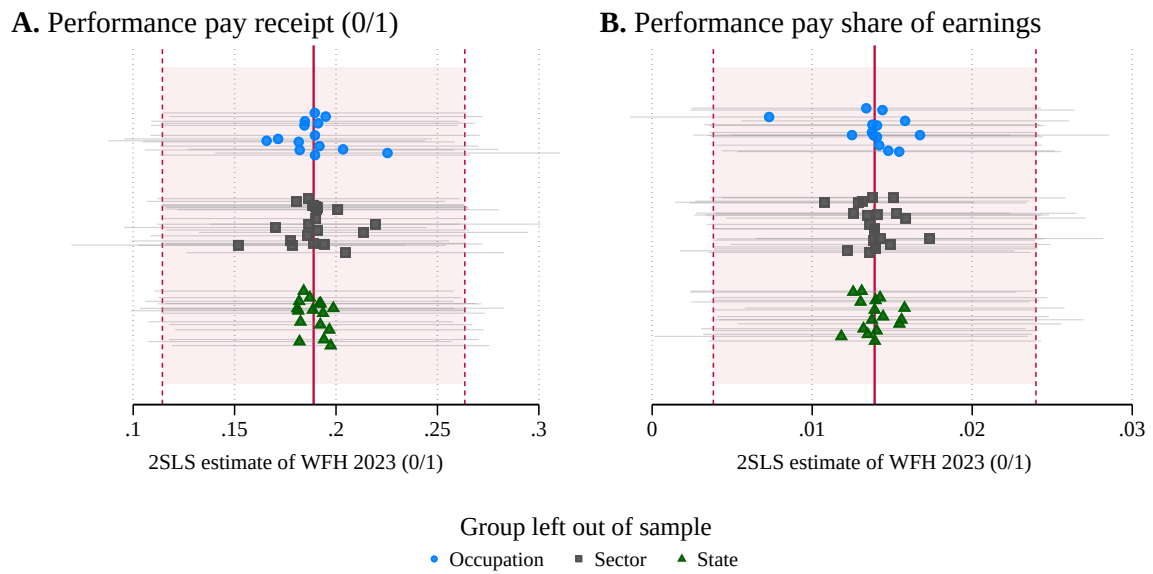
Notes: The table reports 2SLS estimates of the effect of WFH at least once per week in 2023 (0/1) on within-individual 2019–2023 changes (Equations 1 and 2). The outcomes are receipt of performance pay (i) and the PP share of annual earnings (ii). Panel A uses the pooled estimation sample with baseline PP status (0/1) as control; Panel B restricts to workers without PP in 2019. All specs instrument 2023 WFH (0/1) with 2019 WFH feasibility [0,1]. Column (1) is baseline, matching column (2) of [Table B.2](#) (the 2SLS estimates of [Figure 5](#)). Column (2) adds demographic controls measured in 2019 (partner living in household (0/1), presence of children under 14 (0/1), age, age-squared, gender (0/1), and migration background (0/1)). Column (3) adds job controls (large-firm (0/1), tenure (years), supervisor status (0/1), collective-bargaining coverage (0/1), and weekly hours actually worked, all 2019). Column (4) adds occupation-segment and federal-state fixed effects. Column (5) trims obs. with WFH-feasibility scores above the 90th percentile. Column (6) clusters SEs at the three-digit occupation level. Standard errors clustered by individual (columns 1–5) and by three-digit occupation (column 6) are reported in parentheses; p -values from a wild cluster bootstrap- t procedure with the null imposed (10,000 replications, Rademacher weights) reported in brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table F.2: Doubly-robust (IPWRA) LATET estimates of WFH on performance pay

	(1) Baseline	(2) +Demog. controls	(3) +Job controls	(4) +Occup. & State FE	(5) p90 WFH feas. trim	(6) Occup. 3-digit cluster
Panel A: Pooled sample						
<i>(i) Received performance pay (0/1)</i>						
WFH 2023 (0/1)	0.258*** (0.071)	0.271*** (0.067)	0.271*** (0.067)	0.219** (0.087)	0.230*** (0.067)	0.258*** (0.073)
<i>(ii) Performance pay share of earnings</i>						
WFH 2023 (0/1)	0.020*** (0.006)	0.021*** (0.006)	0.020*** (0.006)	0.014** (0.005)	0.016*** (0.006)	0.020** (0.008)
Panel B: Entry margin (Baseline PP = 0)						
<i>(i) Received performance pay (0/1)</i>						
WFH 2023 (0/1)	0.163*** (0.048)	0.158*** (0.048)	0.117** (0.058)	0.201** (0.081)	0.156*** (0.049)	0.163*** (0.057)
<i>(ii) Performance pay share of earnings</i>						
WFH 2023 (0/1)	0.019*** (0.004)	0.018*** (0.003)	0.015*** (0.003)	0.020*** (0.004)	0.017*** (0.004)	0.019*** (0.006)
Individuals (Pooled)	3,443	3,443	3,123	2,652	3,093	3,443
Individuals (Entry)	2,813	2,813	2,524	2,161	2,576	2,813
Industry FE	✓	✓	✓	✓	✓	✓
Demographic controls		✓	✓	✓		
Job controls			✓	✓		
Occupation + State FE				✓		
Cluster	Individual	Individual	Individual	Individual	Individual	Occup. 3-digit

Notes: The table reports doubly-robust LATET estimates of the effect of WFH at least once per week in 2023 (0/1) on within-individual 2019–2023 changes in two outcomes (receipt of performance pay and the PP share of annual earnings), using the inverse-probability-weighted regression adjustment (IPWRA) approach proposed by [Śloczyński et al. \(2022\)](#). The estimator requires a binary instrument, so 2023 WFH status is instrumented with the binary 2019 WFH-feasibility indicator. Observations outside the propensity-score overlap region are trimmed. The panels, outcomes, columns, and control variables mirror [Table F.1](#) (see its notes for the list of covariates). Standard errors clustered by individual (columns 1–5) and by three-digit occupation (column 6) are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Figure F.1: Leave-one-out 2SLS estimates of WFH on performance pay



Notes: The figure shows 2SLS estimates (and 95% confidence intervals) of the effect of 2023 WFH adoption (Equations 1 and 2). The outcomes are within-individual changes (2019–2023) in the receipt of performance pay (Panel A) and the PP share of annual earnings (Panel B). Each marker is a separate regression that repeats the baseline specification, conditional on industry fixed effects and baseline PP status, while leaving out one broad occupation (*KldB* segment), one NACE section, or one federal state. The solid red line marks the full-sample baseline estimate (Figure 5, Panel I) and the dashed lines its 95% confidence interval, based on standard errors clustered by individual. The smallest Kleibergen–Paap first-stage *F*-statistic is 383 across occupations, 331 across sectors, and 376 across states.