Managers and the Persistence of Productivity*

Aroon Narayanan
MIT

Roi Orzach
Boston University

November 2, 2026

Abstract

When productivity is persistent, the gains from having a better manager build up over time. We show theoretically that this build-up leads static specifications to understate the long-run effect of managers. To measure this effect empirically, we develop and apply two methods. First, for event studies, we adapt the fused lasso method to obtain a nonparametric estimate of the effect at the end of the event window. In three influential event studies, this estimate exceeds the original static estimate in magnitude, and in two of them it is larger by 38 to 49 percent. The second method identifies persistence in a parametric model with an instrumental variables strategy and can be applied when an event study is not available, for example, when managerial quality must be inferred from output. Among U.S. public companies, profitability and firm scale are both persistent, and scale far more so. Conservative estimates of persistence imply that the long-run dispersion of CEO effects on sales and employment is 38 to 46 percent larger than static specifications suggest.

*We are thankful for helpful conversations with Charles Angelucci, Mert Demirer, Miguel Espinosa, Robert Gibbons, Andy Newman, Vishan Nigam, Fiona Paine, Andrea Prat, Raffaella Sadun, and Antoinette Schoar. We thank Guo Xu for making his data publicly available, and Henrik Cronqvist, Frank Yu, and Cristóbal Otero for graciously sharing their data with us. We also thank seminar and conference participants at Boston University, MIT, and the CEPR IMO-ESF Workshop for helpful comments.

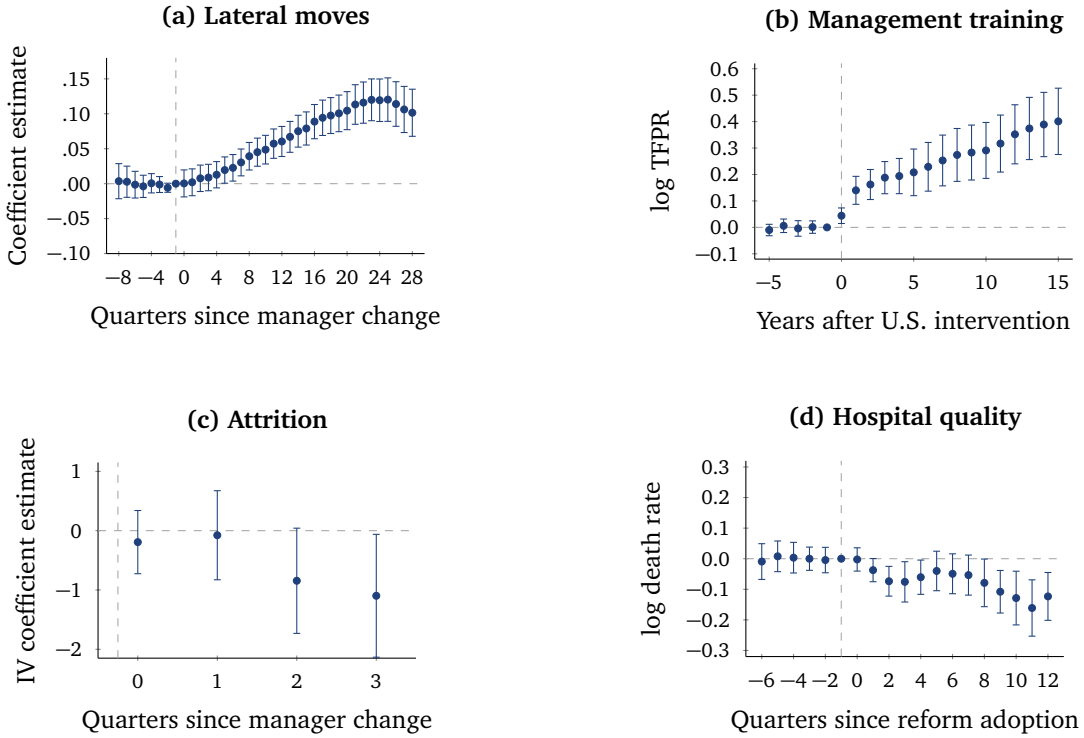
1 Introduction

Productivity is the cornerstone of business performance and a first-order object for economic analysis. A large and growing literature has shown that even among seemingly similar establishments, productivity differences are large, persistent, and economically consequential (Syverson, 2011; Hsieh and Klenow, 2009). A leading explanation for these performance differences is the effect of managers. This empirical literature builds on Bertrand and Schoar (2003), who isolated managerial style by tracking executives across firms to document their significant impact on corporate performance. Subsequent research has reproduced this finding across a wide variety of settings, including service jobs (e.g., Lazear et al., 2015), public-sector administration (e.g., Fenizia, 2022), multinationals (e.g., Minni, 2026), hospitals (e.g., Muñoz and Otero, 2025), and laboratory experiments (e.g., Weidmann et al., 2026).

Most of this work assumes that managerial effects are both immediately present when a new manager arrives and independent of the quality of the predecessor. This assumption runs counter to much of what managers do: hiring, implementing managerial practices, and developing client relations, all of which are slow and path-dependent processes. For instance, Bloom et al. (2016) find that managerial practices are twice as durable as physical capital.[1] More broadly, many strategic decisions leave a durable imprint on organizations: investments in physical capital, hiring choices, and organizational routines all outlast the tenure of the manager who implemented them. Such persistence implies that the positive effects of hiring a better manager will not be immediately realized, as we show in Figure 1, reproducing event studies from influential papers on the effect of managers. Relatedly, Bandiera et al. (2020) find that the performance advantage of leader-type CEOs over manager-type CEOs emerges only after approximately three years in office. Understanding this persistence is important for its own sake, for correctly estimating managerial productivity, and, as we discuss below, for optimal contracting and the evaluation of policy interventions.

[1] Relatedly, Cai et al. (2026) show that organizational capital, as measured through Glassdoor reviews, is highly persistent and influenced by managers. The key difference from Cai et al. (2026) or Bloom et al. (2016) is that managerial practices (or organizational capital as measured in Cai et al., 2026) are observed by the econometrician, whereas in the analysis that follows managerial capital will have to be estimated.

Figure 1: Event-study evidence on the dynamic effects of managers on organizational outcomes



Notes: Panel (a): from Minni (2026). The panel plots the differential cumulative number of lateral job moves for workers who transition from a "low-flyer" to a "high-flyer" manager relative to those who transition between low-flyer managers. Panel (b): from Giorcelli (2019). The panel plots log revenue total factor productivity (TFPR) for Italian firms whose managers received U.S.-sponsored training under the Marshall Plan, relative to firms excluded by an unexpected budget cut. Panel (c): from Hoffman and Tadelis (2021). The panel plots IV estimates of the effect of a one-standard-deviation increase in manager people-management quality on employee attrition, by quarter since the manager change, at a large technology firm. Panel (d): from Muñoz and Otero (2025). The panel plots the dynamic effect on log in-hospital death rates of adopting a meritocratic selection reform for the CEOs of Chilean public hospitals. In all panels, bars are 95 percent confidence intervals.

As evident in the event studies of Figure 1, the question "what is the effect of a manager on output?" is ambiguous in the presence of persistence. Is it the effect in the first period after a change (henceforth, the *short-run* effect), or the eventual steady-state gain (henceforth, the *long-run* effect)? For an organization deciding how to select its managers, the latter is arguably the more relevant object. We develop a theoretical framework that can differentiate these effects, allowing output to depend on arbitrarily many past lags of output. Within this framework, we compare the true long-run effect with the effect recovered by a static analysis in which each manager has an effect independent of time in office or inherited performance. The static analysis understates the long-run effect. For intuition, a static specification summarizes the evolving effects in Figure 1 with a single "average" effect, combining the smaller initial responses with the larger responses observed later. The static analysis also understates the share of productivity differences attributable to managers. Finally, in a parametrized model,

we prove that both understatements are proportionally larger when persistence is higher and when the tenure of managers is shorter.

We apply our methods to several influential studies of managerial effects. When an event study is available, the estimates suggest a natural way to measure the long-run effect: identify where the effects settle and estimate the level there. The difficulty is doing this without choosing a stretch of the graph by eye.[2] The fused lasso, developed by Tibshirani et al. (2005), offers a way to group consecutive post-event periods into blocks with a common fitted level. Tibshirani and Taylor (2011) provide an algorithm for computing these fits and a rule for choosing how much to pool from the data.[3] The procedure allows the earlier effects to evolve flexibly without specifying a parametric production process. We take the level of the last block, which we call the end-of-window effect, as a conservative estimate of the long-run effect. If the gains are still building when the data end, the end-of-window effect may understate the eventual long-run effect. We apply this procedure to several influential event studies in the literature for which we could access data. In every application, the end-of-window effect exceeds the static estimate reported in the original paper in absolute value. In two of the three applications, it is 38 to 49 percent larger.

The event studies trace the effects of known policies or managerial characteristics that are observed by the econometrician. In each case, the goal is to estimate an average response to that policy or characteristic and how it evolves over time. Other settings, such as that studied by Bertrand and Schoar (2003), seek to recover the distribution of effects across individual managers. Here, managerial quality is itself unobserved and must be inferred from output. Nonparametrically estimating both an unrestricted distribution of managerial quality and an arbitrary path of adjustment is a substantially more demanding problem. We therefore impose a parametric structure on persistence while allowing managers to differ freely in their per-period contributions. Plant output evolves as $y_{t,p} = \delta\, y_{t-1,p} + \theta_{m_{t,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}$, where $\delta \in [0, 1)$ measures the persistence of productivity. In this model, which generalizes the static specification of Bertrand and Schoar (2003), a manager's per-period contribution θ_m corresponds to a long-run effect of $\theta_m/(1-\delta)$, which can be interpreted as the gain in productivity if they or a similarly qualified manager remained in place in perpetuity. The standard two-way fixed effects model is the special case $\delta = 0$, in which these two effects coincide. We give conditions on the hiring processes of firms under which δ is point-identified using an instrumental variables strategy in the spirit of Anderson and Hsiao (1982).

[2] Event-study methods also allow tests of whether effects have leveled off over a chosen set of late periods (Freyaldenhoven et al., 2025). Researchers can likewise compare average effects in the first few and last few post-event periods. These approaches still require choosing the relevant periods and how to combine their estimates.

[3] We use the rule suggested in their paper, which allows more blocks only when they improve the fit enough to justify the added complexity.

We then estimate this model with U.S. public-company CEOs, in an ExecuComp–Compustat panel built in the spirit of Bertrand and Schoar (2003). We employ two estimation procedures: an IV estimation in levels and forward orthogonal deviations (FOD; Arellano and Bover, 1995). The levels procedure results in conservative estimates due to Nickell bias. FOD imposes additional restrictions on the data but exhibits a smaller bias in the Monte Carlo simulations we run. The levels estimates show lower persistence for profitability ($\hat{\delta} \approx 0.27$–$0.38$) and higher persistence for firm scale ($\hat{\delta} \approx 0.74$ for log sales, $\hat{\delta} \approx 0.78$ for log employment). Intuitively, competitive forces should erode profit margins quickly, while decisions that shape firm scale do not unwind as fast. Applying our methodology, we find that the long-run dispersion of CEO effects on firm scale is larger than the static estimate by 38–46 percent at the levels estimates and by 57–89 percent at the FOD estimates.

Implications for the literature. Our framework yields several implications for the literature on manager and leader effects. First, static estimates both understate the long-run value of leadership and misrank managers. Therefore, incorporating persistence may help explain the large compensation premiums observed in competitive executive labor markets, which standard models often struggle to justify (Gabaix and Landier, 2008; Edmans et al., 2017).

Second, when firms or researchers conduct randomized controlled trials to evaluate interventions (e.g., Bloom et al., 2013), they typically measure short-to-medium-term treatment effects. If persistence is high, these short-run evaluations will understate the true long-run gains. Interventions that appear marginal over a short horizon, such as hiring a higher-quality manager, may be worthwhile once the full dynamic effect is accounted for. This mechanism may also help sustain persistent productivity differences (e.g., Syverson, 2011) across firms. If lagging firms judge managerial reforms by their short-run effects, they may underestimate the long-run returns and forgo improvements that would otherwise help them catch up.

Third, persistence has implications for optimal performance contracts. Dessein and Prat (2022) model a setting in which a CEO chooses between building organizational capital and boosting short-run performance. An implication of their analysis is that when organizational capital is persistent and imperfectly observed, hiring and firing decisions should weigh a CEO's contribution to organizational capital and not only short-run performance. Our estimates provide empirical support for their assumption of high persistence that drives this implication.[4] More broadly, the optimal contracting literature frequently advocates evaluating managers on residual current performance or relative performance (Holmström, 1982; Murphy, 1999). If current output depends in part on the cumulative effect of past managers, however, these con-

[4]Our econometric framework further suggests that firms can, in principle, estimate and contract on organizational capital, but the signal would be noisy and require several periods of observation.

tracts add unnecessary variance to a manager's evaluation and penalize managers who inherit a low output level from weak predecessors. The same problem can also distort managerial selection and retention. A weak manager who inherits high productivity from a strong predecessor may appear more capable than a stronger manager who inherits a poorly performing organization.

Finally, we acknowledge the ongoing econometric debate regarding the validity of two-way fixed effects (TWFE) models for estimating manager fixed effects. Our approach to estimating persistence is orthogonal to this debate: our identification operates on a sample of firms where the manager stays, uses deeper lags as instruments, and remains valid even if managers never move across plants. Moreover, this concern does not apply to the event-study evidence.

The remainder of the paper proceeds as follows. Section 2 proves that static frameworks will result in understatements in dynamic environments. Section 3 conducts the event-study analysis. Section 4 provides the time-series evidence. Section 5 concludes, and the Appendix provides proofs, the Monte Carlo evidence, and additional robustness checks.

2 Persistence and the Understatement of Manager Effects

This section develops a theoretical framework for studying the effects of managers on productivity in the presence of persistence and establishes the biases that result from estimating a static model that ignores persistence.

The starting point for the analysis is the practical question that the literature has tried to address: how much does a better manager raise output? Recall from the introduction that with persistence, the gains from a better manager build up over time. We focus on the resulting steady-state gain in output (the long-run effect) from employing such a manager as opposed to the immediate benefit (the short-run effect). We show that a model without persistence (a static model) understates the long-run effect.

Throughout, we will index plants by p, time by t, and the manager assigned to plant p at date t by $m_{t,p}$. Manager m has productivity θ_m, and the θ's are drawn from a distribution with compact support and standard deviation σ_θ across managers. Output satisfies

$$y_{t,p} = \sum_{k\geq 1} \delta_k y_{t-k,p} + \theta_{m_{t,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}, \qquad \delta_k \geq 0, \qquad \sum_{k\geq 1} \delta_k < 1. \tag{1}$$

Here α_p and γ_t are plant and time fixed effects, and $\epsilon_{t,p}$ is a production shock. Output is increasing in manager quality and nondecreasing (i.e., $\delta_k \geq 0$) in each lagged output.[5] We

[5] Corollary 2 in the Appendix shows that our results hold when output instead depends on lags of managers' qualities. This follows almost immediately, as inverting the autoregressive form of Equation (1) yields its moving-

assume $\sum_k \delta_k < 1$ so that the effect of a temporary shock eventually subsides. We additionally assume that for each plant, the mean and variance of output remain bounded over time (including any pre-sample dates that are needed given the lag formulation).

Nonzero persistence implies that if a manager increases output by one unit today, then output will be higher by δ_1 tomorrow and by $\delta_1^2 + \delta_2$ two periods from today, even if manager quality returns to its baseline from tomorrow onward. Iterating Equation (1) forward shows that a one-unit increase in manager quality for a single period raises output j periods later by ψ_j, where

$$\psi_0 := 1, \qquad \psi_j := \sum_{k=1}^{j} \delta_k \psi_{j-k} \quad (j \geq 1).$$

Without persistence, $\psi_j = 0$ for all $j \geq 1$, so a manager who is one standard deviation better raises output by exactly σ_θ in each period of their tenure and by nothing once they leave. With persistence, the first-period gain is still σ_θ, but the gains accumulate as long as the plant keeps employing a manager who is one standard deviation better (either this exact manager or a similarly able manager). The gain after t periods is $\sigma_\theta \sum_{j=0}^{t-1} \psi_j$, which converges to the long-run effect $\sigma_\theta \sum_{j\geq 0} \psi_j$.

Finally, we impose the following assumption on manager assignment.

Assumption 1 (Manager Assignment)**.** *For every $j \geq 0$,*

$$\mathbb{E}[\theta_{m_{t-j,p}} \mid m_{t,p} = m, p, t] = a_{j,p} + b_{j,t} + r_j \theta_m, \tag{2}$$

$$\mathbb{E}[\epsilon_{t-j,p} \mid m_{t,p} = m, p, t] = 0, \tag{3}$$

where $0 \leq r_j \leq 1$, $r_0 = 1$, and $a_{0,p} = b_{0,t} = 0$.

The sequence $\{r_j\}_j$, in which a one-unit difference in current manager quality corresponds to a difference of r_j in expected quality j periods earlier, captures the hiring process. For example, a process whereby, each period, a manager is rehired with probability q and otherwise replaced with a manager from a pool of potential managers results in $r_j = q^j$. Instead, if an appointment lasts exactly L periods, with the current manager's tenure so far equally likely to be any length up to L, then $r_j = \max\{1 - j/L, 0\}$.

Assumption 1 permits better plants to hire better managers ($a_{j,p}$) and some periods to bring better managers on average ($b_{j,t}$), but it rules out hiring and retention that respond to past performance, as well as better managers staying longer. Equation (3) allows production shocks to be serially correlated, provided current and past shocks have mean zero conditional on the current manager, plant, and date. Nonetheless, Assumption 1 is strong. We use it to obtain

average representation, in which output is a weighted sum of current and past managers' qualities.

closed forms for the understatements of static estimates, and the estimation procedures in Sections 3 and 4 do not rely on it.

We are now ready to compare the true effect of managers to what a researcher would estimate if they assumed a static model. To set sampling error aside, we work with the population static regression (i.e., assuming infinite data). Let β_m^{static} denote the manager effects and σ_{static} their standard deviation across managers. We assume the static regression meets the standard conditions for identifying manager fixed effects (e.g., the connected-set assumption in Bertrand and Schoar, 2003). The manager effects are then identified up to a common additive constant, which we denote by C; the value of this constant does not affect the comparison across managers.

To build intuition behind our first result, suppose that a new manager with $\theta = 1$ takes over at an "average" plant for two periods. Then, in period 1, the productivity improvement relative to the previous manager is equal to 1. The productivity gain in the next period will be $1 + \psi_1$, as the plant also benefits from the productivity improvement in the last period. A static model averages these two improvements and recovers the manager's effect as $1 + \psi_1/2$, which is larger than the true short-run effect. However, the long-run effect (i.e., the effect if the plant continues to employ similarly able managers) is not only larger than $1 + \psi_1/2$, but (weakly) larger than $1 + \psi_1$, as the productivity gains will continue to compound over time. The following proposition formalizes this basic intuition and provides closed-form expressions for the magnitude of the bias.

Proposition 1 (Biased Productivity)**.** *Suppose the output follows Equation* (1) *and Assumption 1 holds. Define*

$$B_{\text{static}} = \sum_{j \geq 0} \psi_j r_j, \qquad B_{\text{long run}} = \sum_{j \geq 0} \psi_j, \quad \textit{so that} \quad 1 \leq B_{\text{static}} \leq B_{\text{long run}}. \tag{4}$$

In the population-level static regression, the manager effects and their standard deviation satisfy

$$\beta_m^{\text{static}} = B_{\text{static}} \theta_m + C, \qquad \sigma_{\text{static}} = B_{\text{static}} \sigma_\theta. \tag{5}$$

Therefore, the static framework would imply a $\sigma_{\text{static}} = B_{\text{static}} \sigma_\theta$ output gain from permanently hiring managers whose quality is one standard deviation higher. Instead, the true production process implies an eventual gain of $B_{\text{long run}} \sigma_\theta$.

The proposition implies that the static estimate weakly overstates the short-run effect and weakly understates the long-run effect. Further, the long-run effect strictly exceeds the static estimate except in two degenerate cases. In the first, output is not persistent ($\delta_k = 0$ for all k).

In the second, current manager quality predicts one-for-one the quality of every past manager whose contribution still carries over to current output ($r_j = 1$ for all j with $\psi_j > 0$).

The following corollary provides a simpler expression of these results when the manager hiring process is geometric, as in the first example above ($r_j = q^j$), and an even simpler one when production is also AR(1).

Corollary 1 (Geometric Assignment and Productivity Bias)**.** *Under the conditions of Proposition 1, suppose $r_j = q^j$, where $q \in (0,1)$. Then,*

$$\frac{B_{\text{static}}}{B_{\text{long run}}} = \frac{1 - \sum_{k\geq 1} \delta_k}{1 - \sum_{k\geq 1} \delta_k q^k}. \tag{6}$$

When further restricting to AR(1) production ($\delta_1 = \delta$ and $\delta_k = 0$ for $k \geq 2$), and letting $\mu := 1/(1-q)$ denote the average tenure, then

$$\frac{B_{\text{static}}}{B_{\text{long run}}} = \frac{1-\delta}{1-\delta+\delta/\mu}. \tag{7}$$

This corollary expresses the bias of the static estimate in closed form, as a function of persistence and average manager tenure. Intuitively, when persistence is lower, the magnitude of the bias from using a static estimation procedure is lower. Further, when average tenure, μ, is longer, then the effects of past managers gradually disappear from the current manager's performance. As this occurs, the static estimates of managerial effects approach the true long-run effect. These results imply that a static framework understates both the dispersion in managers' long-run effects on output and the eventual gains from policies that permanently improve manager selection. In the empirical analyses, we show that these biases can be substantial as the average tenure is low and the estimated persistence is high.

The same logic carries over to the share of output differences attributable to managers, which the static model also understates (Appendix A.1). Just as σ_{static} credits the current manager with only the part of past managers' contributions that the current manager's quality predicts, the static R^2 assigns the remaining part, which differs across plants with equally good current managers, to production shocks. The two understatements have the same qualitative forces, and under geometric hiring and AR(1) production, both share the same comparative statics: both are proportionally larger when persistence is higher and average tenure is shorter.

3 Persistence: Event-Study Evidence

Section 2 shows that when output is persistent, a static estimate understates the long-run effect of a manager. The benefit of an event study is that it allows for a clean visual of how the effect builds up over time after an event, by reporting a separate coefficient for every period after the event. The cost of this flexibility is that no single coefficient captures the long-run effect. How can a researcher report a single point estimate of the long-run effect using the event study, without seeming ad hoc? In this section, we describe a method from statistics that solves this problem, and then apply it to three event studies.

3.1 Methodology

The fused lasso of Tibshirani et al. (2005) provides a way to group consecutive post-event periods into blocks with a common estimated effect, without imposing a functional form on how the effect evolves. The fused lasso works much like the standard lasso, which estimates a regression with a penalty, scaled by a parameter λ, on the absolute size of each coefficient, and so sets the least useful coefficients exactly to zero. The fused lasso instead fits a path to the event-study coefficients and penalizes the jumps between consecutive periods of that path, at a cost of λ per unit of jump. It keeps a jump only when the jump improves the fit to the event-study coefficients by more than its penalty; otherwise it sets the two adjacent fitted coefficients exactly equal and merges their periods into one block. A larger λ therefore produces fewer, longer blocks. Formally, given the K post-event coefficients $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)$ and their covariance matrix $\hat{V}$, the fused lasso chooses the path $\theta = (\theta_1, \ldots, \theta_K)$ that minimizes $\frac{1}{2}(\hat{\theta} - \theta)'\hat{V}^{-1}(\hat{\theta} - \theta) + \lambda \sum_{k=2}^{K} |\theta_k - \theta_{k-1}|$.

The practical question is then how to choose λ. Much as with the standard lasso, too small a λ overfits, keeping jumps that reflect only noise in the event-study coefficients, while too large a λ underfits, merging periods whose effects truly differ. We follow Tibshirani and Taylor (2011), who show how to choose the λ that best balances the two: the one whose fitted path is estimated to be closest to the true path of effects.[6] The procedure is also fast: their algorithm computes the fit for every value of λ at once. We use this procedure for two reasons. First, it is a well-known procedure within the statistics literature with good estimation properties. Second, it ties our hands: with no tuning choice left to the researcher, any event study, meaning its coefficients and their covariance matrix, maps into a single estimate that

[6] Specifically, we use their dual path algorithm (Algorithm 2, Section 6.1), which computes the fused-lasso fit for every value of λ, together with their rule for choosing λ (Section 10.2). That rule picks the λ that minimizes Mallows's C_p, a standard criterion that trades off how closely the fit tracks the event-study coefficients against how many blocks it uses. They show that, for any given λ, C_p is an unbiased estimate of the fit's precision-weighted mean squared error relative to the true path of effects, so the rule picks the λ with the smallest estimated error.

anyone can reproduce.

We call the level of the last block the *end-of-window effect*. If the effect has stopped growing by then, this is the long-run effect; if it is still growing, the long-run effect is larger. Moreover, the penalty pulls the level of each block toward those of its neighbors, so when effects build up, the level of the last block is smaller in absolute value than most of the coefficients it pools, resulting in yet more conservative estimates. In simulations calibrated to our three applications (Appendix B.1), this level is close to unbiased when the true effect reaches its full size immediately after the event, and it understates the true effect at the end of the window, typically by 10 to 30 percent, when the true effect instead builds up gradually.[7] Therefore, we view this as a conservative estimate of the long-run effect.

Because λ is chosen from the data, a standard error for the end-of-window effect would ignore the choice of λ. We therefore construct confidence intervals with a parametric bootstrap in the spirit of Efron (2014). We draw the event-study coefficients from a normal distribution centered on the point estimates, with their estimated covariance matrix, and repeat the whole procedure, including the choice of λ, on each draw. Separately, to test whether the effect builds up within the window, we compare the mean of the last m post-event coefficients with the mean of the first m, where m is one third of the number of post-event coefficients, rounded down (Table 2).

3.2 Hospital CEO Selection in Chile

To see how the effect builds up over time after an event, we first apply the fused lasso to an event study from Muñoz and Otero (2025), shown in Panel (d) of Figure 1. Their headline coefficient, which is the static estimate, averages the effect over the twelve quarters after adoption. We instead fit the fused lasso to the quarter-by-quarter event-study coefficients. The fit suggests that the effect grows over these quarters and ends larger in absolute value than their headline coefficient.

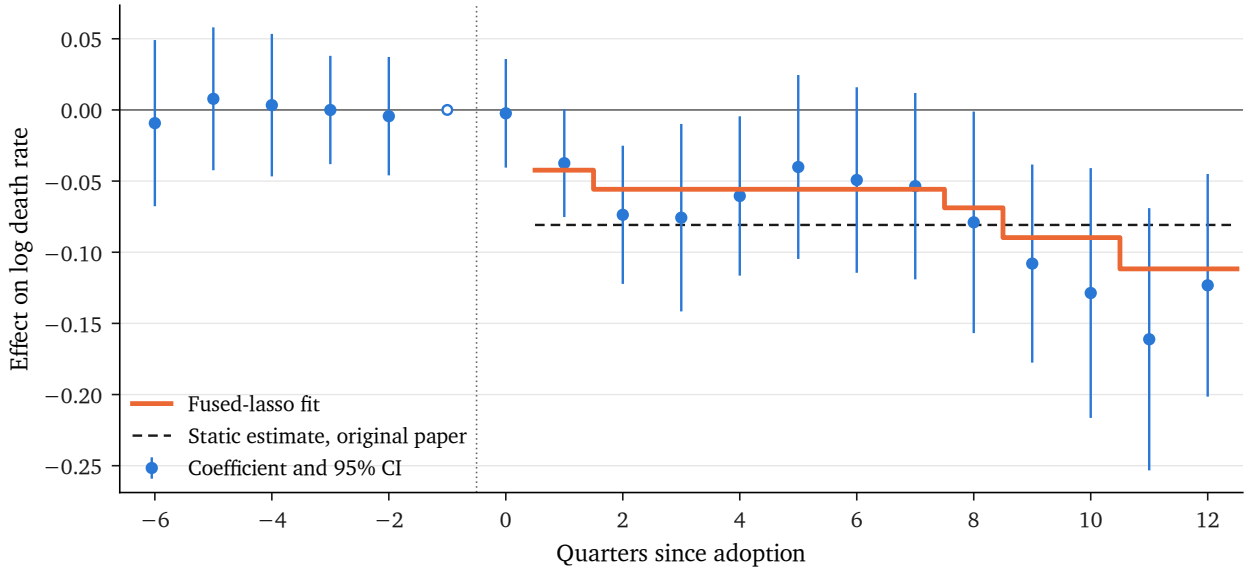
As background, in 2003, Chile replaced the discretionary appointment of senior public executives with a public, competitive selection process and higher pay. Muñoz and Otero (2025) define a hospital as adopting the reform when it first appoints a CEO through the new system. From then on, every later CEO of that hospital must also be selected through the system, so adoption is permanent. The outcome studied in the paper is the log in-hospital death rate. They estimate a stacked event study, from six quarters before to twelve quarters after adoption. Each adopting hospital is compared with its own control group, which excludes hospitals that

[7] If the true effect instead fades over time, the end-of-window effect overstates the long-run effect. However, visually, none of the event studies we consider shows sustained fading effects.

have already adopted. Their headline coefficient is −0.081, which corresponds to an average 8.1 percent decline in in-hospital death rates over the three years after adoption.

The authors kindly shared their replication files with us. Using these files, we reproduce their event study and confirm that our post-adoption coefficients match those reported in Figure 3 of Muñoz and Otero (2025). The replication also produces the covariance matrix of these coefficients, which is the key input to the fused lasso. A new CEO usually takes office partway through a quarter, so the quarter of adoption is only partly treated. As in their analysis, we therefore count the quarter after adoption as the first treated quarter. We also keep their event-study window, which ends twelve quarters after adoption, and fit the fused lasso to the coefficients for quarters 1–12.

Figure 2: Hospital CEO selection in Chile: event study and fused-lasso fit



Notes: The outcome is the log in-hospital death rate. Event time is quarters since a hospital first appointed a CEO through the new selection system (Muñoz and Otero, 2025); quarter −1 is the reference period, normalized to zero. Standard errors are clustered by hospital, and bars are 95 percent confidence intervals. The step line is the fused-lasso fit to the post-adoption coefficients, quarters 1–12 (quarter 0 is only partly treated). The dashed line is the static estimate, −0.081 (their Table 1, column 1), drawn over the same quarters.

Figure 2 shows that the estimated reduction in death rates grows after adoption. The coefficients are close to zero before adoption and become more negative over the twelve quarters after it. Had the reform's full effect been realized immediately, as a static specification assumes, the coefficients would be flat, apart from noise. The fused lasso fits five blocks, each more negative than the one before. Its end-of-window effect (defined as the level of the last block) is −0.112, 38 percent larger in absolute value than the static estimate of −0.081. Intuitively, a static specification summarizes these evolving effects with a single "average" effect, combining the smaller initial responses with the larger responses observed later. The static

estimate, the dashed line in the figure, therefore overstates the reduction early in the window and understates it late.

The gap between the static estimate and the end-of-window effect is economically meaningful. By the end of the window, the static estimate misses more than a quarter of the reform's effect on mortality. The end-of-window effect is not estimated precisely enough to rule out the static estimate, however, and we return to this point in Section 3.5.[8] The mean of the last third of the coefficients is larger in absolute value than the mean of the first third, and the difference is significant at 5 percent (Table 2).

3.3 CEOs with Daughters and Corporate Social Responsibility

For our next application, we use data from Cronqvist and Yu (2017), who ask whether CEOs with daughters run more socially responsible firms. Here the event is a change of CEO, and the treatment is whether the incoming CEO has a daughter. Because a firm's record on social responsibility reflects policies that accumulate over time, a new CEO's influence on it may take years to show. Their headline coefficient is an average over the new CEO's years in office. We estimate the effect separately for each year after the change and fit the fused lasso to these coefficients. The estimates suggest that the effect is larger late in the window than early, and that by the end of the window it exceeds their estimate, although the estimates are noisy.

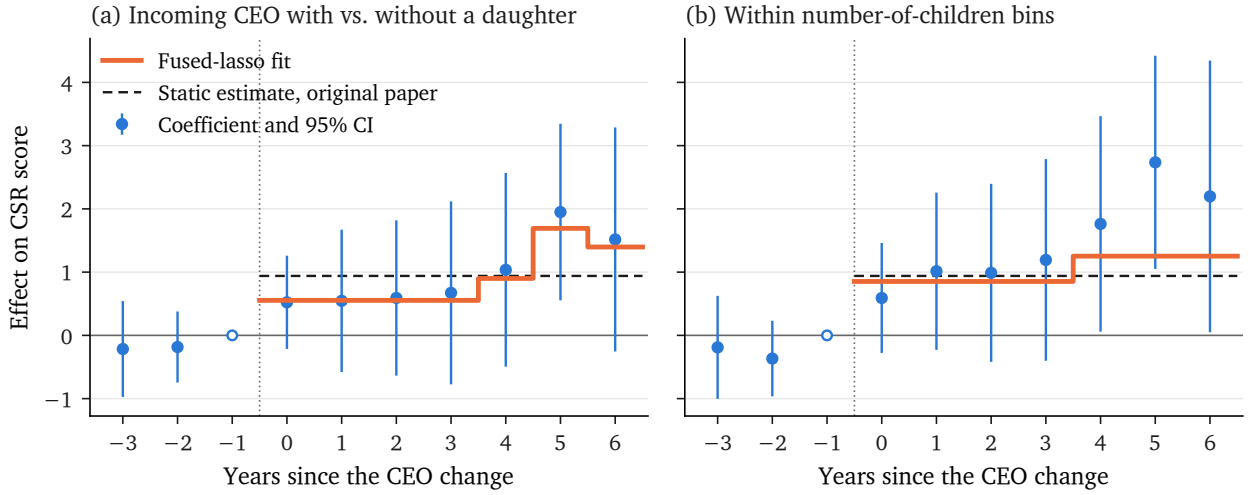
As background, Cronqvist and Yu (2017) study the corporate social responsibility (CSR) of S&P 500 firms over 1992–2012. Their premise is that daughters shape their parents' values. Conditional on the number of children a CEO has, whether any of them is a daughter is a matter of chance, so they treat having a daughter as exogenous. They measure CSR with the KLD (Kinder, Lydenberg, Domini) score, the number of strengths minus the number of concerns across six categories. Their headline estimate (Cronqvist and Yu, 2017, Table 11, column 1) is that a CEO with a daughter is associated with a CSR score that is 0.94 higher, in a panel of 2,662 firm-years. The variation comes from CEO changes within firms, and they compare CEOs who have the same number of children. We estimate an event study around these changes, which gives the coefficient for each year and a variance-covariance matrix that the fused lasso takes as input.

The authors shared with us the data on CEOs' children used in their paper. Our panel has 153 CEO changes at 110 firms, 34 of them to a CEO without a daughter. In our main variant, we compare changes to a CEO with a daughter with changes to a CEO without one, from three years before to six years after the change, with standard errors clustered by firm. Panel (a) of Figure 3 shows this comparison, and Panel (b) repeats it within bins of the incoming

[8] The 95 percent interval for the end-of-window effect is $[-0.190, -0.043]$ and contains the static estimate.

CEO's number of children, comparing CEOs who have the same number of children, as in their headline estimate.

Figure 3: CEOs with daughters and CSR: event studies and fused-lasso fits



Notes: The outcome is the KLD CSR score. Event time is years since the CEO change; year −1 is the reference period, normalized to zero. The sample has 153 CEO changes at 110 firms. Standard errors are clustered by firm, and bars are 95 percent confidence intervals. Panel (a): incoming CEOs with a daughter versus those without. Panel (b): the same comparison within bins of the incoming CEO's number of children. The step line is the fused-lasso fit to the post-change coefficients, years 0–6. The dashed line is the static estimate of 0.94 (Cronqvist and Yu, 2017, Table 11, column 1), from their own panel of 2,662 firm-years.

Figure 3 shows that the effect on CSR is larger later in the new CEO's time in office. In both panels, the coefficients are close to zero before the change and rise after it. A pre-trend would have suggested that firms hire CEOs with daughters in anticipation of higher CSR scores, which is unlikely. The end-of-window effects are 1.40 in Panel (a) and 1.25 in Panel (b), which are 49 and 33 percent larger, respectively, than the static estimate of 0.94. This is consistent with the visual evidence that the effect is increasing over time, resulting in static estimates understating the long-run effect.

The wide confidence intervals in Figure 3 reflect a small sample. Only 34 of the 153 CEO changes are to a CEO without a daughter, and only 12 of these CEOs are still observed in year 6; the others leave the position or reach the end of the data in 2012 first. Even so, the late coefficients are larger than the early ones in both panels. The mean of years 5–6 exceeds the mean of years 0–1 by more than the static estimate itself, and the difference is significant at 10 percent in both panels (Table 2). The bootstrap intervals for the end-of-window effect contain the static estimate, and in Panel (b) the interval excludes zero (Table 1). Taken together, the CSR estimates point to an effect that builds up over the new CEO's tenure, although the effect sizes are noisy.

3.4 Colonial Governors and Patronage

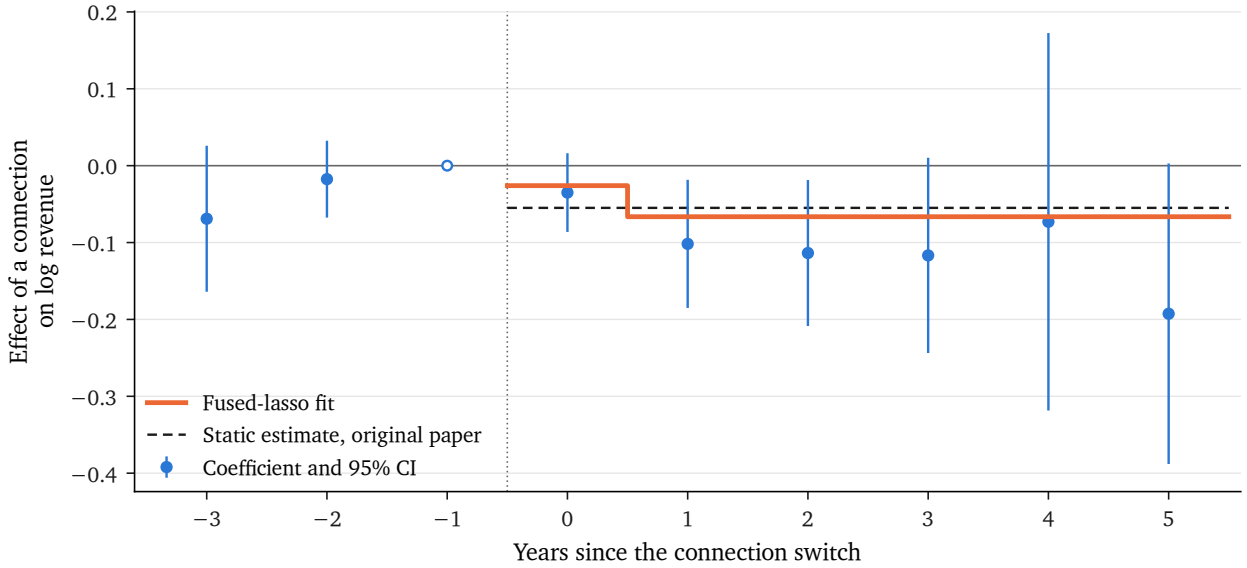
For our last application, we use data from Xu (2018), who studies the costs of patronage in the British Colonial Office. Here the event is a change during the tenure of a manager, rather than a new appointment: a governor gains or loses a connection to his superior in London, the sitting Secretary of State for the Colonies, during a posting. Xu's (2018) estimate is an average over the years in which a governor is connected. We estimate the effect separately for each year after the change and fit the fused lasso to these coefficients. Unlike in the first two applications, the fused lasso fits a single step, one year after the change, and returns a flat path thereafter. The estimates for years 1–3 are close to one another, and those for years 4 and 5 are too noisy to distinguish from them, so the fused lasso produces no further steps.

As background, Xu studies 456 governors of 70 colonies over 1854–1966. Until 1930, governors were appointed at the discretion of the Secretary of State. A governor is connected to the Secretary of State if the two share ancestors, are both aristocrats, both attended Eton, or both studied at Oxford or both at Cambridge. Because the Secretary of State changes while a governor serves a fixed term, a governor can gain or lose a connection during a posting. Xu compares the same governor in the same colony across connected and unconnected years.[9] His estimate for the years before 1930 (his Table 5, column 2) is −0.055. This corresponds to colony revenue about 5.5 percent lower in years when the governor is connected, and he finds no gap after 1930, consistent with the idea that under meritocratic promotions, individuals with connections will continue to work hard. Using his publicly available replication data, we estimate an event study around the 109 changes before 1930, when Xu documents the effect. It runs from three years before to five years after each change. Each window ends at the governor's next connection switch, so that in our sample, every post-event year has the same connection status. This gives the coefficient for each year and the variance-covariance matrix that the fused lasso takes as input.[10]

[9] Over the whole period, only 15 percent of the 729 appointments include such a change.

[10] Following Cengiz et al. (2019), we stack the events with clean controls, and we reverse the sign for events in which the governor loses a connection, so that each coefficient is the effect of gaining one. The design differs from Xu's Table 5 in its controls and clustering, since we cluster by colony and include no spell-length effects.

Figure 4: Colonial governors' connections and colony revenue: event study and fused-lasso fit



Notes: The outcome is log colony revenue. Event time is years since a change in a governor's connection to the Secretary of State during a posting, for the changes before 1930. Changes that end a connection enter with the sign reversed, so each coefficient is the effect of gaining a connection. Each window ends at the governor's next switch. Year −1 is the reference period, normalized to zero. Standard errors are clustered by colony, and bars are 95 percent confidence intervals. The step line is the fused-lasso fit to the post-change coefficients, years 0–5. The dashed line is the static estimate of Xu (2018), −0.055 (his Table 5, column 2).

Here, the fused lasso fits a single step, from the year of the change to the next year, and the fitted effect stays flat after that (Figure 4). Its end-of-window effect is −0.066, 21 percent larger in absolute value than the static estimate of −0.055. Because the path is flat, this gap comes from slightly different designs. Xu (2018) estimates one coefficient for all connected years. To estimate a separate effect for each year, we compare each change with its own control group and end each window at the governor's next change, and with these choices our coefficients average almost twice his estimate. Estimating each year separately also leaves few changes behind each coefficient, only five to nine in the late years, so our estimates are noisy and can differ from his for that reason alone. Our confidence interval for the end-of-window effect includes his estimate, and also zero.

3.5 Taking Stock

Taken together, the three studies suggest that effects can take time to develop, although the pattern differs across settings. In the hospital and CSR applications, the estimated effects grow over the observed window, while in the colonial study, the fitted effect takes a single step and then remains flat. The evidence for gradual build-up is strongest in the hospital study, where the late effects are significantly larger than the early effects at the 5 percent level. The CSR

results point in the same direction but are less precise, while the colonial study provides little evidence of further growth after the initial change (Table 2). One possible explanation is that colonial revenue responds more directly to current tax collection and enforcement, whereas hospital performance and CSR depend more on organizational practices that accumulate over time.

In all four specifications, the end-of-window effects exceed the static estimates reported in the original papers in absolute value (Table 1). The difference is 38 percent for hospitals and 33–49 percent for CSR. However, in all cases, the end-of-window estimates remain relatively noisy: every bootstrap interval contains the static estimate (Table 1).

Finally, the observed windows do not necessarily tell us where the effects settle.[11] If effects are still growing when the data end, as visual inspection suggests may be the case in the hospital and CSR applications, the end-of-window estimates are likely to be conservative estimates of their eventual long-run effects.

Table 1: End-of-window effects and static estimates in the three event studies

	Last block	End-of-window effect	Static estimate	Ratio	Own avg. / static	Bootstrap 95% interval
A. Hospitals (Muñoz and Otero, 2025): log in-hospital death rate, quarters						
Quarters 1–12	11–12	−0.112	−0.081	1.38	1.02	[−0.190, −0.043]
B. CSR (Cronqvist and Yu, 2017): KLD CSR score, years						
Daughter vs. no daughter	6	1.40	0.94	1.49	1.04	[−0.02, 2.92]
Within children bins	4–6	1.25	0.94	1.33	1.59	[0.20, 3.90]
C. Colonial governors (Xu, 2018): log revenue, years						
Before 1930	1–5	−0.066	−0.055	1.21	1.92	[−0.362, 0.005]

Notes: The end-of-window effect is the level of the last block of the fused-lasso fit to the post-event coefficients, with λ chosen by Mallows's C_p (Section 3.1). The static estimates are from Table 1, column 1, of the hospital paper; Table 11, column 1, of the CSR paper; and Table 5, column 2, of the colonial paper. Own avg. / static is the average of the post-event coefficients over the same window divided by the static estimate. The bootstrap interval is the 2.5th to 97.5th percentile of the last block's level over 100,000 draws, repeating the whole procedure, including λ, on each draw with the covariance matrix treated as known. The ratios and intervals treat the static estimate as known. In the colonial row, each window ends at the governor's next connection switch.

[11] We attempted to estimate the long-run effect parametrically by fitting $\theta_k = D\sum_{j=0}^{k}\delta^j$, the path implied by a per-period effect D under AR(1) persistence δ, by minimum distance, using the inverse covariance matrix as the weight and restricting $0 \le \delta \le 1$. Unfortunately, in every specification the confidence set for δ includes $\delta = 1$, at which the long-run effect, $D/(1-\delta)$, is undefined.

Table 2: Tests of build-up within the event window

	Equal coefficients		Late block minus early block			
	Wald χ^2 (df)	p	Blocks	Difference (SE)	p	p over block lengths
A. Hospitals: log in-hospital death rate, quarters						
Quarters 1–12	19.4 (11)	0.054	1–4, 9–12	−0.068 (0.030)	0.022	0.012–0.031
B. CSR: KLD CSR score, years						
Daughter vs. no daughter	9.0 (6)	0.18	0–1, 5–6	1.20 (0.69)	0.080	0.080–0.25
Within children bins	8.3 (6)	0.22	0–1, 5–6	1.66 (0.86)	0.053	0.048–0.14
C. Colonial governors: log revenue, years						
Before 1930	8.6 (5)	0.12	0–1, 4–5	−0.064 (0.095)	0.50	0.098–0.55

Notes: Equal coefficients is a Wald test that all post-event coefficients are equal, using each study's clustered covariance matrix, treated as known. Late block minus early block is the mean of the last m post-event coefficients minus the mean of the first m, with m a third of the coefficients rounded down (four quarters for the hospital study, two years otherwise). The standard error comes from the covariance matrix, and p is from a two-sided z-test. A negative difference in Panels A and C means a larger effect in absolute value late. The last column gives the range of p over block lengths from one to half the number of coefficients, rounded down. The p-values are not adjusted for multiple testing.

4 Persistence: Time-Series Evidence

In various settings, there is no natural way to estimate event studies, and thus a different methodology is needed. In this section, we first show persistence can be parametrically identified in time-series data. Then, we provide evidence for persistence in the setting of Bertrand and Schoar (2003) and show that the effects of managers are larger once persistence is accounted for.

4.1 Identifying Persistence

We specialize the general production process of Section 2 to an AR(1) specification. In this case, only the first lag of output enters directly, and persistence is summarized by a single parameter, δ:

$$y_{t,p} = \delta\, y_{t-1,p} + \theta_{m_{t,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}, \qquad \delta \in [0,1). \tag{8}$$

We first show that δ can be identified from this equation in levels, then discuss estimation and consistency. Our identification argument uses periods in which the same manager remains at a plant between $t-1$ and t, an event we assume occurs with positive probability. Managerial assignment may depend on prior plant and manager histories, subject to the following timing restriction.

Assumption 2. *For each plant p and date $t \geq 2$, the production shock satisfies*

$$\mathbb{E}\left[\epsilon_{t,p} \mid y_{t-2,p}, p, t, m_{t,p}, m_{t-1,p}\right] = 0.$$

The main restriction in Assumption 2 is that output two periods earlier and current hiring, retention, and replacement decisions cannot predict the current shock. This allows shocks to be correlated across adjacent periods, as in a first-order moving average. It also allows better plants to hire better managers and some periods to bring better managers on average. Further, firms can replace managers based on the plant's performance through $t-2$, or through $t-1$ when shocks are serially independent, including after an unexpectedly bad year. What is ruled out is correlation of shocks beyond adjacent periods and decisions that respond to news about the current shock.

We now define the "no manager change" indicator as

$$S_{t,p} := \mathbf{1}\{m_{t,p} = m_{t-1,p}\}.$$

For any variable $V_{t,p}$, let $V^*_{t,p}$ denote the residual from its population regression on plant, time, and current-manager indicators, using observations where $S_{t,p} = 1$. Lagged outputs are regressed on the same indicators as current output, namely those for plant p, date t, and manager $m_{t,p}$. Expectations below are taken across plants and time periods with $t \geq 2$.

Proposition 2 (Identification of δ with Plant and Time Fixed Effects)**.** *Let $y_{t,p}$ satisfy (8) with $0 < \delta < 1$, and define $S_{t,p}$ and $(\cdot)^*$ as above. Suppose Assumption 2 holds, the variables have finite second moments,* $\Pr(S_{t,p} = 1) > 0$*, and*

$$\mathbb{E}\left[y^*_{t-2,p} y^*_{t-1,p} \mid S_{t,p} = 1\right] \neq 0.$$

Then δ is point-identified and satisfies

$$\delta = \frac{\mathbb{E}\left[y^*_{t-2,p} y^*_{t,p} \mid S_{t,p} = 1\right]}{\mathbb{E}\left[y^*_{t-2,p} y^*_{t-1,p} \mid S_{t,p} = 1\right]}. \tag{9}$$

We identify δ using output two periods ago as an instrument for last period's output. We use an instrumental variables estimate because least-squares estimates of Equation (8) can be biased: last period's output may be correlated with the current shock, even under Assumption 2. However, the assumption ensures that output two periods ago is uncorrelated with the current shock after accounting for manager, plant, and date effects. Its covariance with current output therefore equals δ times its covariance with last period's output. Dividing the former

by the latter gives the expression in the proposition.

This division requires a nonzero denominator, so output two periods ago must still predict last period's output after accounting for the fixed effects. When $\delta > 0$, higher output in one period, all else equal, predicts higher output in the next period. A negative association between output two periods ago and last period's shock could weaken this relationship, or eliminate it if it exactly offsets the effect from δ. The first-stage F-statistic tests whether this covariance is zero, and the large values in our main specifications (Table 3) provide strong evidence that it is not.

Practical implementation. Proposition 2 identifies δ from population moments. In a sample, we must also estimate the fixed effects, often from only a few years per manager. We use two estimators. The *levels estimator* is the sample counterpart of the proposition and allows hiring, retention, and replacement to respond to past performance under Assumption 2. However, the short spells in our data introduce Nickell bias with this estimator. *Forward orthogonal deviations* (FOD) avoids this demeaning bias, but its validity requires stronger restrictions on when spells end. Since turnover in our panel responds to performance, we report both estimators and use simulations to assess their bias.

The levels estimator is two-stage least squares of $y_{t,p}$ on $y_{t-1,p}$, instrumented by $y_{t-2,p}$, estimated on stayer observations ($S_{t,p} = 1$), with plant, date, and manager effects absorbed. A CEO's spell at a firm covers only 5.5 years of our panel on average.[12] Estimating the fixed effects from these short spells introduces Nickell bias, which does not disappear simply by adding more firms or more short spells to the panel. To see why, consider the simple case in which each manager has one spell at one plant, tenure lengths are fixed in advance, and shocks are independent over time. Removing the manager and plant effects then means subtracting from $y_{t-2,p}$ the average of $y_{s-2,p}$ over the current manager's stayer years s. That average includes later output affected by the current shock, so subtracting it makes the instrument negatively correlated with the shock. The levels estimator therefore understates δ when the first-stage coefficient is positive, which we observe in our main specifications. Proposition 4 in the Appendix gives an exact expression for the bias at finite T, which is approximately $-(1+\delta)/(\delta T)$ for large T, where T is the number of stayer observations per manager.

FOD subtracts from each variable its average over later usable years of the same spell and rescales the difference (Arellano and Bover, 1995; Alvarez and Arellano, 2003). This removes the CEO and firm effects. We instrument the transformed $y_{t-1,p}$ with the untransformed $y_{t-2,p}$. Leaving the instrument untransformed avoids the Nickell bias described above, while the trans-

[12] A *spell* is a run of consecutive fiscal years in which one CEO runs one firm. It ends when the firm's next year has a different CEO or is missing.

formed error contains only current and later shocks. A sufficient condition for FOD's validity is that these shocks have mean zero conditional on $y_{t-2,p}$ and the spell's timing, a stronger restriction than Assumption 2. This rules out performance-based turnover, because a year's inclusion in FOD depends on the CEO remaining in office for at least one later usable year. Unsurprisingly, turnover in our panel responds to performance in this way (Table 10), so FOD can still be biased. The bias can go in either direction because retention can make lagged output positively or negatively correlated with the difference between the current shock and the average of later shocks. For instance, suppose firms fire CEOs when output falls below a given threshold. Firms with low past output need better current shocks to keep their CEOs than otherwise similar firms with high past output. Since FOD uses a year only if the CEO stays for at least one later year, this selection can make lagged output negatively correlated with the current shock and push the estimate of δ downward.

To assess the size of these biases, we simulate both estimators using the structure of our panel. The levels estimator understates δ in every design, including when turnover is fitted to our data, and the bias can be large (Table 6). As predicted by the theory, FOD has little bias in simulations where CEO assignment is independent of the shocks (Table 6). When turnover is fitted to our data, FOD continues to have little bias in some regressions but understates δ in others. Under a rule that fires every CEO in the bottom tenth of performance, a response far stronger than in our data, FOD overstates δ in some regressions by about 11–12 percent (Appendix B.2). Taken together, these checks lead us to report FOD estimates alongside the conservative levels estimates (Table 3, Panel B).

4.2 Empirical Application: U.S. Public-Company CEOs

We now revisit the variance decomposition of Bertrand and Schoar (2003), whose static specification implicitly assumes $\delta = 0$. Our theory predicts that, when firm outcomes are persistent, this assumption results in understatements of the long-run dispersion of CEO effects (Proposition 1). To assess the magnitude of this understatement, we estimate persistence for five firm outcomes using a panel of U.S. public-company CEOs constructed from ExecuComp and Compustat as in Bertrand and Schoar (2003). Given these estimates of δ, we quantify how much the long-run dispersion of CEO effects exceeds the dispersion recovered by a static regression. Even our conservative estimates imply modest corrections for profitability and large ones for firm scale, due to the greater persistence of scale outcomes.

4.2.1 Data

We merge ExecuComp annual CEO identifiers with Compustat firm-level financials for 1992–2025. As in Bertrand and Schoar (2003), we exclude financial firms (SIC 6000–6999) and utilities (SIC 4900–4999), and we divide the three profitability measures by lagged total assets.[13] Our panel has 47,140 firm-year observations on 3,226 firms and 7,757 CEOs. Each firm-year is a *stayer* observation (the same CEO serves in consecutive years) or a *mover* observation (a new CEO takes over), unless it is the first of a firm's panel or follows a gap.[14] A CEO's spell at a firm, as defined in Section 4.1, lasts 5.5 years on average in the panel, and 433 CEOs appear in more than one firm. Finally, we remind the reader that identifying persistence does not require between-firm movers, only within-firm variation in the lagged dependent variable.

In the language of our model, firms correspond to plants (p) and CEOs to managers (m). We consider five outcomes: return on assets (ROA, income before extraordinary items over lagged total assets), operating ROA (operating cash flow over lagged total assets), EBITDA over lagged total assets, log sales, and log employment. The first three are profitability ratios, and the last two measure firm scale.

4.2.2 Persistence Estimates

We next estimate δ, the parameter governing persistence, for each of the five outcomes. We use the following specification, which nests the standard two-way fixed effects model as the special case $\delta = 0$:

$$y_{t,f} = \delta\, y_{t-1,f} + \alpha_f + \gamma_t + \theta_{m_{t,f}} + \zeta_{i(f),t} + u_{t,f}, \tag{10}$$

where f indexes firms, t indexes fiscal years, $m_{t,f}$ is the CEO in place at firm f in year t, and $\zeta_{i(f),t}$ is a fixed effect for the firm's two-digit SIC industry × year cell. These effects absorb shared industry shocks that could otherwise appear as firm-level persistence, as discussed in Section 4.2.6. We apply the levels and FOD estimators described above and cluster standard errors by firm. Table 3 reports the levels estimates on the stayer subsample in Panel A and the FOD estimates in Panel B.

[13] We winsorize all outcome variables at the 1st and 99th percentiles. We construct the lags used as instruments and as the endogenous regressor from the winsorized series, so the dependent variable, the endogenous regressor, and the instruments share the same support.

[14] Of the firm-years, 38,558 are stayer observations and 5,101 are mover observations. Another 3,481 are the first of a firm's panel or follow a gap in coverage, so their stayer or mover status is undefined.

Table 3: Persistence of five firm outcomes, U.S. CEO-firm panel

	ROA	Operating ROA	EBITDA / Assets	Log sales	Log employment
Panel A. Levels: FE-2SLS on stayer rows					
Estimate $\hat{\delta}$	0.269	0.378	0.338	0.744	0.777
Clustered SE	(0.064)	(0.073)	(0.039)	(0.016)	(0.011)
First-stage F	160	102	415	2,743	4,095
Observations	34,134	34,105	34,087	34,338	33,781
Firms	2,758	2,755	2,755	2,772	2,745
CEOs	5,357	5,353	5,350	5,383	5,319
Panel B. Forward orthogonal deviations, with constant					
Estimate $\hat{\delta}$	0.583	0.707	0.647	0.813	0.872
Clustered SE	(0.040)	(0.045)	(0.028)	(0.025)	(0.021)
First-stage F	402	282	461	208	225
Observations	32,389	32,360	32,338	32,575	32,037
Firms	2,799	2,797	2,794	2,812	2,787
CEOs	5,903	5,899	5,891	5,930	5,859

Notes: An observation is a firm-year. The table reports estimates of δ in Equation (10), with $y_{t-1,f}$ instrumented by $y_{t-2,f}$. ROA, operating ROA, and EBITDA over assets are income before extraordinary items, operating cash flow, and EBITDA, each divided by lagged total assets. A usable year has y_t and y_{t-1} observed. Panel A: levels 2SLS on stayer firm-years (the same CEO at $t-1$ and t), with firm, year, CEO, and SIC2×year fixed effects. Panel B: forward orthogonal deviations within CEO-firm spells, with SIC2×year fixed effects; an observation is a usable year with y_{t-2} observed and a later usable year in the same spell. Standard errors (in parentheses) are clustered by firm. First-stage F is the squared clustered t-statistic of $y_{t-2,f}$. Data: ExecuComp and Compustat, 1992–2025.

In Panel A, the levels estimate of δ is 0.74 for log sales and 0.78 for log employment, against 0.27 to 0.38 for the three profitability measures. Every FOD estimate in Panel B is higher, consistent with a downward bias of the levels estimator, and the ordering across outcomes is the same. Within our framework, these estimates also govern the pace at which a new CEO's impact unfolds. At the levels estimates, a new CEO who permanently replaces a predecessor realizes only 26 percent of the long-run differential in log sales in the first year, against 73 percent for ROA. Because the levels estimates are conservative, these figures likely understate how long it takes managerial effects to approach their long-run magnitude. This is in contrast to the static two-way fixed effects model, which would imply that a CEO's full effect materializes immediately upon appointment.

Under the AR(1) model, a successor's first years reflect both the predecessor's recent growth and the successor's own contribution, since the change in y_t equals δ times the change in y_{t-1} plus new terms. Our estimates of δ therefore imply that a board that assesses a new CEO on one to two years of sales growth overvalues CEOs who succeed a predecessor who grew the firm. The same board undervalues CEOs who succeed a predecessor who shrank it. In the

model, new leadership thus needs time, most strongly for scale outcomes. We return to this point in Section 4.2.5.

Finally, one might naturally expect firm scale to be more persistent than profitability, as we find in the data. A CEO who expands or contracts the firm may leave a lasting imprint on sales, while competitive pressures can erode gains from pricing or cost discipline more quickly. Changes to the workforce may be especially slow to reverse, consistent with the high persistence we estimate for employment.

4.2.3 Heterogeneity in δ across Firms

The point estimates of persistence also vary across firms in intuitive ways (Appendix C.3). Splitting firms at the median of average log employment, we find higher point estimates of δ in larger firms for all five outcomes under both estimators. An additional check using assets to measure size gives the same result for sales and employment but reverses the differences for profitability. While the usual caveats about multiple testing apply, sales and employment are where we would expect this pattern to be more consistent: larger firms have more layers between the CEO and operations, so changes to staffing and the scale of production may take longer to implement or reverse.

4.2.4 How Much CEOs Matter: Revisiting the Bertrand–Schoar Variance Decomposition

Bertrand and Schoar (2003) ask how much CEOs matter using a static regression of firm outcomes on firm, year, and manager effects,

$$y_{t,f} = \alpha_f + \gamma_t + \theta_{m_{t,f}} + \varepsilon_{t,f}, \tag{11}$$

which is our model with $\delta = 0$. The dispersion of the estimated CEO effects, $\widehat{\sigma}_{\text{static}} := \sqrt{\text{Var}(\hat{\theta}_m)}$, then measures the output gain from a CEO who is one standard deviation better. When $\delta > 0$, this gain builds over the CEO's tenure, and the static specification captures only part of it. Proposition 1 gives a closed form: in large samples, $\widehat{\sigma}_{\text{static}}$ converges to $\sigma_{\text{static}} = B_{\text{static}}\sigma_\theta$, with $B_{\text{static}} = \sum_{j\geq 0} \delta^j r_j$. Here r_j measures how strongly the current CEO's quality predicts the quality of the CEO in office j years earlier (Assumption 1). A natural interpretation is the probability that today's CEO was also in office j years earlier. Intuitively, the CEO in office j years ago still contributes δ^j times their quality to today's output, and the static regression credits this contribution to today's CEO only when the two are the same person. Hence σ_{static} lies between two benchmarks. The first-year gain σ_θ includes only the current year's contribution. The long-run gain $\sigma_\theta/(1-\delta)$ includes every past year's.

Scaling σ_{static} by the *multiplier* $M = B_{\text{long run}}/B_{\text{static}}$, where $B_{\text{long run}} = 1/(1-\delta)$, recovers the long-run dispersion $\sigma_\theta/(1-\delta)$. We compute M from $\hat{\delta}$ and the panel's tenure profile. Today's CEO was in office j years earlier exactly when the CEO is in at least the $(j+1)$th year of tenure, so we measure r_j as the share of CEO-years with tenure of at least $j+1$ years. Table 4 reports M at the levels (Panel A) and FOD (Panel B) estimates of δ, with 95 percent confidence intervals.

Table 4: Long-run multipliers of CEO effects

	ROA	Operating ROA	EBITDA / Assets	Log sales	Log employment
Panel A. At the levels estimate					
$\hat{\delta}$	0.269	0.378	0.338	0.744	0.777
Long-run effect $1/(1-\hat{\delta})$	1.368	1.608	1.511	3.902	4.494
Static effect	1.306	1.491	1.418	2.827	3.086
Multiplier M	1.047	1.078	1.066	1.380	1.456
95% interval, lower	1.017	1.031	1.043	1.319	1.400
95% interval, upper	1.078	1.126	1.088	1.442	1.512
Panel B. At the FOD estimate					
$\hat{\delta}$	0.583	0.707	0.647	0.813	0.872
Long-run effect $1/(1-\hat{\delta})$	2.401	3.416	2.832	5.350	7.816
Static effect	2.031	2.598	2.287	3.406	4.135
Multiplier M	1.182	1.315	1.238	1.571	1.890
95% interval, lower	1.123	1.179	1.180	1.390	1.567
95% interval, upper	1.241	1.451	1.296	1.752	2.214

Notes: The long-run effect is $1/(1-\hat{\delta})$ and the static effect is $\sum_{j=0}^{40} \hat{\delta}^j r_j$, both per unit of a CEO's per-period contribution, where r_j is the share of the outcome's CEO-years with tenure of at least $j+1$ years. The static effect is implied by the model, not estimated by a regression. The multiplier M is their ratio (Proposition 1). Tenure counts a CEO's years in office at the firm, back to the CEO's start date where ExecuComp reports it. Panel A uses the levels estimates of δ and Panel B the FOD estimates of Table 3. The 95 percent interval for M uses the delta method. Data: ExecuComp and Compustat, 1992–2025.

At the levels estimates, the long-run dispersion of CEO effects exceeds the static dispersion by 38 percent for log sales and 46 percent for log employment. The correction is smaller for the three profitability measures, at 5 to 8 percent, because profitability is less persistent than firm scale (Corollary 1). Every confidence interval in Table 4 excludes $M = 1$, the value without persistence. Further, because the levels estimator understates δ in our simulations (Appendix B.2) and M increases with δ, the multipliers at the levels estimates are also conservative. At the FOD estimates, the long-run dispersion exceeds the static dispersion by 57 percent for log sales and 89 percent for log employment.

The multiplier gives a simple rule for converting the static dispersion into the long-run one. The rule relies on two simplifying assumptions. First, boards do not keep or replace CEOs

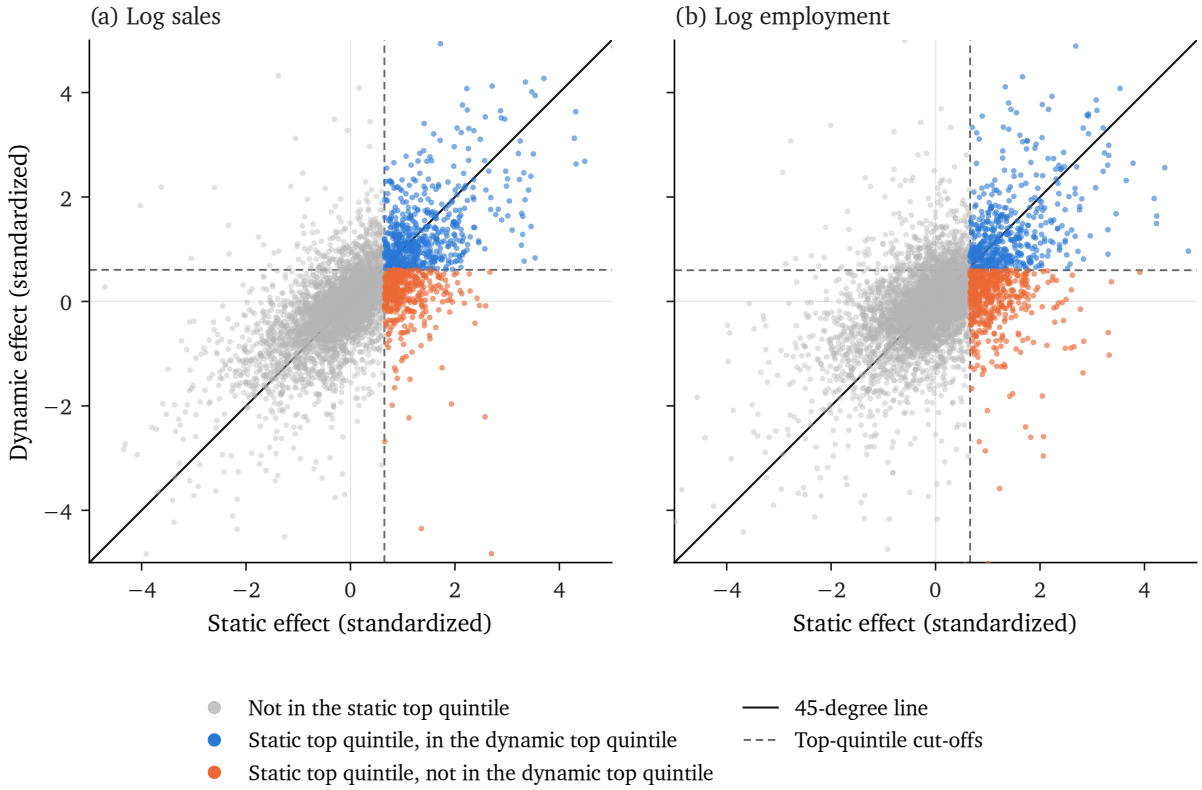
based on the firm's performance. Second, we know each CEO's static effect exactly, whereas in practice each effect is estimated from the CEO's few years in office and is off by some error. We next discuss how to estimate each CEO's long-run effect more directly.

4.2.5 Which CEOs Matter: Static vs. Dynamic Estimates

So far, accounting for persistence has changed how much CEOs appear to matter. With infinite data, it does not change the ordering of which CEOs matter most (Proposition 1). In our panel it does, because we observe most CEOs for only a few years. A CEO's static effect will then be biased by how good or bad the CEO's predecessors were and by how long the CEO stayed. The effects that allow for persistence avoid these biases.

Figure 5 compares these two estimates. The static ranking uses the CEO effects from the specification of Bertrand and Schoar (2003), Equation (11), with industry-by-year effects added. The ranking that allows for persistence uses each CEO's long-run effect, $\theta_m/(1-\delta)$, which we call the *dynamic effect*. Moving $\delta y_{t-1,f}$ to the left-hand side of (10) shows that the CEO effects from a regression of $y_{t,f} - \hat{\delta} y_{t-1,f}$ on firm, year, CEO, and industry-by-year effects estimate θ_m. Dividing them by $1-\hat{\delta}$, with $\hat{\delta}$ the FOD estimate, gives the dynamic effects. The rank correlation between the two is 0.66 for log sales and 0.52 for log employment. Roughly 40 percent of the CEOs in the static top quintile are not in the dynamic top quintile for log sales, and roughly 50 percent for log employment.

Figure 5: Static and dynamic CEO effects within firms



Notes: Each point is a CEO observed for at least two usable years at one firm, in firms with at least two CEOs (4,955 CEOs for log sales and 4,924 for log employment). The horizontal axis is the static effect, the CEO effect from a regression of the outcome on firm, year, CEO, and SIC2×year fixed effects. The vertical axis is the dynamic effect, the CEO effect from the same regression of $y_t - \hat{\delta} y_{t-1}$, divided by $1 - \hat{\delta}$, with $\hat{\delta}$ the FOD estimate of Table 3. Each effect is measured relative to the mean of the firm's CEOs, shrunk in proportion to its sampling noise (empirical Bayes, Morris, 1983), and standardized. The solid line is the 45-degree line, the dashed lines are the top-quintile cutoffs, and orange points are CEOs in the static but not the dynamic top quintile. In Panels (a) and (b), respectively, 23 and 25 points lie outside the plotted range. Data: ExecuComp and Compustat, 1992–2025.

Figure 5 shows that the two estimates disagree about which CEOs are better, but it cannot show which estimate is more accurate. Both estimates are noisy, since each comes from a CEO's few years in office, which may have been lucky or unlucky, and the dynamic effects also contain the estimation error in $\hat{\delta}$. Since we do not observe the CEOs' true effects, we cannot tell in real data whether these errors outweigh the biases of the static effects. Therefore, in Appendix B.3, we use simulations of our model, in which the CEOs' true effects are known, to show that the dynamic effects capture 53 percent of the within-firm variation in the true effects for log sales and 65 percent for log employment, against 37 and 34 percent for the static effects, suggesting that, if our model holds, the dynamic methods are much more accurate at identifying which CEOs are good or bad.

Finally, which CEOs count as good matters well beyond this paper. In research, many studies

use estimated manager effects to ask whether better managers are paid more (e.g., Bertrand and Schoar, 2003), what makes a good manager, and how firms find one. If firm outcomes are persistent, static estimates partly credit CEOs for what they inherited, so these studies may rest on a flawed measure of who is good.[15] In practice, boards face the same problem when they judge CEOs by how the firm performs on their watch. Accounting for persistence changes not only how much CEOs matter, but which ones do.

4.2.6 Alternative Mechanical Explanations for Persistence

A natural concern is that our estimates of δ reflect mechanical features of the data rather than a property of the production process that amplifies managerial effects. We consider three such channels.

A first concern is that δ captures persistence in the economic environment (e.g., installed capital or long-term contracts) rather than persistence in the managerial contribution to output. This interpretation is consistent with our conceptual framework, in which δ is a property of the production process, not of managers per se. Our central argument does not require managers themselves to be persistent. Because the production process is persistent, a manager's per-period contribution accumulates into a long-run effect (Section 2) that exceeds the flow effect. Any channel through which past output feeds into current output, such as capital, organizational routines, or workforce composition, would make managerial effects durable. For instance, serially correlated demand due to the signing of long-term contracts or building relationships is consistent with our interpretation of δ. However, serially correlated demand shocks that are exogenous to the firm are not, and we treat them as a threat in the third concern below.

A second concern is endogenous manager assignment. Poor firm performance strongly predicts CEO departures (Jenter and Lewellen, 2021).[16] The key restriction on selection for the levels estimator is that hiring, retention, and replacement must not systematically predict the current production shock, as formalized in Assumption 2. FOD also requires restrictions on spell endings, as discussed above. Appendix C.1 documents the response of turnover to performance, and Appendix B.2 simulates both estimators with replacement probabilities fitted to the observed turnover rate by performance decile. These simulations quantify how selective

[15] The theoretical results of Dessein and Prat (2022) align with this distinction: in their model, a CEO's later employment and pay in equilibrium depend on the growth of performance and organizational capital at the CEO's previous firm rather than on their level.

[16] Such selection also affects estimates of the variance of CEO effects. In Dessein and Prat (2022), bad CEOs are fired quickly and never rehired, so estimates that rely on CEOs observed for several years or at more than one firm censor the lower tail of CEO quality. This suggests that the dispersions of CEO effects in Bertrand and Schoar (2003) and in our analysis understate their true values.

retention, which Assumption 2 permits, affects both estimators. Appendix C.1.1 discusses the assumption's remaining restriction, that retention not anticipate the current shock.

A third concern is serially correlated unobserved shocks. If demand, technology, or regulatory shocks persist for multiple periods and are not absorbed by the fixed effects, the estimated δ could reflect shock persistence rather than organizational persistence. The bias can then run in either direction, and our instrumental variables strategy addresses this concern in part. On the stayer subsample, the current-manager fixed effect absorbs any time-invariant manager-specific component, and the industry-by-year fixed effects in the CEO panel absorb shocks shared by firms of a two-digit industry. The instrument $y_{t-2,p}$ is valid provided the production innovation $\epsilon_{t,p}$ is uncorrelated with the earlier innovations $\epsilon_{t-2,p}, \epsilon_{t-3,p}, \ldots$ that determine it.[17]

We also verify that FOD with industry-by-year effects has little bias in simulations with first-order moving-average shocks or a persistent industry shock, in which CEO assignment is independent of the shocks (Table 7). Using deeper lags as instruments provides a check on this condition. With FOD, the Hansen J test that $y_{t-2,p}$ and $y_{t-3,p}$ identify the same δ rejects at the 5 percent level for four of the five outcomes (Table 11). The FOD estimates with $y_{t-3,p}$ are also higher than those with $y_{t-2,p}$ for ROA, operating ROA, EBITDA over assets, and log sales (Appendix C.2). If the shocks have a negative autocovariance at the second lag, the FOD estimates understate persistence for four of the five outcomes. Differences in δ across firms could also produce the higher estimates: both instruments can be valid and still identify different averages of the firms' δ, because deeper lags weight the more persistent firms more.

5 Conclusion

This paper develops two econometric tools for measuring the effects of managers when productivity is persistent. With persistence, a better manager's gains build up over time, and static specifications understate both the long-run gain from better managers and the share of output differences attributable to them. The first tool applies to event studies of observed managerial characteristics or policies. We apply the fused lasso of Tibshirani et al. (2005) to the event-study estimates. This allows for a conservative plug-and-play estimate of the long-run effect, without imposing any parametric assumptions. The second tool applies when managerial quality is unobserved and must be inferred from output. We summarize persistence with a single parameter δ, which equals zero in the standard two-way fixed effects model, and show that δ is point-identified by an instrumental variables strategy adapted from Anderson

[17] When retention does not respond to performance, this condition allows a first-order moving average but not an autoregressive process. When firms replace managers after poor performance, the condition must also hold for $\epsilon_{t-1,p}$.

and Hsiao (1982).

We apply both tools to several influential studies estimating the effects of managers. When applied to event studies of hospitals, corporate social responsibility, and colonial governors, the first shows that the estimated effect at the end of each window exceeds the static estimate reported in the original paper in absolute value. In the hospital application and the main CSR design, where the average of the event-study coefficients is close to that static estimate, it is 38 to 49 percent larger in absolute value. These differences are not significant, and if the gains are still building when the data end, the eventual long-run gain may be larger still. When applied to U.S. public-company CEOs, the second shows that persistence varies sharply across outcomes. Profitability measures exhibit lower persistence ($\hat{\delta} \approx 0.27$–$0.38$), while firm scale persists at $\hat{\delta} \approx 0.74$ for log sales and $\hat{\delta} \approx 0.78$ for log employment. These estimates are conservative, and forward orthogonal deviations (FOD), which avoids Nickell bias, gives higher persistence for every outcome. Combined with observed CEO tenures, these estimates imply that the long-run dispersion of CEO effects on firm scale is 38–46 percent larger than the Bertrand and Schoar (2003) variance decomposition implies at the conservative estimates, and 57–89 percent larger at the FOD estimates, so CEOs matter more for firm performance than the static literature reports.

Several open questions follow from this analysis. Persistence itself varies across outcomes and institutional settings, and our results hint at why: outcomes that reflect accumulated institutional capacity and organizational capital (firm scale) are more persistent than outcomes tied to short-run performance (profitability). A more formal analysis of the determinants of persistence is an important direction for future work, as is understanding the channels through which managers build durable organizational capital and the tools that accelerate adjustment after a poorly performing predecessor. In settings with richer plant-level characteristics, one could let δ vary with the firm's environment. Because different types of managers may work in different environments, this would separate how quickly organizations adjust from how much managers themselves differ. Further, with longer time series, one could relax the AR(1) form of the persistence process and estimate richer lag dynamics. At a broader level, the literature has long documented that managers matter; our analysis suggests it has also understated the long-run scale of their influence.

References

Alvarez, Javier and Manuel Arellano, "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 2003, *71* (4), 1121–1159.

Anderson, Theodore Wilbur and Cheng Hsiao, "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 1982, *18* (1), 47–82.

Arellano, Manuel and Olympia Bover, "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 1995, *68* (1), 29–51.

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

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

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

_ , Raffaella Sadun, and John Van Reenen, "Management as a Technology?," Working Paper 22327, National Bureau of Economic Research 2016. Revised October 2017.

Cai, Wei, Andrea Prat, and Jiehang Yu, "Measuring Organizational Capital," Working Paper 35039, National Bureau of Economic Research 2026.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The Effect of Minimum Wages on Low-Wage Jobs," *Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Cronqvist, Henrik and Frank Yu, "Shaped by Their Daughters: Executives, Female Socialization, and Corporate Social Responsibility," *Journal of Financial Economics*, 2017, *126* (3), 543–562.

Dessein, Wouter and Andrea Prat, "Organizational Capital, Corporate Leadership, and Firm Dynamics," *Journal of Political Economy*, 2022, *130* (6), 1477–1536.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter, "Executive Compensation: A Survey of Theory and Evidence," in Benjamin E. Hermalin and Michael S. Weisbach, eds., *The Handbook of the Economics of Corporate Governance*, Vol. 1, Elsevier, 2017, pp. 383–539.

Efron, Bradley, "Estimation and Accuracy After Model Selection," *Journal of the American Statistical Association*, 2014, *109* (507), 991–1007.

Fenizia, Alessandra, "Managers and Productivity in the Public Sector," *Econometrica*, 2022, *90* (3), 1063–1084.

Freyaldenhoven, Simon, Christian Hansen, Jorge Pérez Pérez, and Jesse M. Shapiro, "Visualization, Identification, and Estimation in the Linear Panel Event-Study Design," in Victor Chernozhukov, Johannes Hörner, Eliana La Ferrara, and Iván Werning, eds., *Advances in Economics and Econometrics: Twelfth World Congress*, Vol. 2, Cambridge University Press, 2025, pp. 225–268.

Gabaix, Xavier and Augustin Landier, "Why Has CEO Pay Increased So Much?," *The Quarterly Journal of Economics*, 2008, *123* (1), 49–100.

Giorcelli, Michela, "The Long-Term Effects of Management and Technology Transfers," *American Economic Review*, 2019, *109* (1), 121–152.

Hoffman, Mitchell and Steven Tadelis, "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis," *Journal of Political Economy*, 2021, *129* (1), 243–285.

Holmström, Bengt, "Moral Hazard in Teams," *The Bell Journal of Economics*, 1982, *13* (2), 324–340.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Jenter, Dirk and Katharina Lewellen, "Performance-Induced CEO Turnover," *The Review of Financial Studies*, 2021, *34* (2), 569–617.

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

Minni, Virginia, "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs," *The Quarterly Journal of Economics*, 2026, *141* (3), 1871–1920.

Morris, Carl N., "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 1983, *78* (381), 47–55.

Muñoz, Pablo and Cristóbal Otero, "Managers and Public Hospital Performance," *American Economic Review*, 2025, *115* (11), 4040–4074.

Murphy, Kevin J., "Executive Compensation," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3B, Elsevier, 1999, pp. 2485–2563.

Syverson, Chad, "What Determines Productivity?," *Journal of Economic Literature*, 2011, *49* (2), 326–365.

Tibshirani, Robert, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight, "Sparsity and Smoothness via the Fused Lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, *67* (1), 91–108.

Tibshirani, Ryan J. and Jonathan Taylor, "The Solution Path of the Generalized Lasso," *The Annals of Statistics*, 2011, *39* (3), 1335–1371.

Weidmann, Ben, Joseph Vecci, Farah Said, Sonia Bhalotra, Achyuta Adhvaryu, Anant Nyshadham, Jorge Tamayo, and David J. Deming, "How Do You Identify a Good Manager?," *The Quarterly Journal of Economics*, 2026, *141* (2), 1581–1633.

Xu, Guo, "The Costs of Patronage: Evidence from the British Empire," *American Economic Review*, 2018, *108* (11), 3170–3198.

A Theory

Proof of Proposition 1. We first establish two properties of the weights ψ_j, and then write output as a weighted sum of current and past contributions $c_{s,p} := \theta_{m_{s,p}} + \alpha_p + \gamma_s + \epsilon_{s,p}$, with these weights. The static estimates then follow from the mean of this sum conditional on the current manager, and the long-run effect follows from the sum of the weights. Since each $\delta_k \geq 0$, induction on the recursion gives $\psi_j \geq 0$. Summing the recursion over $j = 1, \ldots, J$ gives $\sum_{j=0}^{J} \psi_j \leq 1 + \sum_k \delta_k \sum_{j=0}^{J} \psi_j$, so the partial sums are bounded by $1/(1 - \sum_k \delta_k)$ and $\sum_j \psi_j$ is finite. Summing the recursion over all $j \geq 1$ then gives $\sum_j \psi_j - 1 = \sum_k \delta_k \sum_j \psi_j$. Hence

$$\psi_j \geq 0, \qquad \sum_{j \geq 0} \psi_j = \frac{1}{1 - \sum_{k \geq 1} \delta_k} = B_{\text{long run}} < \infty.$$

The terms of $B_{\text{static}} = \sum_{j \geq 0} \psi_j r_j$ are nonnegative, $r_0 \psi_0 = 1$, and $r_j \leq 1$, so the bounds $1 \leq B_{\text{static}} \leq B_{\text{long run}}$ in (4) follow.

For the expansion, fix $J \geq 1$ and substitute $c_{s,p} = y_{s,p} - \sum_k \delta_k y_{s-k,p}$, which is (1), into $\sum_{j<J} \psi_j c_{t-j,p}$. By the recursion for ψ_j, the outputs at lags $1, \ldots, J-1$ cancel, which leaves

$$y_{t,p} = \sum_{j=0}^{J-1} \psi_j c_{t-j,p} + R_J, \qquad R_J = \sum_{j=0}^{J-1} \psi_j \sum_{k \geq J-j} \delta_k y_{t-j-k,p}.$$

The weights on the outputs in R_J are nonnegative and sum to $\sum_{j<J} \psi_j \sum_{k \geq J-j} \delta_k$. This sum tends to zero as $J \to \infty$ by dominated convergence, since $\sum_j \psi_j < \infty$ and $\sum_{k \geq m} \delta_k \to 0$ as $m \to \infty$. The expected square of R_J is at most the square of this sum times $\sup_s \mathbb{E}[y_{s,p}^2]$, which is finite because output has bounded mean and variance. Therefore R_J tends to zero in mean square, and

$$y_{t,p} = \sum_{j \geq 0} \psi_j \left(\theta_{m_{t-j,p}} + \alpha_p + \gamma_{t-j} + \epsilon_{t-j,p} \right), \tag{12}$$

where the series converges in mean square.

Taking conditional expectations term by term, which this convergence allows, and applying Assumption 1,

$$\mathbb{E}[y_{t,p} \mid m_{t,p} = m, p, t] = \sum_{j \geq 0} \psi_j \left(\alpha_p + \gamma_{t-j} + a_{j,p} + b_{j,t} \right) + B_{\text{static}} \theta_m. \tag{13}$$

Each partial sum of the first term is additive in plant and date, so its limit can be written as $A_p + G_t$. The conditional mean is therefore the sum of plant, time, and manager fixed effects. Since the manager effects are identified up to a common constant, the population static regression has $\beta_m^{\text{static}} = B_{\text{static}} \theta_m + C$, where C reflects the normalization. Taking the

variance across managers gives

$$\sigma^2_{\text{static}} = \text{Var}_m(B_{\text{static}}\theta_m + C) = B^2_{\text{static}}\sigma^2_\theta.$$

The common constant drops out of the variance. Taking square roots proves (5).

Now let date 1 be the first appointment under an improved hiring rule that raises manager quality by $\Delta\theta$ in every period from then on, holding the preceding history and all other determinants of output fixed. The output difference at date $t \geq 1$ is

$$\Delta y_t = \Delta\theta \sum_{j=0}^{t-1} \psi_j \longrightarrow B_{\text{long run}}\Delta\theta.$$

Setting $\Delta\theta = \sigma_\theta$ gives the eventual gain $B_{\text{long run}}\sigma_\theta$. □

Proof of Corollary 1. The recursion for ψ_j and $\sum_{j\geq 0}\psi_j < \infty$ imply, for $z \in [0,1]$,

$$\sum_{j\geq 0}\psi_j z^j = \left(1 - \sum_{k\geq 1}\delta_k z^k\right)^{-1}.$$

Now evaluate at $z = q$ and $z = 1$, and apply Proposition 1. For AR(1), substitute $\delta_1 = \delta$, $\delta_k = 0$ for $k \geq 2$, and $q = 1 - 1/\mu$. □

The following corollary follows almost immediately from the previous results, as mentioned in Footnote 5.

Corollary 2 (Production as a Sum of Past Manager Qualities)**.** *Replace* (1) *with*

$$y_{t,p} = \theta_{m_{t,p}} + \sum_{j\geq 1} c_j \theta_{m_{t-j,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}, \qquad c_j \geq 0, \qquad \sum_{j\geq 1} c_j < \infty. \tag{14}$$

Under Assumption 1, Proposition 1 holds with

$$B_{\text{static}} = 1 + \sum_{j\geq 1} c_j r_j, \qquad B_{\text{long run}} = 1 + \sum_{j\geq 1} c_j, \qquad 1 \leq B_{\text{static}} \leq B_{\text{long run}}. \tag{15}$$

Proof of Corollary 2. Taking the conditional expectation of (14) gives

$$\mathbb{E}[y_{t,p} \mid m_{t,p} = m, p, t] = \alpha_p + \gamma_t + \sum_{j\geq 1} c_j(a_{j,p} + b_{j,t}) + \left(1 + \sum_{j\geq 1} c_j r_j\right)\theta_m.$$

The first three terms are additive in plant and date, and the coefficient on manager quality is

B_{static}. The same variance argument therefore gives $\sigma_{\text{static}} = B_{\text{static}}\sigma_\theta$. The bounds follow from $c_j \geq 0$ and $0 \leq r_j \leq 1$.

Increasing quality at every date from 1 onward by $\Delta\theta$ raises output at date t by

$$\Delta y_t = \Delta\theta \left(1 + \sum_{j=1}^{t-1} c_j\right) \longrightarrow \left(1 + \sum_{j\geq 1} c_j\right)\Delta\theta. \qquad \square$$

Proof of Proposition 2. Exogeneity. The two manager identities in Assumption 2 determine $S_{t,p}$. The law of iterated expectations therefore gives

$$\mathbb{E}\Big[\epsilon_{t,p} \mid y_{t-2,p}, p, t, m_{t,p}, S_{t,p} = 1\Big] = 0.$$

On the stayer subsample, the current innovation is consequently uncorrelated with the second lag and with each plant, time, and current-manager indicator. Its fitted value in the population regression on these indicators is zero, so $\epsilon^*_{t,p} = \epsilon_{t,p}$. Subtracting the part explained by the indicators from the second lag preserves its orthogonality to the current innovation:

$$\mathbb{E}\Big[y^*_{t-2,p}\epsilon_{t,p} \mid S_{t,p} = 1\Big] = 0. \tag{16}$$

Identification. Applying the same population regressions to the current output equation and using $\epsilon^*_{t,p} = \epsilon_{t,p}$ gives

$$y^*_{t,p} = \delta\, y^*_{t-1,p} + \epsilon_{t,p}. \tag{17}$$

Multiplying by $y^*_{t-2,p}$ and taking expectations,

$$\mathbb{E}\Big[y^*_{t-2,p}y^*_{t,p} \mid S_{t,p} = 1\Big] = \delta\, \mathbb{E}\Big[y^*_{t-2,p}y^*_{t-1,p} \mid S_{t,p} = 1\Big],$$

where the innovation term is zero by (16). Dividing by the nonzero first-stage covariance yields (9). $\square$

A.1 Persistence and the Share of Output Attributable to Managers

With persistence, current output depends not only on the current manager but on every manager the plant has had, through their accumulated contribution $\Theta_{t,p} = \sum_{j\geq 0} \psi_j \theta_{m_{t-j,p}}$. The natural measure of how much of productivity is attributable to managers is therefore the share of output variance due to $\Theta_{t,p}$, after removing plant and date effects.[18]

[18] This measure corresponds to the share of output variance attributable to the manager fixed effects, and is the standard notion in the literature. Proposition 3 allows for production shocks to be correlated with the plant's

Proposition 3 (Biased Manager R^2)**.** *Under the conditions of Proposition 1, let β_m^{static} denote the manager effects from the population-level static regression, and for any variable $V_{t,p}$ let $V_{t,p}^{\dagger}$ denote the residual from its population regression on plant and time indicators. Define*

$$R^2_{\text{static}} := \frac{\mathbb{E}\big[(\beta^{\text{static}\dagger}_{m_{t,p}})^2\big]}{\text{Var}(y_{t,p})}, \qquad R^2_{\text{long run}} := \frac{\mathbb{E}\big[(\Theta^{\dagger}_{t,p})^2\big]}{\text{Var}(y_{t,p})}, \qquad \Theta_{t,p} := \sum_{j\geq 0} \psi_j \, \theta_{m_{t-j,p}}. \tag{18}$$

Then,

$$R^2_{\text{long run}} - R^2_{\text{static}} = \frac{\mathbb{E}\big[\text{Var}(\Theta_{t,p} \mid m_{t,p}, p, t)\big]}{\text{Var}(y_{t,p})} \geq 0,$$

with equality if and only if $\text{Var}(\Theta_{t,p} \mid m_{t,p}, p, t) = 0$.

Proof of Proposition 3. Since both measures share the same denominator, it suffices to compare their numerators. By Assumption 1 and the same limit argument as in the proof of Proposition 1,

$$\mathbb{E}[\Theta_{t,p} \mid m_{t,p} = m, p, t] = \sum_{j\geq 0} \psi_j \left(a_{j,p} + b_{j,t} + r_j \theta_m\right) = A_p + G_t + B_{\text{static}} \theta_m,$$

where A_p and G_t are the limits of the plant and date parts of the sum, defined as in the proof of Proposition 1 but without α_p and γ_{t-j}. Let $\eta_{t,p} := \Theta_{t,p} - \mathbb{E}[\Theta_{t,p} \mid m_{t,p}, p, t]$. Since $\eta_{t,p}$ has mean zero conditional on $(m_{t,p}, p, t)$, it is uncorrelated with every function of $(m_{t,p}, p, t)$. This includes the plant and date indicators, so $\eta^{\dagger}_{t,p} = \eta_{t,p}$. Residualizing also removes $A_p + G_t$. Thus

$$\Theta^{\dagger}_{t,p} = B_{\text{static}} \theta^{\dagger}_{m_{t,p}} + \eta_{t,p}, \qquad \mathbb{E}[(\Theta^{\dagger}_{t,p})^2] = B^2_{\text{static}} \, \mathbb{E}[(\theta^{\dagger}_{m_{t,p}})^2] + \mathbb{E}[\eta^2_{t,p}],$$

where the cross term vanishes because $\theta^{\dagger}_{m_{t,p}}$ is a function of $(m_{t,p}, p, t)$.

For the numerator of R^2_{static}, the proof of Proposition 1 shows that the population fitted values of the static regression are $F_{t,p} = \mathbb{E}[y_{t,p} \mid m_{t,p}, p, t]$, with manager component $\beta^{\text{static}}_{m_{t,p}} = B_{\text{static}} \theta_{m_{t,p}} + C$. Because $y_{t,p} - F_{t,p}$ is orthogonal to all indicators, regressing $y_{t,p}$ or $F_{t,p}$ on plant and date indicators alone gives the same fitted values. Adding manager indicators therefore raises the explained variance by

$$\mathbb{E}[(F^{\dagger}_{t,p})^2] = \mathbb{E}\big[(\beta^{\text{static}\dagger}_{m_{t,p}})^2\big] = B^2_{\text{static}} \, \mathbb{E}[(\theta^{\dagger}_{m_{t,p}})^2],$$

where the constant C drops out because the plant and date indicators absorb it.

Subtracting the two numerators and using $\mathbb{E}[\eta^2_{t,p}] = \mathbb{E}[\text{Var}(\Theta_{t,p} \mid m_{t,p}, p, t)]$ gives the result. The difference is zero if and only if $\eta_{t,p} = 0$ almost surely. □

history of managers, so there may be other effects of managers per se. The measure is therefore easiest to interpret as the share of output attributable to managers absent such correlation.

Proposition 3 shows that the static measure understates this share unless today's manager, together with the plant and date, fully determines the contributions of past managers. The reason is that the static regression explains output using current-manager, plant, and date fixed effects. However, two plants with equally good current managers can perform differently because one benefited from better managers in the past. A static framework cannot attribute this to managers and instead attributes it to the production shocks, causing the understatement. As before, restricting to geometric hiring yields a closed form that shows how the understatement depends on persistence and average tenure.

Corollary 3 (Geometric Assignment and Manager R^2 Bias)**.** *Under the conditions of Proposition 3, suppose $r_j = q^j$, where $q \in (0,1)$, and the within-plant mean and variance of manager quality are constant over time, with positive variance. Then*

$$\frac{R^2_{\text{static}}}{R^2_{\text{long run}}} = \frac{\left[\sum_{j\geq 0} \psi_j q^j\right]^2}{\sum_{j,k\geq 0} \psi_j \psi_k q^{|j-k|}}. \tag{19}$$

For AR(1) production under geometric hiring, with mean tenure $\mu = 1/(1-q)$, this becomes

$$\frac{R^2_{\text{static}}}{R^2_{\text{long run}}} = \frac{1-\delta^2}{1-\delta^2(1-1/\mu)^2}. \tag{20}$$

Proof of Corollary 3. For $j \geq k$, apply Assumption 1 at date $t-k$. Since the within-plant mean and variance of manager quality are constant over time, this gives

$$\text{Cov}(\theta_{m_{t-j,p}}, \theta_{m_{t-k,p}} \mid p,t) = q^{j-k}\,\text{Var}(\theta_{m_{t,p}} \mid p,t).$$

Since the within-plant mean $\mathbb{E}[\theta_{m_{t,p}} \mid p,t]$ depends only on the plant, so does $\mathbb{E}[\Theta_{t,p} \mid p,t]$, and removing plant and date effects removes both means. The two variance numerators in the proof of Proposition 3 therefore share the factor $\mathbb{E}[\text{Var}(\theta_{m_{t,p}} \mid p,t)]$; canceling it gives the first expression. For AR(1), $\psi_j = \delta^j$, so

$$\sum_{j,k\geq 0} \delta^{j+k} q^{|j-k|} = \frac{1}{1-\delta^2}\left(1 + \frac{2\delta q}{1-\delta q}\right) = \frac{1+\delta q}{(1-\delta^2)(1-\delta q)}.$$

Substituting this sum and $q = 1-1/\mu$ gives the AR(1) expression. □

A.2 Nickell Bias of the Levels Estimator

Section 4.1 explains why the levels estimator is biased when spells are short. To remove the manager effect, the estimator subtracts from each instrument $y_{t-2,p}$ its mean over the manager's stayer years. That mean includes later output, which contains part of the current shock, so the demeaned instrument is negatively correlated with the current shock. Proposition 4 measures the resulting bias. To obtain a closed form and a clean proof, it uses a stylized version of the model of Section 4.1: every manager serves one spell of the same length at a single plant, all plants have the same appointment dates, and shocks are independent over time. These assumptions are much stronger than those of Proposition 2 and are not meant to describe our data. After the proof, we explain which of them can be relaxed, and Appendix B.2 studies the bias in simulations that keep the structure of our panel.

Proposition 4 (Nickell Bias of the Levels Estimator)**.** *In the model of Section 4.1, let $\delta > 0$ and suppose that:*

(a) There are N plants with the same nonrandom appointment dates. Each plant has J successive managers, who work only at that plant and serve one spell each: an arrival year followed by $T \geq 3$ stayer years. Output is observed from the year before the first arrival to the end of the last spell. The plants are independent and identically distributed: each plant's effect, its managers' qualities, its output before the first arrival, and its shocks are drawn independently across plants from the same distribution.

(b) Each shock $\epsilon_{t,p}$ has mean zero and variance $\sigma_\epsilon^2 > 0$, and is jointly independent of the plant effect, of earlier output and shocks at that plant, and of the qualities of the managers appointed there by date t. Output, manager qualities, and plant effects have second moments bounded uniformly in T.

Let $\hat{\delta}$ be the levels estimator, and let κ_T be the first-stage covariance: the probability limit, as $N \to \infty$, of the sample covariance between the instrument $y_{t-2,p}$ and the regressor $y_{t-1,p}$ after removing plant, date, and manager effects. If $\kappa_T \neq 0$, then $\delta_T^* := \operatorname{plim}_{N\to\infty} \hat{\delta}$ exists and

(i) $\delta_T^* - \delta = -\dfrac{\sigma_\epsilon^2}{T^2 \kappa_T} \displaystyle\sum_{k=0}^{T-3} (T-2-k)\,\delta^k$*, which is negative whenever $\kappa_T > 0$;*

(ii) as $T \to \infty$, $\kappa_T > 0$ for all large T, and $T(\delta_T^* - \delta) \to -(1+\delta)/\delta$.

Proof. Because each manager works at only one plant, the manager effects absorb the plant effects. Let $\tilde{z}_{t,p}$ denote the instrument $y_{t-2,p}$ after removing manager and date effects on the

stayer observations. Since $\tilde{z}_{t,p}$ is orthogonal in the sample to every manager and date indicator, substituting (8) into the estimator gives

$$\hat{\delta} - \delta = \frac{\sum_{p,t:S_{t,p}=1} \tilde{z}_{t,p}\, \epsilon_{t,p}}{\sum_{p,t:S_{t,p}=1} \tilde{z}_{t,p}\, y_{t-1,p}}.$$

After division by NJT, the denominator converges to κ_T, so it remains to find the limit of the numerator.

Since all plants have the same appointment dates, a given date falls in the same year of the same spell at every plant. Removing manager and date effects is therefore equivalent to subtracting from each variable its mean over the manager's stayer years, and then subtracting the mean of the result across plants at each date. The second step removes the date effects γ_t, whose contribution to output is the same at every plant on a given date. It changes the numerator only through the mean of the shocks across plants at each date, which converges to zero as $N \to \infty$. We may therefore set $\gamma_t = 0$ and, for the numerator, work with the manager-demeaned instrument one spell at a time.

Fix a spell, drop the plant index, and number its arrival year 0, so that its stayer years are $1, \dots, T$. Let $c = (\theta_m + \alpha_p)/(1-\delta)$ be the steady-state level of output under this manager, and let $g = y_{-1} - c$ be the deviation of output from c in the year before the manager arrives. Iterating (8) within the spell gives

$$v_r := y_r - c = \delta^{r+1} g + \sum_{q=0}^{r} \delta^{r-q} \epsilon_q, \qquad -1 \le r \le T. \tag{21}$$

By (b), the spell's shocks are independent of one another and of (c, g).

For part (i), write $z_t = y_{t-2}$ for the instrument and $\bar{z}$ for its mean over the spell's stayer years. By (21), $\mathbb{E}[z_s \epsilon_t] = \sigma_\epsilon^2 \delta^{s-t-2}$ if $s - 2 \ge t$, and $\mathbb{E}[z_s \epsilon_t] = 0$ otherwise. Hence

$$\mathbb{E} \sum_{t=1}^{T} (z_t - \bar{z}) \epsilon_t = -\frac{\sigma_\epsilon^2}{T} \sum_{t=1}^{T-2} \sum_{s=t+2}^{T} \delta^{s-t-2} = -\frac{\sigma_\epsilon^2}{T} S_T, \qquad S_T := \sum_{k=0}^{T-3} (T-2-k)\delta^k.$$

By the law of large numbers across plants, the numerator of $\hat{\delta} - \delta$, divided by NJT, converges to $-\sigma_\epsilon^2 S_T / T^2$. Dividing by κ_T gives (i).

For part (ii), since $\sum_{k \ge 0} \delta^k = 1/(1-\delta)$ and $\sum_{k \ge 0} k\delta^k < \infty$, we have $S_T = T/(1-\delta) + O(1)$. For the denominator, let $C = \sigma_\epsilon^2/(1-\delta^2)$, $a_t = \delta^{t-1}$, and $\bar{a} = T^{-1} \sum_t a_t$. In the limit, subtracting the mean across plants at each date replaces v_r by $v_r - \mathbb{E}[v_r]$, so κ_T depends on the covariances

$$\mathrm{Cov}(v_r, v_s) = C\delta^{|r-s|} + (\mathrm{Var}(g) - C)\delta^{r+s+2},$$

which follow from (21). Let G_T be the mean of $\mathrm{Var}(g)$ over the J spells of a plant. Evaluating the first-stage covariance in each spell with these expressions, and averaging over the J spells, gives

$$\kappa_T = \delta C - \frac{C}{T^2}\sum_{s,t=1}^{T}\delta^{|s-t-1|} + \frac{\delta(G_T - C)}{T}\sum_{t=1}^{T}(a_t - \bar{a})^2.$$

By (b), G_T is bounded uniformly in T. The double sum is at most $T(1+\delta)/(1-\delta)$, and the last sum is at most $1/(1-\delta^2)$. Hence $\kappa_T = \delta C + O(T^{-1})$, which is positive for large T. Substituting into (i),

$$\delta_T^* - \delta = -\frac{\sigma_\epsilon^2/(1-\delta)}{T\,\delta\sigma_\epsilon^2/(1-\delta^2)} + O(T^{-2}) = -\frac{1+\delta}{\delta T} + O(T^{-2}). \qquad \square$$

Part (i) shows where the bias comes from. Its numerator is the covariance between the demeaned instrument and the current shock, which is negative for every tenure length. Its denominator is the first-stage covariance κ_T, which has the same sign as the first-stage coefficient. The levels estimator therefore understates δ whenever the first stage is positive. With very short tenures the first stage can be negative, but in our main specifications it is positive. Together with the simulations of Appendix B.2, this is why we call the levels estimates conservative. The bias does not fall as plants are added, since δ_T^* is already the limit as $N \to \infty$. However, by part (ii), it falls in proportion to $1/T$ as tenures lengthen.

The result extends to tenures of different lengths, as long as appointment dates remain nonrandom, as in (a). In that case, T is replaced by an "average" of the managers' numbers of stayer years; for long tenures, this is simply their mean. Similarly, if a manager works at several plants, T is, qualitatively, the manager's total number of stayer years across those plants. The reason is that the manager effect is estimated from all the years in which the manager is observed, regardless of when or at which plant. The formal argument is more involved, because plant effects are then no longer absorbed by manager effects and must be removed separately, which adds a bias of the same kind that shrinks with the number of years per plant. This is why the proposition does not consider this case.

A.3 Heterogeneous Persistence

We show that heterogeneity in δ can be identified and estimated in two cases: when δ varies with observable plant characteristics, and when it differs between stayers and movers.

A.3.1 Heterogeneity by Observable Characteristics

We first extend the baseline data-generating process to allow the persistence parameter to vary with observable plant characteristics, such as the industry code or firm-size tercile. Let

$X_{t,p} \in \mathscr{X}$ denote an observable characteristic of plant p at time t, where $\mathscr{X}$ is a finite set. We refer to the plant-years with $X_{t,p} = x$ as group x. We modify the output equation from Section 4.1 so that δ becomes a function of this characteristic:

$$y_{t,p} = \delta(X_{t,p})\, y_{t-1,p} + \theta_{m_{t,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}, \qquad t \geq 1. \tag{22}$$

To identify the group-specific persistence $\delta(x)$ for each $x \in \mathscr{X}$, we adapt the argument from Proposition 2. Define the group-specific stayer indicator $S_{t,p}(x) := \mathbf{1}\{m_{t,p} = m_{t-1,p}, X_{t,p} = x, X_{t-1,p} = x\}$, which isolates the subsample where both the manager and the observable characteristic remain the same between $t-1$ and t. Here $(\cdot)^*$ denotes residuals from the same regressions on plant, time, and current-manager indicators as in Proposition 2, but computed within this subsample. Suppose the conditional-mean restriction in Assumption 2 holds after additionally conditioning on $(X_{t,p}, X_{t-1,p})$:

$$\mathbb{E}\Big[\epsilon_{t,p} \mid y_{t-2,p}, p, t, m_{t,p}, m_{t-1,p}, X_{t,p}, X_{t-1,p}\Big] = 0.$$

Since $S_{t,p}(x)$ is a function of $m_{t,p}$, $m_{t-1,p}$, $X_{t,p}$, and $X_{t-1,p}$, all of which are in the conditioning set, restricting to this subsample selects only on conditioning variables. By the law of iterated expectations, the current innovation is therefore uncorrelated with the second lag and with the plant, time, and current-manager indicators on this subsample. The indicators explain none of it, so $\epsilon^*_{t,p} = \epsilon_{t,p}$, and residualizing the second lag preserves orthogonality:

$$\mathbb{E}\Big[y^*_{t-2,p}\epsilon_{t,p} \mid S_{t,p}(x) = 1\Big] = 0.$$

Since $X_{t,p} = x$ on this subsample, accounting for the fixed effects in (22) gives

$$y^*_{t,p} = \delta(x)\, y^*_{t-1,p} + \epsilon_{t,p}. \tag{23}$$

As in Proposition 2, assume that the variables have finite second moments, that $\Pr(S_{t,p}(x) = 1) > 0$, and that the first-stage covariance $\mathbb{E}[y^*_{t-2,p} y^*_{t-1,p} \mid S_{t,p}(x) = 1]$ is nonzero. Multiplying (23) by $y^*_{t-2,p}$, using its orthogonality to the current innovation, and dividing by the nonzero covariance yields

$$\delta(x) = \frac{\mathbb{E}\Big[y^*_{t-2,p} y^*_{t,p} \mid S_{t,p}(x) = 1\Big]}{\mathbb{E}\Big[y^*_{t-2,p} y^*_{t-1,p} \mid S_{t,p}(x) = 1\Big]}. \tag{24}$$

Hence $\delta(x)$ is point-identified for each $x \in \mathscr{X}$, as in Proposition 2.

Since (24) is (9) restricted to group x, the baseline levels estimator carries over directly, applied separately to each group. For a given x, we restrict the estimation to observations

where $X_{t,p} = X_{t-1,p} = x$ and the manager remains the same. We then estimate

$$y_{t,p} = \delta(x)\, y_{t-1,p} + \alpha_p + \gamma_t + \theta_{m_{t,p}} + \epsilon_{t,p}, \tag{25}$$

instrumenting $y_{t-1,p}$ with $y_{t-2,p}$ and absorbing plant, time, and current-manager fixed effects via iterative demeaning.

A.3.2 Heterogeneity by Manager Switches

Persistence may also differ depending on whether the current manager is the same as last period's manager or a new arrival. We now show how both types of persistence can be separately identified. We allow for heterogeneous persistence by assuming that the persistence parameter δ depends on the mover indicator $D_{t,p} := \mathbf{1}\{m_{t,p} \neq m_{t-1,p}\}$:

$$y_{t,p} = \delta(D_{t,p})\, y_{t-1,p} + \theta_{m_{t,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}. \tag{26}$$

We write $\delta_S := \delta(0)$ for stayer persistence and $\delta_M := \delta(1)$ for mover persistence. The current output equation has coefficient δ_S on stayer observations and δ_M on mover observations. We apply the levels argument separately to each subsample, requiring that the instrument predict lagged output after accounting for its fixed effects. We state the result for the mover subsample, where $D_{t,p} = 1$.

For any scalar variable $V_{t,p}$, let $V^{\dagger}_{t,p}$ denote the residual from its population regression on plant, time, and current-manager indicators, computed on the mover subsample. All lags are regressed on the indicators for the current observation, as in Proposition 2.

Proposition 5 (Identification of δ_M on Switch Observations)**.** *Let $y_{t,p}$ satisfy (26) with $\delta_M < 1$. Suppose Assumption 2 holds, the variables have finite second moments,* $\Pr(D_{t,p} = 1) > 0$*, and*

$$\mathbb{E}\Big[\, y^{\dagger}_{t-2,p} y^{\dagger}_{t-1,p} \mid D_{t,p} = 1 \Big] \neq 0.$$

Then δ_M is point-identified and satisfies

$$\delta_M = \frac{\mathbb{E}\Big[\, y^{\dagger}_{t-2,p} y^{\dagger}_{t,p} \mid D_{t,p} = 1 \Big]}{\mathbb{E}\Big[\, y^{\dagger}_{t-2,p} y^{\dagger}_{t-1,p} \mid D_{t,p} = 1 \Big]}. \tag{27}$$

Proof of Proposition 5. The two manager identities in Assumption 2 determine $D_{t,p}$. The law of iterated expectations therefore gives

$$\mathbb{E}\Big[\, \epsilon_{t,p} \mid y_{t-2,p}, p, t, m_{t,p}, D_{t,p} = 1 \Big] = 0.$$

The current innovation is consequently uncorrelated with the second lag and with each plant, time, and current-manager indicator on the mover subsample. Residualizing the instrument preserves this orthogonality. Since the residualized instrument is also orthogonal to the part of the innovation explained by the indicators,

$$\mathbb{E}\Big[y^{\dagger}_{t-2,p}\epsilon^{\dagger}_{t,p} \mid D_{t,p}=1\Big]=\mathbb{E}\Big[y^{\dagger}_{t-2,p}\epsilon_{t,p} \mid D_{t,p}=1\Big]=0. \tag{28}$$

On the mover subsample, accounting for the fixed effects in (26) gives

$$y^{\dagger}_{t,p}=\delta_M\, y^{\dagger}_{t-1,p}+\epsilon^{\dagger}_{t,p}.$$

Multiplying by $y^{\dagger}_{t-2,p}$, taking expectations, and using (28),

$$\mathbb{E}\Big[y^{\dagger}_{t-2,p}y^{\dagger}_{t,p} \mid D_{t,p}=1\Big]=\delta_M\,\mathbb{E}\Big[y^{\dagger}_{t-2,p}y^{\dagger}_{t-1,p} \mid D_{t,p}=1\Big].$$

Dividing by the nonzero first-stage covariance proves (27). □

For practical implementation, the mover persistence δ_M can be estimated by restricting the sample to mover firm-years ($m_{t,p}\neq m_{t-1,p}$) and estimating

$$y_{t,p}=\delta_M\, y_{t-1,p}+\alpha_p+\gamma_t+\theta_{m_{t,p}}+\epsilon_{t,p}, \tag{29}$$

instrumenting $y_{t-1,p}$ with $y_{t-2,p}$ and absorbing plant, time, and current-manager fixed effects via iterative demeaning. In our CEO panel, however, this regression is not feasible. A CEO typically has only one mover year per spell, which the CEO fixed effect absorbs, so almost no mover years remain. We therefore estimate δ_S and δ_M in a single regression on stayer and mover years together, with the fixed effects of Table 3 and an indicator for mover years. In it, $y_{t-1,p}$ enters separately for stayer and mover years, each instrumented by $y_{t-2,p}$ for the same years. Because the CEO fixed effect is now also estimated from the CEO's stayer years, it no longer absorbs the mover year. The two estimates are similar for log sales and log employment, and δ_M is somewhat higher than δ_S for the three profitability measures (estimates not reported).

B Simulations

In this appendix, we use three sets of simulations to check how well our methods recover the true effects and parameters they are designed to measure. First, we ask how well the fused lasso, as we apply it to event studies in Section 3.1, recovers the true effect at the end of

the event window. Second, we ask how well our estimators of persistence recover the true δ when we simulate the panel AR(1) model on the structure of the CEO panel. Third, on panels simulated in the same way, we ask whether the static or the dynamic CEO effects of Figure 5 are closer to the CEOs' true effects (Appendix B.3).

B.1 Event-Study Monte Carlos

Section 3 finds that, in every application, the end-of-window effect is larger in absolute value than the static estimate reported in the original paper (Table 1). Persistence would produce this gap, because a static specification combines the smaller initial responses with the larger responses observed later. But so would a fused lasso that overstates the effect at the end of the window. To check this second explanation, we simulate event studies whose true path of effects we know, and ask how close the fused-lasso estimate comes to the true effect at the end of the window. In no design does it overstate that effect by more than 1 percent on average, and it understates effects that build up after the event.

To keep the exercise realistic, we build every simulation from one of four sets of event-study coefficients. These are the hospital study over quarters 1–12 (Section 3.2), the two CSR comparisons over years 0–6 (Section 3.3), and the colonial study over years 0–5 (Section 3.4). We draw the simulated coefficients using each set's covariance matrix as estimated in Section 3, so they are as noisy, and as correlated with one another, as the actual ones.

True paths. For each set, we consider five true paths of effects. Four follow the AR(1) model of Section 4.1. In that model, the event raises the outcome directly by D in every period from the event onward, and each period also carries over a share δ of the previous period's effect. The effect k periods after the event is then

$$\theta_k = D + \delta\,\theta_{k-1} = D\,(1 + \delta + \cdots + \delta^k) = D\,\frac{1-\delta^{k+1}}{1-\delta}.$$

For the hospital study, $k = 0$ is quarter 1, the first treated quarter. With $\delta = 0$, the effect reaches its full size at once and stays flat, and larger values of δ mean a slower build-up. We use $\delta \in \{0, 0.3, 0.6, 0.9\}$. We set D equal to the estimated effect in the first post-event period, so each of these paths starts at the first actual coefficient. Because that estimate is a single noisy coefficient, we check that it is a sensible scale: in each of the four sets, it is at least one standard error from zero and has the same sign as the average post-event coefficient. The fifth path takes as the truth the step-shaped path that the fused lasso fits to the actual coefficients, so that the simulated effects resemble those in the data.

Estimation. For each of the 20 designs (five paths for each of the four sets), we draw the event-study coefficients 2,000 times from a normal distribution centered on the true path, with that set's estimated covariance matrix. We treat each draw exactly as we treat the actual coefficients in Section 3.1. The fused lasso groups consecutive periods into blocks with a common level, and Mallows's C_p chooses how much to pool. The estimate of the end-of-window effect is the level of the last block. For comparison, we also compute for each draw the average over a late window fixed in advance, which covers the last three or four periods. We compare every estimate with the true effect in the last period of the window, θ_K, which is what the end-of-window effect measures (Section 3.1). The fused lasso uses only the coefficients in the window, so it cannot detect growth after the window ends, resulting in a conservative estimate of the long-run effect. For instance, when $\delta = 0.9$, the true effect is still rising at the end of the window, so even θ_K is smaller in absolute value than the path's steady state $D/(1-\delta)$.

Results. Table 5 reports the fused-lasso estimate (Panel A) and the late-window average (Panel B). Each entry is the average estimate (Mean) or the root mean squared error (RMSE), both as shares of θ_K.

Panel A of Table 5 reports that, with $\delta = 0$, the average fused-lasso estimate is 96–100 percent of the true effect in all four sets. Intuitively, the fused lasso keeps a jump only when the jump improves the fit by more than its penalty. When $\delta = 0$, the true effect does not change after the event, so every jump reflects only noise. The fused lasso therefore pools the periods into a few blocks. With $\delta = 0.9$, more jumps pass this test, and it fits more blocks in all four sets, so the procedure responds to the shape of the path.[19]

In all 16 designs with a build-up (i.e., every path except $\delta = 0$), the average fused-lasso estimate falls short of the true effect (Panel A of Table 5). In 12 of them, it is 70–90 percent of the true effect.

In Panel B of Table 5, the average over a late window fixed in advance is closer to the true effect than the fused-lasso estimate in all 16 designs with a build-up and has a lower RMSE in all four with $\delta = 0.9$ but a higher one in all eight with δ of 0 or 0.3. Its weakness is that the researcher must choose the window, which makes it susceptible to p-hacking, and the best choice depends on the unknown path.

Taken together, our simulations suggest that the fused lasso gives a rule-based estimate of the end-of-window effect, with the blocks chosen by the data rather than by the researcher. Because no design overstates the effect by more than 1 percent on average, the end-of-window effects we estimate in our applications are unlikely, on average, to be overstated by more than a

[19] Across sets, the average number of blocks rises from 1.6–2.2 when $\delta = 0$ to 2.4–8.4 when $\delta = 0.9$.

few percent, and may be understated, provided the true effects do not fade over the window.[20]

Table 5: Simulated estimates of the end-of-window effect

	$\delta = 0$		$\delta = 0.3$		$\delta = 0.6$		$\delta = 0.9$		Fitted path	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
A. Fused-lasso estimate: level of the last block										
Hospital, quarters 1–12	0.99	0.66	0.88	0.51	0.84	0.40	0.91	0.19	0.82	0.37
CSR, daughter vs. no daughter	0.97	1.04	0.85	0.73	0.69	0.58	0.74	0.45	0.89	0.50
CSR, within children bins	1.00	1.11	0.85	0.81	0.72	0.60	0.72	0.48	0.89	0.55
Colonial, before 1930	0.96	1.51	0.81	1.20	0.68	0.83	0.61	0.66	0.73	0.99
B. Average of the late-window coefficients										
Hospital, quarters 1–12	0.99	1.00	1.00	0.71	1.03	0.40	0.93	0.16	0.91	0.35
CSR, daughter vs. no daughter	0.98	1.39	1.02	0.95	0.96	0.58	0.89	0.29	0.95	0.51
CSR, within children bins	0.98	1.43	1.00	1.00	0.98	0.59	0.89	0.30	1.01	0.67
Colonial, before 1930	0.96	2.32	0.96	1.67	0.95	1.00	0.86	0.52	0.97	1.21

Notes: Each row is one of the four sets of event-study coefficients of Section 3: the hospital study (quarters 1–12), the two CSR comparisons, and the colonial study. Panel A reports the fused-lasso estimate, computed on each draw as in Section 3.1, and Panel B the average of the coefficients over a late window fixed in advance: quarters 9–12 for the hospital study, years 4–6 for CSR, and years 3–5 for the colonial study. Each entry summarizes 2,000 draws of the event-study coefficients from a normal distribution centered on the true path, with that set's estimated covariance matrix. Mean is the average estimate divided by the true effect in the last period of the window, θ_K; RMSE is the root mean squared error divided by $|\theta_K|$. The δ columns use the AR(1) path $\theta_k = D(1-\delta^{k+1})/(1-\delta)$, where D is the estimated first post-event coefficient, δ the share of each period's effect carried into the next, and $k = 0$ the first post-event period (quarter 1 for the hospital study). The last column takes the fused-lasso fit to the actual coefficients as the true path.

B.2 Panel-Data Monte Carlos

Next, we check how well the two estimators of Table 3 recover δ in simulated data with the structure of our CEO panel. The answer matters because the multipliers of Section 4.2.4 increase with δ. The simulations support two conclusions. First, in every design we simulate, the levels estimator understates δ, which is why we call the levels estimates in Panel A of Table 3 conservative for the outcomes we simulate. Second, FOD with industry-by-year effects has little bias when CEO assignment is independent of the shocks, except in the fitted ROA model below. Under turnover fitted to the data, FOD has little bias for log sales and understates δ for ROA (Table 8).

We simulate only the outcomes, from the AR(1) model of Section 4.1, and take everything else from the CEO panel. We keep its 3,226 firms, their fiscal years and gaps, and the years in which each outcome is missing. We also keep the CEO in office in every firm-year (8,582 CEO-firm spells held by 7,757 CEOs). Every simulated CEO thus remains in office as long as

[20] A single estimate can still be far from the truth, as the RMSE columns of Table 5 report. One caveat is that the simulations treat each study's estimated covariance matrix as known, so they omit the extra noise from estimating it with a limited number of clusters.

the real one did, so both estimators select their samples from the simulated data as they do from the real data. In the baseline, this CEO schedule is that of the data and is independent of the simulated shocks.

Design. Each simulated outcome follows

$$y_{t,f} = \delta\, y_{t-1,f} + \theta_{m_{t,f}} + \alpha_f + \gamma_t + \epsilon_{t,f}.$$

In the baseline, the CEO effects θ_m, firm effects α_f, year effects γ_t, and shocks $\epsilon_{t,f}$ are normal, and the shocks are independent over time. A CEO's effect is independent of the firms the CEO manages, and a CEO who moves keeps the same effect. To give each firm a history, the process runs for 100 years before the firm's first panel year, with CEOs turning over at the panel's annual rate of 11.7 percent (5,101 CEO changes over 43,659 firm-years with a prior-year row).

To set the variances of the CEO effects, firm effects, year effects, and shocks, we match moments of our data at the FOD estimates of Table 3, separately for ROA and log sales, one profitability and one scale outcome, and hold them fixed. We then set the true δ to 0.3, 0.5, 0.75, or 0.9. We also set it equal to the FOD estimate itself (0.583 for ROA and 0.813 for log sales). Because FOD is nearly unbiased in these simulations, this design reproduces our FOD estimate. If the model fit the data well, it would also reproduce our levels estimate, and it does so for ROA but not for log sales.[21] For log sales, no δ in the baseline reproduces both estimates, so for that outcome we also check the two conclusions in the alternative designs and the models fitted to our data below.

Estimators. We apply both estimators of Table 3 to every simulated panel, with industry-by-year effects unless noted and standard errors clustered by firm. We compute the levels estimator (Panel A) as two-stage least squares of $y_{t,f}$ on $y_{t-1,f}$, instrumented by $y_{t-2,f}$, on stayer firm-years, absorbing firm, year, CEO, and industry-by-year effects. The FOD estimator (Panel B) applies forward orthogonal deviations within each CEO's spell at a firm (Alvarez and Arellano, 2003). It measures each year's outcome and its lag as deviations from the average of the later years of the same spell, which removes the CEO and firm effects because they are constant within a spell. We instrument the transformed lag with the untransformed $y_{t-2,f}$ and partial out the transformed industry-by-year dummies and a constant.

[21] In Table 3, the levels estimates are 0.269 for ROA and 0.744 for log sales. At the FOD estimates, the simulated levels estimate averages 0.240 and 0.613, and the gap between the two estimators is 0.34 and 0.19 in the simulations, against 0.31 and 0.07 in our data. The first stage of the levels estimator is stronger in the baseline than in our data for both outcomes. At the FOD estimates, its mean F is 3,553 for ROA and 10,052 for log sales, compared with 160 and 2,743 in Table 3.

Results. Table 6 reports that the levels estimator understates δ in every row, by 0.14–1.00, and by more when true persistence is lower, in line with Proposition 4. It also understates δ in every alternative design of Table 7, when turnover responds to performance (Table 8), and in the models fitted to our data below, at each model's own δ and at the FOD estimate of Table 3.

In the baseline, the bias of FOD is between -0.012 and -0.004 in every row of Table 6, and its 95 percent intervals cover the true δ in 92–96 percent of replications. It stays close to unbiased in every alternative design of Table 7, including shocks that are correlated over time and a persistent shock shared by all firms in an industry. With year effects only, however, FOD cannot separate that industry shock from persistence at the firm and overstates δ, which is why our specification includes industry-by-year effects.

Models fitted to our data. Appendix C.2 suggests that the shocks may be correlated at the second lag but not beyond. We therefore also simulate models whose shocks have a second-lag autocovariance, fitted by simulated method of moments to the dynamic moments of ROA, log sales, and log employment. In every fitted model, the levels estimator understates δ at the model's own δ and at the FOD estimate of Table 3, and FOD is close to unbiased except for ROA, which it understates by about 0.14. The ROA model passes the overidentification test, but the log-sales and log-employment models do not ($p = 0.015$ and $p = 0.0015$), so they cannot match all the moments at once. Each fitted model also implies a value of δ. We use these models only to check the two estimators, so our estimates of δ remain those of Table 3, and the multipliers of Section 4.2.4 are computed from them.

Table 6: Simulated estimates of δ on the CEO panel

	Levels FE-2SLS				FOD + constant			
True δ	Bias	RMSE	Cov.	Mean F	Bias	RMSE	Cov.	Mean F
Panel A. ROA calibration								
0.30	−0.982	0.984	0.00	355	−0.011	0.027	0.93	1,386
0.50	−0.426	0.427	0.00	2,206	−0.008	0.019	0.93	1,830
0.75	−0.230	0.230	0.00	7,845	−0.005	0.015	0.94	1,021
0.90	−0.142	0.142	0.00	15,710	−0.004	0.017	0.94	290
0.583 (data)	−0.343	0.343	0.00	3,553	−0.006	0.018	0.92	1,681
Panel B. Log-sales calibration								
0.30	−0.999	1.001	0.00	347	−0.012	0.045	0.95	332
0.50	−0.434	0.435	0.00	2,164	−0.008	0.036	0.94	370
0.75	−0.237	0.237	0.00	7,626	−0.007	0.035	0.96	184
0.90	−0.149	0.149	0.00	14,927	−0.005	0.057	0.96	35
0.813 (data)	−0.200	0.201	0.00	10,052	−0.006	0.040	0.95	115

Notes: Each row summarizes 3,000 simulated panels of the AR(1) model of Section 4.1, with the CEO panel's firms, years, gaps, and CEO spells, iid normal shocks, and CEO assignment independent of the shocks. Panel A uses the ROA calibration and Panel B the log-sales calibration. Rows give the true δ; (data) marks the FOD estimate of Table 3. Both estimators include SIC2×year fixed effects. Bias is the mean estimate minus the true δ, RMSE the root mean squared error, Cov. the coverage of the nominal 95 percent interval with standard errors clustered by firm, and Mean F the average first-stage F.

Table 7: Simulated estimates of δ under alternative designs

	Levels FE-2SLS		FOD + constant			
	SIC2 × year		SIC2 × year		Year only	
Design	Bias	Cov.	Bias	Cov.	Bias	Cov.
Panel A. ROA calibration, $\delta = 0.5$						
iid shocks (baseline)	−0.426	0.00	−0.008	0.93	0.000	0.95
MA(1) shocks	−0.385	0.00	−0.007	0.91	0.000	0.94
Industry-year AR(1), calibrated†	−0.421	0.00	−0.007	0.93	0.044	0.24
Industry-year AR(1), stress	−0.371	0.00	−0.007	0.94	0.264	0.00
$\sigma^2_\theta/\sigma^2_\varepsilon = 0.1$	−0.502	0.00	−0.009	0.92	0.000	0.94
$\sigma^2_\theta/\sigma^2_\varepsilon = 1.0$	−0.286	0.00	−0.004	0.94	0.000	0.95
Firm-specific variance†	−0.428	0.00	−0.009	0.94	−0.002	0.96
Components recalibrated at δ	−0.418	0.00	−0.007	0.93	0.000	0.95
First CEO in office through burn-in	−0.426	0.00	−0.007	0.94	0.000	0.95
Panel B. Log-sales calibration, $\delta = 0.75$						
iid shocks (baseline)	−0.237	0.00	−0.007	0.96	−0.001	0.96
MA(1) shocks	−0.261	0.00	−0.006	0.95	0.000	0.95
Industry-year AR(1), calibrated†	−0.230	0.00	−0.005	0.95	0.080	0.23
Industry-year AR(1), stress	−0.203	0.00	−0.005	0.95	0.198	0.00
$\sigma^2_\theta/\sigma^2_\varepsilon = 0.1$	−0.302	0.00	−0.007	0.95	0.000	0.95
$\sigma^2_\theta/\sigma^2_\varepsilon = 1.0$	−0.128	0.00	−0.003	0.94	0.000	0.94
Firm-specific variance†	−0.237	0.00	−0.007	0.95	−0.001	0.95
Components recalibrated at δ	−0.226	0.00	−0.006	0.95	−0.001	0.96
First CEO in office through burn-in	−0.239	0.00	−0.005	0.95	−0.001	0.95

Notes: Simulations as in Table 6, with true $\delta = 0.5$ in the ROA calibration (Panel A) and 0.75 in the log-sales calibration (Panel B). Each row changes one feature of the baseline: MA(1) shocks with coefficient 0.3; an industry-year AR(1) shock with persistence 0.8, at its calibrated variance or at a stress variance equal to that of the iid shock; the variance of CEO effects relative to that of the shocks (calibrated at 0.311 for ROA and 0.283 for log sales); a lognormal firm-level scale of the shock variance; variance components re-estimated at the row's δ; or the first CEO in office throughout the 100 simulated years before the panel. †: variance calibrated to our data. The third estimator is FOD with year effects only. Each cell has 1,000 or 3,000 replications, and the columns are defined as in Table 6.

Performance-driven turnover. We again keep the panel's firms, years, gaps, and industries, but now simulate the CEO schedule along with the outcome, so that whether a CEO is replaced can depend on the firm's simulated performance. In the first rule, a CEO is replaced with a probability that depends on the firm's performance in the previous year, relative to the year effects and the firm's steady state. Specifically, we fit this hazard to the overall turnover rate in our data and to the turnover rate by decile of industry-adjusted performance. In the second rule, the board also observes part of year t's shock, and the successor holds office that year. We use ROA and log-sales calibrations.

Table 8 reports that performance-driven turnover moves the bias of the levels estimator by approximately 0.01 or less, except in the two stress cases for ROA: by 0.02–0.03 when CEOs are fired only below a cutoff far in the left tail of performance, and by about 0.05 when every CEO whose performance falls in the bottom tenth is fired. Under turnover fitted to the data, FOD understates δ for ROA, but only by 0.03–0.04 against a true δ of 0.58, so it errs on the conservative side. Because FOD is so precise, however, even this small bias is about twice its standard error, and its 95 percent intervals cover the true δ in only 28–45 percent of replications. The intuition behind this bias is that an FOD equation for year t exists only if the CEO is in office in $t+1$, which depends on the shock in t. How large this bias is depends on how boards respond to poor performance, which the turnover we observe cannot fully pin down. Consider two rules that both match how often CEOs leave in our data and how strongly their departures respond to performance. If the chance of replacement rises steadily as performance worsens, FOD is close to unbiased for ROA. If boards replace CEOs only after very poor years, it understates δ by 0.08–0.10.[22] The bias is largest under a rule far harsher than turnover in our data, in which every CEO whose performance falls in the bottom tenth is fired: FOD then understates δ by 0.18–0.24 for ROA and overstates it by 0.09–0.10 for log sales. In every design of Table 8, FOD has a lower RMSE than the levels estimator.

[22] Turnover under our fitted rule responds to performance about 70–77 percent as strongly as in our data for ROA, and more strongly for log sales.

Table 8: Simulated estimates of δ with performance-driven turnover

	Levels FE-2SLS			FOD + constant					
	SIC2 × year			SIC2 × year			Year only		
Design	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.
Panel A. ROA calibration, $\delta = 0.5834$									
iid; exits exogenous	−0.328	0.329	0.00	−0.007	0.018	0.92	−0.001	0.016	0.95
iid; exits on performance in $t-1$	−0.319	0.319	0.00	−0.038	0.041	0.28	−0.033	0.036	0.41
iid; exits on $t-1$ and year-t news	−0.321	0.321	0.00	−0.031	0.035	0.45	−0.025	0.029	0.60
MA(1); exits exogenous	−0.323	0.324	0.00	−0.007	0.014	0.91	0.000	0.012	0.95
MA(1); exits on performance in $t-1$	−0.317	0.317	0.00	−0.030	0.033	0.28	−0.024	0.027	0.44
MA(1); exits on $t-1$ and year-t news	−0.318	0.319	0.00	−0.025	0.028	0.44	−0.019	0.022	0.62
iid; exits on $t-1$, linear hazard	−0.328	0.328	0.00	0.002	0.016	0.95	0.008	0.017	0.92
MA(1); exits on $t-1$, linear hazard	−0.321	0.322	0.00	0.000	0.012	0.94	0.007	0.014	0.90
iid; exits on $t-1$, cutoff (stress)	−0.295	0.295	0.00	−0.100	0.101	0.00	−0.094	0.095	0.00
MA(1); exits on $t-1$, cutoff (stress)	−0.302	0.303	0.00	−0.076	0.077	0.00	−0.070	0.071	0.00
iid; exits on $t-1$, bottom tenth fired	−0.274	0.274	0.00	−0.241	0.241	0.00	−0.235	0.236	0.00
MA(1); exits on $t-1$, bottom tenth fired	−0.277	0.277	0.00	−0.178	0.179	0.00	−0.173	0.174	0.00
Panel B. Log-sales calibration, $\delta = 0.8131$									
iid; exits exogenous	−0.192	0.192	0.00	−0.006	0.040	0.95	0.000	0.038	0.95
iid; exits on performance in $t-1$	−0.191	0.191	0.00	−0.002	0.039	0.95	0.004	0.038	0.95
iid; exits on $t-1$ and year-t news	−0.190	0.190	0.00	−0.002	0.040	0.95	0.003	0.038	0.95
MA(1); exits exogenous	−0.220	0.220	0.00	−0.006	0.030	0.95	−0.001	0.029	0.96
MA(1); exits on performance in $t-1$	−0.220	0.220	0.00	−0.003	0.030	0.95	0.002	0.030	0.95
MA(1); exits on $t-1$ and year-t news	−0.220	0.220	0.00	−0.004	0.030	0.95	0.001	0.029	0.95
iid; exits on $t-1$, linear hazard	−0.191	0.191	0.00	−0.002	0.042	0.92	0.002	0.041	0.94
MA(1); exits on $t-1$, linear hazard	−0.221	0.221	0.00	−0.003	0.031	0.92	0.002	0.031	0.95
iid; exits on $t-1$, cutoff (stress)	−0.190	0.190	0.00	0.026	0.054	0.89	0.031	0.055	0.85
MA(1); exits on $t-1$, cutoff (stress)	−0.218	0.218	0.00	0.017	0.037	0.91	0.022	0.039	0.85
iid; exits on $t-1$, bottom tenth fired	−0.190	0.190	0.00	0.099	0.115	0.56	0.106	0.120	0.51
MA(1); exits on $t-1$, bottom tenth fired	−0.211	0.212	0.00	0.089	0.098	0.40	0.095	0.103	0.33

Notes: Simulations as in Table 6, except that each firm's CEO schedule is simulated together with the outcome, after 100 pre-sample years. Panel A uses the ROA calibration (true $\delta = 0.5834$) and Panel B the log-sales calibration (true $\delta = 0.8131$). Shocks are iid or MA(1) with coefficient 0.3. Exits are exogenous, with probability 0.117 a year, or depend on performance in $t-1$, measured by the level of ROA or the growth of log sales: through a logistic hazard fitted to the turnover rate by decile of industry-adjusted performance in our data; through a linear hazard or a firing cutoff in the far left tail, each matched to the overall turnover rate and the linear probability coefficient; or by firing every CEO whose performance falls in the bottom tenth, with other exits at a constant rate that keeps the overall rate near 0.117. The cutoff and bottom-tenth rows are stress cases. In the news rows, the board also observes the first half of year t's shock and the successor holds office in year t. Each cell has 1,000 replications, except 300 in the linear and cutoff rows of Panel B. Estimators are as in Table 7, and Bias, RMSE, and Cov. are defined as in Table 6.

B.3 Static and Dynamic CEO Effects against the Truth

Figure 5 shows that the static and dynamic effects rank a firm's CEOs differently, but our data cannot show which ranking is closer to the CEOs' true effects. In panels simulated from our model, we know the true effects, so we can compare both estimates with them.

Design. We simulate the model of Appendix B.2, now also for log employment. For each outcome, the true δ is the FOD estimate of Table 3, and the variances are matched to that outcome's data at that value. CEO effects are drawn independently of firms, predecessors, and shocks. Besides the baseline, with iid shocks, we simulate three designs of Table 7: MA(1) shocks with coefficient 0.3, a lognormal firm-level scale of the shock variance, and an industry-year AR(1) shock with persistence 0.8 at its calibrated variance.

Estimates. On each simulated panel we repeat every step behind Figure 5. We estimate δ by FOD on that panel, so the dynamic effects contain the error in $\hat{\delta}$. We estimate the static and dynamic CEO effects on the same rows, measure each relative to the mean of the firm's CEOs, and shrink it by empirical Bayes. As in Figure 5, we compare the estimates for CEOs observed for at least two usable years at one firm, in firms with at least two CEOs. For each estimate, we compute the within-firm R^2: the squared correlation, across these CEOs, between the estimates and the true effects θ_m, both measured relative to the firm's mean. A correlation does not change when either variable is rescaled, so dividing by $1-\hat{\delta}$ or $1-\delta$ leaves the R^2 unchanged. The R^2 therefore measures how accurately each estimate compares a firm's CEOs with one another, regardless of the estimate's scale.

Results. With iid shocks, the static effects capture 37 percent of the within-firm variation in the true effects for log sales and 34 percent for log employment, and the dynamic effects capture 53 and 65 percent (Table 9). The dynamic effects do better in each of the 1,000 panels for both outcomes. The ordering is the same in the other three designs and for ROA, where the gap is small. Of the CEOs whose true effects are in the top quintile, the static effects place 51 percent in their own top quintile for log sales and 50 percent for log employment, and the dynamic effects place 59 and 65 percent.

Over a CEO's years in office, the static effect also picks up what remains of the predecessors' contributions, persistent luck from earlier years, and a build-up of the CEO's own contribution that depends on how long the CEO stayed. The dynamic effects remove these terms by subtracting $\hat{\delta} y_{t-1,f}$, at the cost of the error in $\hat{\delta}$, and in these simulations the gain exceeds the cost. The comparison holds in our model at the FOD estimate of δ, in which a CEO's contribution persists at the same rate in every firm and turnover does not respond to performance.

Table 9: Accuracy of static and dynamic CEO effects in simulated panels

Outcome	Shocks	Share of variation in true effects captured: Static	Share of variation in true effects captured: Dynamic	Panels in which dynamic is higher (%)
Log sales	iid	0.37	0.53	100
	MA(1)	0.31	0.46	100
	Firm-specific variance	0.37	0.53	100
	Industry-year shock	0.36	0.53	100
Log employment	iid	0.34	0.65	100
	MA(1)	0.29	0.59	100
	Firm-specific variance	0.33	0.65	100
	Industry-year shock	0.33	0.65	100
ROA	iid	0.53	0.56	99.9
	MA(1)	0.44	0.49	100
	Firm-specific variance	0.53	0.56	97.5
	Industry-year shock	0.53	0.56	100

Notes: Each row averages the simulated panels of the model of Appendix B.2 in which both shares are defined: 1,000 per row, except 999 for log sales with firm-specific variance. The panels have the true δ at the outcome's FOD estimate in Table 3 and variances matched to the outcome's data. Shocks: iid normal; MA(1) with coefficient 0.3; a lognormal firm-level scale of the shock variance; iid plus an industry-year AR(1) shock with persistence 0.8 (Table 7). The static effects come from Equation (11) with SIC2×year fixed effects and the dynamic effects from the same regression of $y_t - \hat{\delta} y_{t-1}$, with $\hat{\delta}$ the FOD estimate on the simulated panel. Both are measured relative to the mean of the firm's CEOs and shrunk by empirical Bayes, as in Figure 5, for CEOs observed for at least two usable years at one firm, in firms with at least two CEOs. The share is the squared correlation, pooled across these CEOs, between an estimate and the true effect, both measured relative to the firm mean. In every row, the Monte Carlo standard error of the difference between the two shares is below 0.001.

C Additional Empirics

This appendix reports additional analyses of the CEO panel of Section 4. We document how CEO turnover responds to performance (Appendix C.1), check our instrument with deeper lags of output (Appendix C.2), and examine how persistence varies with firm size and industry sector (Appendix C.3).

C.1 CEO Turnover and Performance

CEO turnover in our panel responds to performance, so stayer years are selected on performance. Assumption 2 allows boards to replace a CEO after poor performance, provided that two conditions hold. The two conditions are that the CEO for year t is chosen before the shock of year t is realized, and that this shock is unrelated to the information on which the decision

is based.

In this appendix, we document this selection in our panel, and in Appendix B.2 we simulate the AR(1) model with turnover that responds to performance. Controlling for a CEO's last stayer year changes the levels estimates of δ by approximately 1 percent or less. In the simulations, the levels estimator still understates δ in every design. Under turnover fitted to the data, FOD has little bias for log sales and understates δ for ROA by 0.03–0.04.

C.1.1 Selection on Performance in Our Panel

Panel A of Table 10 implies that moving a firm from the 10th to the 90th percentile of industry-adjusted ROA lowers its probability of a turnover by 5.4 p.p., from a base of about 12 percent. Performance in the previous year adds little once the current year is included. Panel B reports that the residual of the levels equation in a CEO's last stayer year before a turnover is lower than in other stayer years for all five outcomes. This is consistent with boards keeping a CEO more often after a good year than after a bad one.

Controlling for the last stayer year moves the levels estimates little. Although small, the changes in Panel C are statistically significant for operating ROA and EBITDA over assets, and borderline for log sales.[23] The changes are small because, once the fixed effects are removed, y_{t-2} is nearly uncorrelated with the indicator (the correlation is between -0.023 and -0.008).

None of these checks tests Assumption 2, because the assumption allows boards to replace a CEO after a bad year; what it rules out is retention that predicts the current year's shock. There are two ways this could happen that our data cannot rule out. The first is timing. Each fiscal year is assigned to the executive who held the post for most of it, and, judging from ExecuComp's dates, in 47 percent of turnovers the outgoing CEO left office during the first fiscal year assigned to the successor, typically three months in. If an early-year departure responds to news about year t, the remaining stayer years are selected on part of the year-t shock, and the dates cannot show whether boards act on such news. The second is serial correlation. If shocks are serially correlated, a board that keeps a CEO after a good year keeps one whose next shock is predictable, which would violate the assumption.

[23] With the standard errors of Table 10, $z = -2.66$, -2.90, and -1.96, computed from unrounded estimates. With bootstrap standard errors, the change for log sales is significant ($p = 0.022$).

Table 10: CEO turnover, performance, and the last stayer year

Panel A. Turnover after t: linear probability model, firm and year effects	
Industry-adj. ROA, t	−0.264
	(0.025)
Industry-adj. ROA, $t-1$	0.002
	(0.023)
Change in log sales, t	−0.040
	(0.009)
Change in log sales, $t-1$	−0.014
	(0.008)
Observations	36,922
Firms	2,683
Turnovers	4,476

	ROA	Operating ROA	EBITDA / Assets	Log sales	Log employment
Panel B. Residual in the last stayer year before a turnover					
Raw difference	−0.0119	−0.0063	−0.0077	−0.0223	−0.0217
	(0.0013)	(0.0010)	(0.0010)	(0.0033)	(0.0033)
FE-adjusted difference	−0.0170	−0.0090	−0.0110	−0.0319	−0.0310
	(0.0020)	(0.0016)	(0.0015)	(0.0051)	(0.0051)
Observations	34,134	34,105	34,087	34,338	33,781
Firms	2,758	2,755	2,755	2,772	2,745
CEOs	5,357	5,353	5,350	5,383	5,319
Last-year rows	3,455	3,451	3,449	3,473	3,406
Panel C. Persistence with a last-year dummy					
$\hat{\delta}$, main	0.2690	0.3782	0.3381	0.7437	0.7775
$\hat{\delta}$ with dummy	0.2669	0.3744	0.3357	0.7430	0.7770
	(0.0640)	(0.0731)	(0.0394)	(0.0158)	(0.0109)
Change in $\hat{\delta}$	−0.0021	−0.0039	−0.0024	−0.0007	−0.0005
	(0.0018)	(0.0015)	(0.0008)	(0.0004)	(0.0004)

Notes: Panel A: an observation is a firm-year with the next year and the regressors observed. The outcome is 1 if the firm's CEO in $t+1$ differs from its CEO in t, and the coefficients are from a linear probability model with firm and year fixed effects. Industry-adjusted ROA is the firm's ROA minus the median ROA in its SIC2×year cell, and the change in log sales is $\log(\text{sales}_t) - \log(\text{sales}_{t-1})$. Panels B and C use the stayer firm-years of Table 3, Panel A; the last stayer year is the stayer year before a turnover. Panel B compares the residual $y_t - \hat{\delta} y_{t-1}$, net of the fixed effects, in the last stayer year with other stayer years, raw and as the coefficient on a last-year indicator added to the levels regression (FE-adjusted). Panel C reports $\hat{\delta}$ with and without that indicator. Standard errors (in parentheses) are clustered by firm.

C.2 Instruments with Deeper Lags

Our instrument, output two periods ago, is valid if the current shock is uncorrelated with the earlier shocks that determine it (Section 4.1). Output three periods ago gives a check: if both lags are valid and firms share one δ, the two should give the same estimate. With FOD, the Hansen J test rejects this at the 5 percent level for all outcomes except operating ROA, but the differences mostly point to more persistence. The third lag raises the estimate of δ by 0.03 to 0.11 for ROA, operating ROA, EBITDA over assets, and log sales, and lowers it by 0.01 for log employment. If the shocks have a negative autocovariance at the second lag, the FOD estimates understate persistence for four of the five outcomes. Differences in δ across firms could also produce these higher estimates.

Panel B of Table 11 reports the FOD estimates with y_{t-3} and with y_{t-2} on the same observations. The J test of y_{t-3} against y_{t-4} does not reject for any outcome, consistent with shocks that are uncorrelated beyond the second lag.

If the shocks are correlated at the second lag and not beyond, output three periods ago is a valid instrument, in line with the test against output four periods ago, and its estimates, higher for four of the five outcomes, are consistent.[24] If instead δ differs across firms, both lags can be valid, and the deeper lag gives higher estimates because it weights the more persistent firms more. Neither first-order moving-average shocks, classical measurement error, nor a random-walk drift in the manager's effect would produce the rejection.

The test cannot distinguish a second-lag correlation of the shocks from differences in δ across firms. Under a second-lag correlation, the multipliers at the FOD estimates (Section 4.2.4) are too low for the four outcomes in which $\hat{\delta}$ rises, so we understate the long-run dispersion of CEO effects for these outcomes. Under differences across firms, both lags can give valid averages of the firms' δ, and our main estimates place less weight on the more persistent firms than the deeper lags do. Either way, the deeper lags give no sign that our FOD estimates overstate persistence for these four outcomes.

[24] An MA(2) fitted to the residual autocovariances matches the signs of all five differences between the two estimates. The autocovariances behind it are individually insignificant ($|t|$ at most 1.5).

Table 11: Persistence with deeper lags as instruments

	ROA	Operating ROA	EBITDA / Assets	Log sales	Log employment
Panel A. Levels FE-2SLS on stayer rows					
Main: y_{t-2} alone, all rows with y_{t-2}					
$\hat{\delta}$ (SE)	0.269 (0.064)	0.378 (0.073)	0.338 (0.039)	0.744 (0.016)	0.777 (0.011)
First-stage F	160	102	415	2,743	4,095
N (firms)	34,134 (2,758)	34,105 (2,755)	34,087 (2,755)	34,338 (2,772)	33,781 (2,745)
(a), (c) Rows with y_{t-2} and y_{t-3} observed					
(a) y_{t-2} alone	0.262 (0.072)	0.364 (0.082)	0.329 (0.044)	0.729 (0.019)	0.771 (0.012)
(a) y_{t-3} alone	0.307 (0.315)	0.320 (0.192)	0.341 (0.112)	0.789 (0.018)	0.772 (0.014)
Difference	0.045 (0.279)	−0.045 (0.174)	0.013 (0.087)	0.061 (0.013)	0.001 (0.006)
(c) y_{t-2} and y_{t-3}	0.262 (0.071)	0.362 (0.082)	0.328 (0.044)	0.727 (0.019)	0.771 (0.012)
F: (a) / (c)	5.2 / 69	14 / 45	41 / 156	642 / 1,230	993 / 1,968
(c) Hansen J $[p]$	0.03 [0.861]	0.08 [0.772]	0.03 [0.873]	29.37 [<0.001]	0.04 [0.841]
N (firms)	31,318 (2,579)	31,287 (2,577)	31,272 (2,577)	31,509 (2,594)	30,956 (2,569)
(b) Rows with y_{t-2}, y_{t-3} and y_{t-4} observed					
(b) y_{t-2} alone	0.251 (0.083)	0.379 (0.090)	0.327 (0.049)	0.726 (0.021)	0.759 (0.014)
(b) y_{t-3} and y_{t-4}	0.413 (0.316)	0.334 (0.215)	0.302 (0.127)	0.791 (0.021)	0.760 (0.016)
Difference	0.162 (0.293)	−0.045 (0.207)	−0.025 (0.099)	0.065 (0.015)	0.002 (0.007)
F (Wald/2)	5.1	5.4	22	285	430
(b) Hansen J $[p]$	0.44 [0.507]	0.28 [0.599]	0.07 [0.790]	1.89 [0.169]	3.14 [0.076]
N (firms)	28,771 (2,420)	28,734 (2,418)	28,720 (2,418)	28,919 (2,435)	28,396 (2,410)
Panel B. Forward orthogonal deviations, with constant					
Main: y_{t-2} alone, all rows with y_{t-2}					
$\hat{\delta}$ (SE)	0.583 (0.040)	0.707 (0.045)	0.647 (0.028)	0.813 (0.025)	0.872 (0.021)
First-stage F	402	282	461	208	225
N (firms)	32,389 (2,799)	32,360 (2,797)	32,338 (2,794)	32,575 (2,812)	32,037 (2,787)
(a), (c) Rows with y_{t-2} and y_{t-3} observed					
(a) y_{t-2} alone	0.585 (0.045)	0.730 (0.052)	0.656 (0.032)	0.812 (0.030)	0.894 (0.024)
(a) y_{t-3} alone	0.699 (0.059)	0.821 (0.065)	0.760 (0.039)	0.838 (0.028)	0.883 (0.022)
Difference	0.114 (0.042)	0.090 (0.058)	0.104 (0.027)	0.026 (0.008)	−0.011 (0.004)
(c) y_{t-2} and y_{t-3}	0.590 (0.045)	0.742 (0.050)	0.641 (0.032)	0.870 (0.029)	0.861 (0.020)
F: (a) / (c)	172 / 162	146 / 113	244 / 190	175 / 99	198 / 151
(c) Hansen J $[p]$	8.32 [0.004]	2.79 [0.095]	16.98 [<0.001]	12.22 [<0.001]	8.10 [0.004]
N (firms)	29,765 (2,631)	29,731 (2,629)	29,714 (2,627)	29,926 (2,647)	29,400 (2,622)
(b) Rows with y_{t-2}, y_{t-3} and y_{t-4} observed					
(b) y_{t-2} alone	0.581 (0.051)	0.731 (0.056)	0.660 (0.035)	0.822 (0.036)	0.881 (0.028)
(b) y_{t-3} and y_{t-4}	0.693 (0.068)	0.829 (0.065)	0.758 (0.045)	0.880 (0.032)	0.864 (0.023)
Difference	0.112 (0.045)	0.098 (0.060)	0.099 (0.031)	0.059 (0.026)	−0.018 (0.013)
F (Wald/2)	72	61	99	105	131
(b) Hansen J $[p]$	0.02 [0.879]	0.30 [0.583]	2.66 [0.103]	2.05 [0.152]	0.52 [0.469]
N (firms)	27,335 (2,468)	27,300 (2,466)	27,283 (2,465)	27,477 (2,483)	26,950 (2,461)

Notes: Estimates of δ with deeper lags of the outcome as instruments for y_{t-1}, with the levels estimator and with FOD. Main is the specification of Table 3, with instrument y_{t-2}. Block (a) instruments with y_{t-3}, block (b) with y_{t-3} and y_{t-4}, and block (c) with y_{t-2} and y_{t-3}. Blocks (a) and (c) use the rows on which y_{t-2} and y_{t-3} are observed, and block (b) the rows on which y_{t-2}, y_{t-3}, and y_{t-4} are observed; each "y_{t-2} alone" row repeats the main specification on those rows. Difference is a block's estimate minus the "y_{t-2} alone" estimate on the same rows. F is the first-stage F of the block's instruments (with two instruments, the Wald statistic divided by two). Hansen J has one degree of freedom and tests whether the two instruments identify the same δ; p-values are in brackets. Standard errors (in parentheses) are clustered by firm. Observations are rows (levels) or FOD equations (FOD), with the number of firms in parentheses.

C.3 Heterogeneity in Persistence across Firms

In this appendix, we report the estimates and tests behind Section 4.2.3, which splits our CEO panel by firm size, and the results of splitting it by industry sector. Appendix A.3.1 shows how persistence that varies with such observable characteristics is identified. Persistence may vary across firms with the durability of their organizational capital. We examine two splits: by firm size, since larger organizations may adjust more slowly, and by SIC sector, since the durability of organizational decisions may differ across sectors. Table 12 reports the size split for the levels estimator (Panel A) and for FOD (Panel B), and the sector split for log sales only, with both estimators (Panel C).

Panel A of Table 12 reports levels estimates for firms split at the median of their average log employment. Large firms have higher levels estimates of δ than small firms for all five outcomes. This difference is significant at 5 percent only for log sales (0.77 against 0.71, $p = 0.034$). Each firm stays in one group in every year, so the log-sales difference is a difference *between* firms.[25] Because the bias of the levels estimator depends on the variance of CEO effects relative to the variance of shocks, which may differ between large and small firms (Table 7), we also compare the groups with FOD.

Panel B reports that the FOD gap is positive for all five outcomes: 0.007 for ROA, 0.009 for operating ROA, 0.070 for EBITDA over assets, 0.064 for log sales, and 0.085 for log employment. Only the last is significant at 5 percent ($p = 0.035$). For log sales, the gap is about the size of the levels gap (0.067), with a larger standard error. After a Holm adjustment for five outcomes, neither the levels gap for log sales nor the FOD gap for log employment is significant (adjusted $p = 0.168$ and 0.173).

For the profitability outcomes, the signs of the gaps depend on the measure of size. When we split firms at the median of their mean log assets, the FOD gaps for log sales and log employment stay positive, the three profitability gaps change sign, and none of the five is significant at 5 percent.[26] For log sales and log employment, differences in spell length do not generate the gaps, since large firms have slightly shorter spells, with a mean of 9.86 years against 10.11 for small firms, weighted by stayer rows. These results suggest that firm scale, measured by log sales and log employment, is more persistent in larger firms, consistent with larger organizations adjusting more slowly.

In Panel C, we estimate the persistence regression for log sales separately within each major

[25] The split uses the 3,224 firms with employment data, and the median firm has roughly 3,900 employees. Splitting stayer rows instead by each firm's log employment two years earlier, which is predetermined, gives a log-sales difference of approximately 0.02, with p above 0.5 for each of three plausible medians.

[26] The FOD gaps are 0.053 for log sales ($p = 0.247$), 0.078 for log employment ($p = 0.082$), and −0.038, −0.016, and −0.044 for ROA, operating ROA, and EBITDA over assets. Splitting at the median of each firm's first observed log employment changes only the sign of the ROA gap (−0.029).

SIC division. In levels, $\hat{\delta}$ is lowest in mining (0.64) and highest in wholesale and retail trade (0.82). We reject equal persistence across the seven sectors ($\chi^2(6) = 16.70$, $p = 0.010$), but not without wholesale and retail ($\chi^2(5) = 5.21$, $p = 0.391$). Dropping mining instead leaves $p = 0.057$, so the rejection comes mainly from wholesale and retail. With FOD, $\hat{\delta}$ ranges from 0.73 in wholesale and retail to 0.83 in light manufacturing, close to the pooled estimate of 0.81, and we do not reject equal persistence across the seven sectors ($\chi^2(6) = 0.96$, $p = 0.99$). The persistence of log sales is thus substantial in every sector, and apart from the high levels estimate for wholesale and retail, we find little evidence that it differs across sectors.

Table 12: Persistence by firm size and by industry sector

	ROA	Operating ROA	EBITDA / Assets	Log sales	Log employment
Panel A. Levels FE-2SLS, by firm size					
Small firms: $\hat{\delta}$	0.253	0.323	0.289	0.706	0.766
	(0.088)	(0.115)	(0.062)	(0.028)	(0.018)
Large firms: $\hat{\delta}$	0.299	0.420	0.424	0.773	0.781
	(0.087)	(0.070)	(0.040)	(0.015)	(0.013)
Large − Small	0.046	0.097	0.135	0.067	0.014
	(0.124)	(0.135)	(0.074)	(0.032)	(0.022)
p-value	0.708	0.472	0.067	0.034	0.524
Observations, small	13,302	13,288	13,280	13,396	13,132
Observations, large	20,558	20,543	20,526	20,671	20,376
Panel B. Forward orthogonal deviations (FOD), by firm size					
Small firms: $\hat{\delta}$	0.594	0.689	0.621	0.775	0.801
	(0.053)	(0.068)	(0.042)	(0.042)	(0.036)
Large firms: $\hat{\delta}$	0.601	0.699	0.691	0.838	0.886
	(0.047)	(0.042)	(0.024)	(0.020)	(0.018)
Large − Small	0.007	0.009	0.070	0.064	0.085
	(0.070)	(0.080)	(0.048)	(0.047)	(0.040)
p-value	0.922	0.906	0.147	0.171	0.035
Observations, small	12,641	12,627	12,621	12,733	12,483
Observations, large	19,748	19,733	19,717	19,842	19,554

Panel C. Log sales, by industry sector

	Levels		FOD		FOD
Sector	$\hat{\delta}$	SE	$\hat{\delta}$	SE	first-stage F
Mining	0.642	(0.052)	0.786	(0.058)	23.5
Construction	0.737	(0.044)	0.821	(0.081)	9.3
Manufacturing–light	0.719	(0.048)	0.826	(0.050)	62.2
Manufacturing–heavy	0.765	(0.021)	0.796	(0.044)	59.3
Transport & Comms	0.725	(0.063)	0.753	(0.170)	5.4
Wholesale & Retail	0.819	(0.019)	0.732	(0.097)	12.6
Services	0.733	(0.027)	0.801	(0.056)	31.2
All firms (pooled)	*0.744*	*(0.016)*	*0.813*	*(0.025)*	*208*
Test of equal δ across sectors: χ^2 statistic and [p-value]					
$\chi^2(6)$, seven sectors	16.70	[0.010]	0.96	[0.987]	
$\chi^2(5)$, excl. Wholesale & Retail	5.21	[0.391]	0.45	[0.994]	

Notes: Estimates of δ as in Table 3, separately for firms above and below the median of their mean log employment (1,612 firms per group). Panel A uses the levels estimator and Panel B FOD. Large − Small is the difference between the two estimates, with a standard error from the two independent samples, and p-values are two-sided. Panel C estimates δ for log sales separately within each SIC division, with both estimators; the tests of equal δ across divisions use the joint covariance of the separate estimates. Standard errors (in parentheses) are clustered by firm. Data: ExecuComp and Compustat, 1992–2025.