Managers and the Persistence of Productivity*

Aroon Narayanan
MIT

Roi Orzach
Boston University

August 16, 2026

Abstract

Standard two-way fixed effects models assume that a manager's effect on performance is fully realized upon assignment. We instead model output as a persistent process to which the current manager contributes each period; the standard model is the special case in which the persistence parameter δ is zero. We prove that δ is point-identified using lagged output as an instrument, and that previous estimates understate the effects of managers. Using four empirical applications, we demonstrate that ignoring persistence understates how much leaders matter for organizational outcomes and thus how much leader selection policies matter. First, applying the framework to public-company CEOs, we find that incorporating persistence results in a 50–60 percent increase in the estimated long-run standard deviation of CEO effects on firm scale (employment and sales) relative to the standard variance-decomposition estimate. Next, incorporating persistence raises the estimated long-run effect of a CEO having a daughter on Corporate Social Responsibility activity by roughly 70 percent. Finally, applying the framework to two policy reforms that changed how senior leaders are selected—the Warren Fisher Reform in the British Colonial Office and the 2003 introduction of merit-based selection for Chilean public hospital CEOs—we find that the long-run gain from improved leader selection is several times larger than the static estimates suggest.

*We are thankful for helpful conversations with Charles Angelucci, Mert Demirer, Robert Gibbons, Andy Newman, Vishan Nigam, Fiona Paine, Andrea Prat, Rafaella Sadun, and Antoinette Schoar. We thank Guo Xu for making his data publicly available, and Henrik Cronqvist, Fang Yu, and Cristóbal Otero for graciously sharing their data with us.

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 comparable 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 (cf. Lazear et al., 2015), public sector administration (cf. Fenizia, 2022), multinationals (cf. Minni, 2026), hospitals (cf. Muñoz and Otero, 2025), and laboratory experiments (cf. Weidmann et al., 2024).

At the same time, a seeming contradiction runs through much of this applied work: managerial effects are assumed to be 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 shown 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) papers is that managerial practices (or organizational capital as measured in Cai et al., 2026) is observed to the econometrician, whereas in the analysis that follows managerial capital will have to be estimated.

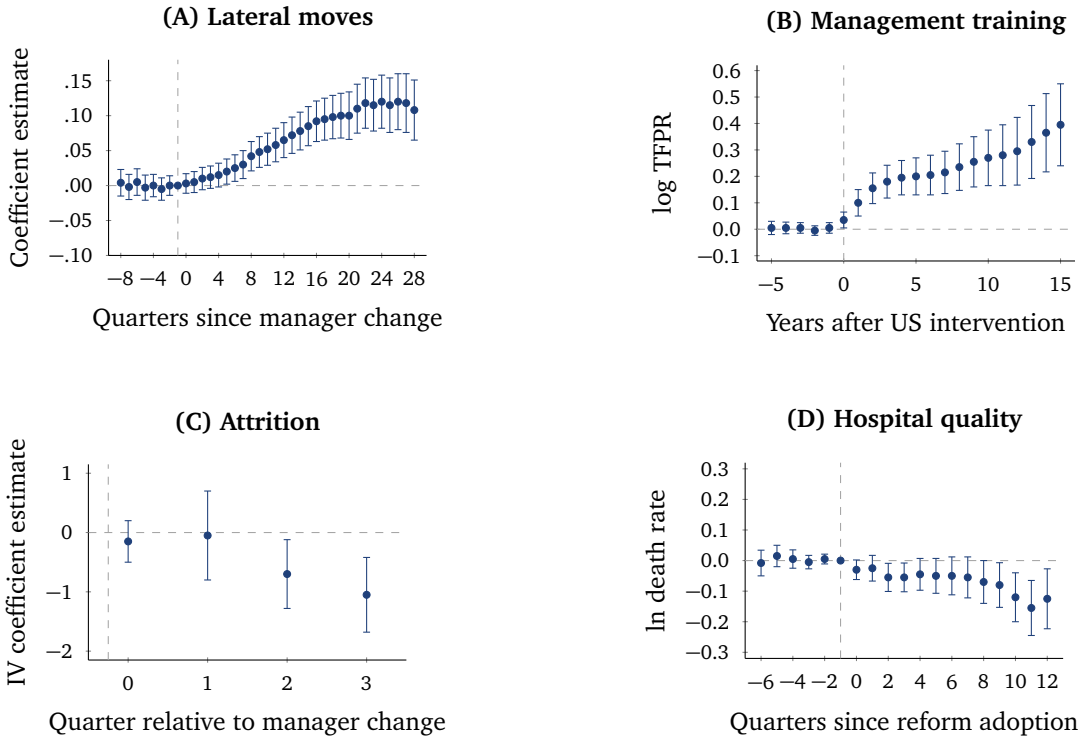


Figure 1: This figure presents event-study evidence on the dynamic effects of managers on organizational performance. **Panel (A)**: from Minni (2026). The figure plots the differential number of lateral job moves for workers who transition from a "low-flyer" to a "high-flyer" manager relative to those who transition between low-flyer managers. **Panel (B)**: from Giorcelli (2019). The figure shows log total factor productivity revenue (TFPR) for Italian firms whose managers received US-sponsored training under the Marshall Plan, relative to firms excluded by an unexpected budget cut. **Panel (C)**: from Hoffman and Tadelis (2021). The figure 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 figure presents the dynamic effect on log hospital death rates of adopting a meritocratic selection reform for hospital CEOs in Chilean public hospitals.

In this paper, we develop an econometric framework in which output is a persistent process to which the current manager contributes each period. Plant output evolves as $y_{t,p} = \delta\, y_{t-1,p} + \theta_{m_{t,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}$, with $\delta \in [0,1)$ the persistence of productivity. The standard two-way fixed effects model is the special case $\delta = 0$, in which a manager's per-period contribution is also their long-run effect. We show that δ is point-identified using an instrumental variables strategy in the spirit of Anderson and Hsiao (1982): on a subsample of plants where the manager does not change between $t-1$ and t, the only systematic link between current and lagged output is the persistence parameter itself, so deeper lags identify δ under strictly weaker assumptions than those required to identify manager fixed effects (e.g., those in Bertrand and Schoar, 2003).

Ignoring persistence has consequences for the measurement of managerial effects. We show that the standard two-way fixed effects (TWFE) approach, which implicitly imposes $\delta = 0$,

understates the share of productivity differences attributable to managers. Intuitively, when a new manager arrives at a plant, prior output reflects the accumulated influence of predecessors, not the current manager's quality. A static regression attributes only the portion of output that covaries with the current manager's identity, treating the inherited component as residual noise. The traditional estimates therefore understate the share of productivity attributable to managers. Across our empirical applications, the true share of output variance explained by managers is several times larger than the TWFE estimate.

The empirical applications document three consequences of ignoring persistence. The first concerns how much leaders matter for firm performance, the second how much observable leader characteristics shape firm choices, and the third how much value is gained by using a better system for selecting leaders. Further, each of these analyses incorporating persistence imposes no new assumptions on the data (besides that the panel must include at least 3 consecutive time periods), and can be implemented from the existing static specification by adding a lagged dependent variable and instrumenting it with the second lag of the outcome.

We begin with U.S. public-company CEOs. In an ExecuComp–Compustat panel built in the spirit of Bertrand and Schoar (2003), we estimate lower persistence for profitability ($\hat{\delta} \approx 0.23$–$0.39$) and higher persistence for firm scale ($\hat{\delta} \approx 0.74$ for log sales, $\hat{\delta} \approx 0.78$ for log employment) as expected: competitive forces should erode profit margins quickly, while decisions that shape firm scale do not unwind as fast. Applying our methodology, we find the long-run dispersion of CEO effects on firm scale is larger than the standard estimate by 50–60 percent. The same correction applies to regressions that ask which leader characteristics shape firm choices. We revisit a seminal paper in this literature: Cronqvist and Yu (2017), which shows that having daughters leads CEOs to increase their firms' Corporate Social Responsibility (CSR) activity. Using our methodology, we estimate a relatively large persistence for CSR activity, $\hat{\delta} \approx 0.60$. Given this persistence, the long-run effect of CEOs with daughters is around 70 percent larger than the static estimate.

We then turn to two settings in which policy reforms changed who got selected as a leader. The first concerns Xu (2018)'s analysis of patronage in the British Colonial Office. Xu (2018) estimates that patronage, the allocation of governorships on the basis of social connections, reduced colony revenue by approximately four percent: a number that also quantifies the gain from the Warren Fisher Reform, which replaced patronage with merit-based selection. We estimate $\hat{\delta} \approx 0.80$ for colony revenue, which implies a long-run cost of patronage of approximately nineteen percent, roughly five times the static estimate. The Warren Fisher Reform was therefore much more valuable than a static analysis of the cost of patronage would suggest.

Second, we apply our framework to Muñoz and Otero (2025)'s analysis of the 2003 reform of senior public-sector hiring in Chile. The reform replaced discretionary appointment with a

competitive merit-based selection system for top public-sector positions, including the CEOs of public hospitals, and Muñoz and Otero (2025) estimates that the reform reduced in-hospital mortality among public hospitals by approximately eight percent. We estimate $\hat{\delta} \approx 0.84$ for log hospital mortality, which implies a long-run mortality reduction of approximately eighteen percent, roughly 2.5 times the static estimate. Further, the long-run gain is concentrated in hospitals that drew high-quality CEOs, for which we estimate a long-run mortality reduction of roughly 25 percent; hospitals that drew lower-quality CEOs show no detectable long-run improvement.

Implications for the literature. Our framework yields several implications for the literature on manager and leader effects. First, standard static variance decompositions understate the long-term value of leadership. We confirm this prediction empirically in our CEO analysis, where persistence ranges from $\hat{\delta} \approx 0.23$ for profitability to $\hat{\delta} \approx 0.74$–$0.78$ for firm scale. That the prior literature implicitly understates the value of good CEOs 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. In our two analyses analyzing the effects of improved selection processes, we estimate effect sizes 2-5 times the static estimates.

Third, persistence has implications for optimal performance contracts. Dessein and Prat (2022) model a setting in which a CEO allocates effort between building organizational capital and boosting short-run performance measures. A central result of their analysis is that when organizational capital is persistent and unobservable, contracts should reward long-run performance as short-run incentives distort effort away from capital accumulation. Our estimates provide empirical support for their assumption of high persistence that drives this result.[2] 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 contracts add unnecessary variance to a manager's evaluation and penalize managers who inherit a low output level from weak predecessors.

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

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 requires only a sample of stayers and deeper lags as instruments, circumventing the need to estimate high-dimensional fixed effects and remains valid even if managers never move across plants. Moreover, three of our four empirical applications—the CSR analysis, the colonial analysis, and the hospital analysis—estimate the effects of observable characteristics or policy reforms rather than of unobservable manager fixed effects themselves, so the TWFE identification concerns do not bear on these conclusions. Only the CEO variance decomposition requires TWFE-style fixed-effect estimation. Appendix F provides Monte Carlo simulations of the persistence estimator's finite-sample performance.

The remainder of the paper proceeds as follows. Section 2 presents the econometric model. Section 3 develops the identification results, characterizes the bias from ignoring persistence, and extends the framework to heterogeneous persistence. Section 4 applies the framework to U.S. public-company CEOs, revisiting the Bertrand and Schoar (2003) variance decomposition (how much CEOs matter) and the Cronqvist and Yu (2017) regression on CEO daughters and CSR (which CEO characteristics matter). Section 5 turns to two policy reforms that changed how senior leaders are selected: the Warren Fisher Reform in the British Colonial Office (Xu, 2018) and the 2003 introduction of competitive recruitment for senior public-sector positions in Chile (Muñoz and Otero, 2025). In both settings, the gain from merit-based selection is several times larger than the static estimate.

2 Econometric Model

Time is discrete, $t = 0, 1, \ldots, T$, and plants are indexed by $p \in \{1, \ldots, P\}$. Each plant p is assigned a manager $m_{t,p} \in \{1, \ldots, M\}$ at date t. Manager m has an unknown productivity effect $\theta_m \in \mathbb{R}$; we write $\theta_{m_{t,p}}$ for the productivity of the manager assigned to plant p at time t. Plant output depends on past productivity, the current manager's contribution, and plant and time effects:

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

Here α_p is a plant fixed effect, γ_t is a time fixed effect, and $\epsilon_{t,p}$ is a mean-zero production innovation. The parameter $\delta \in [0, 1)$ measures the persistence of productivity. We write $\Delta x_{t,p} := x_{t,p} - x_{t-1,p}$ for first differences. We require that, with positive probability, the same manager remains at a plant between $t-1$ and t. We allow managerial assignment to depend on prior plant and manager histories, subject to the timing restriction below. The identification

argument follows the logic of Anderson and Hsiao (1982).

Assumption 1. *Fix $t \geq 2$. Let $D_{t,p} = \mathbf{1}_{m_{t,p} \neq m_{t-1,p}}$. Assume:*

(E1) *(*Exogeneity of Production Innovations*) For each p, $\epsilon_{t,p}$ is mean zero and satisfies*

$$\mathbb{E}[\epsilon_{t-k,p} \mid \mathscr{F}_{t-2}] = 0 \qquad \text{for } k \in \{0,1\},$$

where $\mathscr{F}_{t-2}$ is generated by $\{y_{\tau,p}, m_{\tau,p}\}_{\tau \leq t-2}$ and all fixed effects.

(E2) *(*Predetermined Managerial Assignment*) Conditional on $\mathscr{F}_{t-2}$, the switch event $D_{t,p}$ and the identities $(m_{t,p}, m_{t-1,p})$ are independent of $\{\epsilon_{t,p}, \epsilon_{t-1,p}\}$.*

Interpretation of Assumption 1. Condition (E1) requires that this year's output shocks cannot be predicted from the past. Appendix 6 tests this, with broadly supportive results. Condition (E2) allows manager replacements to depend on plant and time fixed effects, on lagged output, and on a manager's track record, so scheduled turnover, retention of strong managers, and selection on past performance are all permitted. However, it rules out turnover that responds to the output surprise in the year just before the switch: a manager may be replaced for a poor record, but not for a single unexpectedly bad year. Because the innovations are latent, (E2) cannot be tested directly, but its most plausible violation, selection into the stayer sample that reacts to ϵ_{t-1}, would itself compromise the identification checks in Appendix 6.

3 Econometric Analysis

In this section we begin by showing that δ can be identified given Assumption 1. Next, we discuss the theoretical predictions on the biases that will emerge from imposing $\delta = 0$ when estimating managerial effects.

3.1 Identification of δ with Plant and Time Fixed Effects

We begin by introducing the following notation. Define the "no manager change" indicator

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

Further, for any array $\{X_{t,p}\}_{p=1}^{P}$ and date t, define the cross-sectional mean and time-demeaned value as

$$\overline{X}_t := \frac{\sum_{p=1}^{P} X_{t,p}}{P} \quad \text{and} \quad X^*_{t,p} := X_{t,p} - \overline{X}_t.$$

Proposition 1 (Identification of δ with plant and time fixed effects)**.** *Let $y_{t,p}$ satisfy (1) with $\delta < 1$, and define $S_{t,p}$ and $(\cdot)^*$ as above. Suppose there exists a scalar variable $Z_{t,p}$ such that, for some $t \geq 2$,*

$$\mathbb{E}\big[(\Delta\epsilon_{t,p})^* Z_{t,p} S_{t,p}\big] = 0, \tag{2}$$

$$\mathbb{E}\big[(\Delta y_{t-1,p})^* Z_{t,p} S_{t,p}\big] \neq 0. \tag{3}$$

Then δ is point-identified and satisfies

$$\delta = \frac{\mathbb{E}\big[(\Delta y_{t,p})^* Z_{t,p} S_{t,p}\big]}{\mathbb{E}\big[(\Delta y_{t-1,p})^* Z_{t,p} S_{t,p}\big]}. \tag{4}$$

Finally, $Z_{t,p} = y_{t-2,p}$ satisfies (2)–(3) under Assumption 1. The deeper lag $Z_{t,p} = y_{t-3,p}$ also satisfies these conditions and may be used as an additional instrument.

Intuitively, first-differencing removes the plant fixed effect. On the subsample with no manager change between $t-1$ and t, the manager term also drops out, so the only systematic link between current growth $\Delta y_{t,p}$ and lagged growth $\Delta y_{t-1,p}$ is the persistence parameter δ. After accounting for time fixed effects (which shift all plants in a given period equally) and addressing the endogeneity of $\Delta y_{t-1,p}$, regressing $\Delta y_{t,p}$ on $\Delta y_{t-1,p}$ recovers δ. As is common in such differenced regressions, the composite error term, $y_{t,p}$ contains both $\epsilon_{t,p}$ and $-\epsilon_{t-1,p}$, the latter of which is shared with the endogenous regressor, making OLS estimates biased. A natural instrument is $y_{t-2,p}$: under Assumption 1, $y_{t-2,p}$ is predetermined with respect to the shocks entering the current differenced error, and hence is orthogonal to it. Finally, $y_{t-2,p}$ enters $y_{t-2,p}$ with a coefficient of 1 but enters $y_{t-1,p}$ with a coefficient of $\delta < 1$, implying $y_{t-2,p}$ enters $\Delta y_{t-1,p}$) with a non-zero coefficient, satisfying relevance.

3.2 Bias from Ignoring Persistence

Before discussing the bias from not accounting for persistence when estimating the effect of managers on productivity, we note that persistence makes the question of what is the effect of managers on productivity inherently ambiguous. Is the question what is the effect of permanently hiring a better manager on long-run productivity, the short run effect of hiring a better manager for one period, the sum of the effect in all subsequent periods due to persistence? Additionally, how should a firm discount the stream of future performance gains from a better manager today, given that the relevant discount factor differs from δ?

Absent persistence, such ambiguity is removed. Denote by σ the standard deviation of managerial productivity. Then absent persistence hiring a one standard deviation better man-

ager, improves performance by σ in each period they are employed, independent of the prior manager and all other nuances. Instead, in our dynamic framework, hiring a one standard deviation better manager leads to a σ improvement in the first period after being hired. However, permanently changing hiring practices resulting in a permanent one standard deviation increase in manager quality would lead to a $\sigma/(1-\delta)$ improvement in steady-state productivity.[3] We now state the estimated effects if a researcher used the TWFE model to answer this question.

Proposition 2 (Biased Productivity)**.** *Suppose output satisfies*

$$y_{t,p} = \delta y_{t-1,p} + \theta_{m_{t,p}} + \varepsilon_{t,p},$$

with $\mathbb{E}[\theta_{m_{t,p}}] = 0$ *and* $\text{Var}(\theta_{m_{t,p}}) = \sigma^2_\theta$*. Assume that within each plant and each period the manager remains with probability q, so expected within-plant tenure is* $\mu := 1/(1-q)$*. A researcher who estimates this utilizing TWFE imposing* $\delta = 0$ *recovers*

$$\widehat{\sigma}_\theta := \sqrt{\text{Var}(\widehat{\theta}_m)} = \frac{\sigma_\theta}{(1-\delta)+\delta/\mu}. \tag{5}$$

Therefore, this researcher would overstate the short-run effect of hiring a better manager by a factor of $\frac{1}{(1-\delta)+\delta/\mu}$*, but understate the long-run effects of policy reforms that result in permanently better managers by a factor of* $1/a = 1 + \frac{\delta}{\mu(1-\delta)}$.

Let us build intuition behind this result. Set $\delta = 1/2$ and assume that a new manager with $\theta = 1$ takes over at an "average" firm for two periods. Therefore, in period 1 the productivity improvement relative to the previous manager is equal to 1 and the next period the productivity improvement is $1 + \frac{1}{2} \cdot 1 = 3/2$. The TWFE model averages these two improvements and states that the managers effect is 1.25. Using this estimate, one would conclude that the immediate effect of hiring this manager would be $1.25 > 1$, in the true data generating process. Further, this estimate would imply that permanently hiring managers of this type would result in a productivity improvement of 1.25, rather than the true long-run effect of $1/(1-\delta) = 2$. When μ, the expected number of periods a manager is present in a firm, is smaller, the TWFE has fewer sample periods to capture the true effect of a manager, and thus understates the dynamic effect more. Similarly, when δ is larger, it takes even longer to see the true effect of a manager, and the misestimate is larger in the TWFE estimation.

The first implication of the proposition is that the standard two-way fixed effects model underestimates the true dispersion in long-run managerial productivity, as can be seen from

[3] This notion of $\sigma/(1-\delta)$ would also correspond to the undiscounted sum of future productivity gains from hiring a better manager in only one period.

the attenuation factor a being less than one in Equation (5). This is because good managers inherit lagged output that reflects the accumulated influence of their predecessors, and since predecessors are on average of mean productivity, the inherited output compresses toward the mean. A static regression attributes only the portion of output that covaries with the current manager's identity, treating the inherited component as residual noise. As a result, by understating the true variance in long-run productivity, policy reforms which alter the average managerial quality will have an understated effect on firm performance.

A related question in this literature is what share of productivity differences is attributable to managers. The next proposition shows that traditional estimates understate this share as well.

Proposition 3 (Biased Manager R^2)**.** *Consider the same data generating process as in Proposition 2. Let R^2_{TWFE} denote the population R^2 from a two-way fixed effects regression that imposes $\delta = 0$, which attributes variation to the current manager only:*

$$R^2_{\text{TWFE}} := \frac{\text{Var}(\mathbb{E}[y_{t,p} \mid m_{t,p}])}{\text{Var}(y_{t,p})}. \tag{6}$$

Second, let R^2_{true} denote the share of variance explained by managers assigned to the plant:

$$R^2_{\text{true}} := \frac{\text{Var}(\mathbb{E}[y_{t,p} \mid \{m_{\tau,p}\}_{\tau \le t}])}{\text{Var}(y_{t,p})}. \tag{7}$$

The ratio of the static R^2 to the true manager-driven share is

$$\frac{R^2_{\text{TWFE}}}{R^2_{\text{true}}} = \frac{1-\delta^2}{1-(\delta q)^2} = \frac{1-\delta^2}{1-\delta^2\left(1-\frac{1}{\mu}\right)^2} < 1.$$

This ratio is strictly increasing in μ and is strictly decreasing in δ for $\mu > 1$.

Therefore, the conventional methodologies underestimate the share of productivity dispersion attributable to managers. With persistence, the manager contribution to output is not just the current manager's effect but an accumulated sum $\Theta_{t,p} = \sum_{j\ge 0} \delta^j \theta_{m_{t-j,p}}$ of all past manager effects that reflects the entire history of managers at plant p, with recent managers weighted more heavily. Thus the natural benchmark for "how much productivity is attributable to managers" is the fraction of output variance explained by $\Theta_{t,p}$, namely $R^2_{\text{true}} = \text{Var}(\Theta_{t,p})/\text{Var}(y_{t,p})$. A two-way fixed effects regression that imposes $\delta = 0$ instead forces productivity to move one-for-one with the current manager dummy. Under random assignment, the inherited component $\sum_{j\ge 1} \delta^j \theta_{m_{t-j,p}}$ is unrelated to the identity of the current manager and therefore cannot

be attributed to that manager. The static regression implicitly treats this inherited variation as exogenous noise. As a result, the regression imposing $\delta = 0$ explains only the portion of $\Theta_{t,p}$ that is systematically related to the current manager, and its fit satisfies $R^2_{\text{TWFE}} < R^2_{\text{true}}$ whenever $\delta > 0$ and within-plant tenure is finite. In Sections 4 and 5, we show that these biases can be substantial.

3.3 Practical Implementation and Consistency

In practice, the differenced identification strategy described above loses much of the cross-sectional variation and adds noise from compounded errors. We now describe the estimation procedure we use throughout the paper. The regression takes the form

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

instrumenting $y_{t-1,p}$ with $y_{t-2,p}$, and including plant, time, and current-manager fixed effects. This levels specification is asymptotically equivalent to the Anderson–Hsiao-style differenced estimator as $T \to \infty$: in a long panel the within transformation that absorbs the plant fixed effect converges to first differencing (see Appendix Section F for an extended discussion).[4] Since all the datasets we employ have long time horizons, this approximation is appropriate. The levels approach has the practical advantage of preserving more cross-sectional variation and avoiding the additional noise introduced by differencing. Standard errors are clustered at the plant level throughout. In Appendix Section F, we provide Monte Carlo simulations which confirm that this estimation procedure performs well on data with similarly long time horizons.

Finally, for the estimation of δ, we restrict to observations where the manager does not change between $t-1$ and t (i.e., $m_{t,p} = m_{t-1,p}$). On this subsample the current-manager fixed effect absorbs $\theta_{m_{t,p}}$, and since the manager has not changed, the manager's contribution drops out of the relationship between $y_{t,p}$ and $y_{t-1,p}$, leaving the persistence parameter δ as the only systematic link between current and lagged output. Appendix B.1 shows that δ is separately identified on mover observations as well, and that both parameters can be estimated using the same IV levels strategy. Throughout the paper we use an AR(1) specification; Appendix A reports model-selection diagnostics showing that AR(0) is too sparse for the primary outcomes and that AR(2) specifications add little beyond the AR(1) model.

[4] Formally, the levels specification differs from the Anderson–Hsiao differenced estimator only in how the fixed effects are removed. First-differencing eliminates α_p exactly, whereas the within transformation introduces a correlation of order $O(1/T)$ between the demeaned regressor and the demeaned error, referred to as Nickell bias. This is the only source of divergence between the two estimators, and it vanishes as $T \to \infty$.

3.4 Alternative Mechanical Explanations for Persistence.

A natural concern is that the estimated persistence parameter δ may reflect mechanical features of the data rather than a property of the production process that amplifies managerial effects. We consider several such channels and argue that each either biases our estimates downward, strengthening our conclusions, or is directly addressed by the empirical design.

A first concern is that δ captures persistence in the economic environment—demand conditions, installed capital, or long-term contracts—rather than persistence in the managerial contribution to output. We note that this interpretation is consistent with our conceptual framework. The parameter δ is a property of the production process, not of managers, per se. Our central argument does not require managers themselves to be persistent; rather, because the production process is persistent, a manager's per-period contribution accumulates into a long-run effect that exceeds the contemporaneous flow effect. Any channel through which past output feeds into current output (e.g., capital, organizational routines, or workforce composition) is part of the mechanism that makes managerial effects durable. Moreover, the documented patterns of persistence estimates across outcomes is difficult to reconcile with a pure environmental-inertia story: outcomes that reflect accumulated organizational decisions (firm scale, CSR policies, fiscal revenue) exhibit high persistence, while outcomes tied to short-run performance and competitive pressure (profitability, social unrest) do not.

A second concern is endogenous manager assignment. If higher-quality managers are systematically assigned to plants on the basis of recent performance, a spurious correlation between lagged and current output could arise through the selection channel rather than through organizational persistence. Two features of the design mitigate this concern. First, the identification of δ relies on the stayer subsample, where the manager does not change between $t-1$ and t, so the assignment mechanism that governs managerial transitions is not operative on the estimation sample. Second, even on mover observations, the direction of any bias works against our findings. The empirical literature on executive turnover documents that CEO departures are strongly predicted by poor firm performance (Jenter and Lewellen, 2021), so new managers disproportionately arrive at declining plants. If boards replace departing managers with higher-quality successors, incoming manager quality is negatively correlated with inherited output, which would bias $\hat{\delta}$ downward rather than upward.[5]

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

[5] A parallel logic applies to CEO fixed-effect estimates. In Dessein and Prat (2022), good CEOs build organizational capital and are retained while poor matches are terminated, so observed quality is censored at the lower tail and assortative matching lets firm fixed effects absorb part of the CEO contribution. Both forces attenuate estimated CEO effects, making them lower bounds on the true dispersion of CEO quality.

δ could reflect shock persistence rather than organizational persistence. This is the one channel that could, in principle, bias $\hat{\delta}$ upward. The instrumental variables strategy addresses it directly: on the stayer subsample, the current-manager fixed effect absorbs any time-invariant manager-specific component, and the instrument $y_{t-2,p}$ is valid provided the production innovation $\epsilon_{t,p}$ is uncorrelated with $\epsilon_{t-2,p}$. The Arellano–Bond AR(2) test, reported in Appendix 6, examines precisely this condition and fails to reject at conventional levels for the main CEO profitability and scale outcomes, supporting the view that the estimated persistence reflects the production process rather than residual serial correlation in unobservables. The inclusion of industry-by-year fixed effects in the CEO panel further absorbs sector specific cyclical variation that would otherwise contaminate the persistence estimate.

A fourth concern is Nickell bias. The inclusion of a lagged dependent variable in a fixed-effects panel introduces a mechanical negative correlation between the demeaned regressor and the demeaned error term, biasing the autoregressive coefficient downward by a term of order $1/T$. The panels used in this paper have long time dimensions—over thirty years for the CEO panel, over a century for the colonial panel, and sixty quarters for the hospital panel—so the bias is small in magnitude. Further, because it is negative, any remaining Nickell bias causes us to understate persistence. This is confirmed by the Monte Carlo simulations in Appendix F, which show that $\hat{\delta}$ exhibits negative bias across nearly all configurations, consistent with the asymptotic analysis of Alvarez and Arellano (2003), and that the bias shrinks rapidly as T grows.

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

The U.S. public-company CEO setting allows us to revisit two influential static estimates that ignore persistence. Bertrand and Schoar (2003)'s variance decomposition is the standard quantitative answer to the question of how much CEOs matter for firm performance. We show that the implicit assumption in Bertrand and Schoar (2003) that $\delta = 0$ results in understatements of the long-run dispersion of CEO effects. Similarly, we show that previous papers which ask how much CEO characteristics affect firm performance will also understate the long-run effects of these characteristics. Here, we show that Cronqvist and Yu (2017)'s estimate of the effects of CEOs with daughters on corporate social responsibility is understated by roughly 70 percent. We first describe the CEO–firm panel built from ExecuComp and Compustat, present persistence estimates across firm outcomes, and then revisit each estimate in turn.

4.1 Data

We merge ExecuComp annual CEO identifiers with Compustat firm-level financials for the period 1992–2025. We exclude financial firms (SIC 6000–6999) and utilities (SIC 4900–4999), following Bertrand and Schoar (2003), and construct profitability and scale measures using lagged total assets as the denominator.[6] The resulting panel contains 47,102 firm-year observations covering 3,226 firms and 7,756 unique CEOs. Of these, 38,505 are stayer observations (the same CEO serves in consecutive years) and 5,090 are mover observations (a new CEO takes over); the remaining 3,507 are the first observation of each firm's panel or years following a gap in coverage, for which the stayer/mover classification is undefined. The mean CEO tenure is approximately five years, and 434 CEOs appear in more than one firm, though 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 correspond to managers (m). We consider five outcome variables: return on assets (net income over lagged total assets), operating return on assets (operating cash flow over lagged total assets), EBITDA over lagged total assets, log sales, and log employment. The first four span the profitability and scale measures central to the Bertrand and Schoar (2003) analysis; log employment provides a direct measure of organizational scale and captures the hiring-decisions channel through which managerial effects may persist.

4.2 Persistence Estimates

We estimate the following specification, which nests the standard two-way fixed effects model as the special case $\delta = 0$:

$$y_{t,f} = \delta\, y_{t-1,f} + \alpha_f + \gamma_t + \theta_{m_{t,f}} + \zeta_{i(f),t} + u_{t,f}, \tag{8}$$

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. The industry–year terms absorb sector-level demand and supply variation that would otherwise generate serial dependence in y at the firm level; without them, δ would conflate firm-specific persistence with industry-wide cyclicality. We instrument $y_{t-1,f}$ with $y_{t-2,f}$, absorb fixed effects via iterative demeaning, and cluster standard errors at the firm level. The estimation sample is consecutive

[6] All outcome variables are winsorized at the 1st and 99th percentiles. Lags used as instruments and as the endogenous regressor are then constructed from the winsorized series, so the dependent variable, the endogenous regressor, and the instruments all share the same support, preventing extreme outliers from dominating the first-stage relationship.

firm-years where the same CEO is in place at $t-1$ and t, defined as the stayer subsample. On this stayer subsample the current-CEO fixed effect absorbs both contemporaneous and lagged manager productivity, leaving δ identified from within-CEO output dynamics. Table 1 reports the results.

Table 1: Persistence estimates: U.S. CEO–firm panel (ExecuComp/Compustat, 1992–2025)

	$\hat{\delta}$
Outcome: Return on Assets	
Estimate	0.231
Clustered SE	(0.056)
First-stage F	2,559
Observations (IV sample)	35,107
Outcome: Operating Return on Assets	
Estimate	0.386
Clustered SE	(0.059)
First-stage F	1,900
Observations (IV sample)	35,077
Outcome: EBITDA / Assets	
Estimate	0.319
Clustered SE	(0.036)
First-stage F	7,417
Observations (IV sample)	35,053
Outcome: Log Sales	
Estimate	0.744
Clustered SE	(0.015)
First-stage F	58,166
Observations (IV sample)	35,355
Outcome: Log Employment	
Estimate	0.784
Clustered SE	(0.010)
First-stage F	61,772
Observations (IV sample)	34,979

Notes: IV estimates from the levels specification (8), with $y_{t-1,f}$ instrumented by $y_{t-2,f}$. Firm, year, CEO, and SIC2×year fixed effects are absorbed via iterative demeaning. The sample restricts to firm-years in which the same CEO was in place in the prior period. Standard errors are clustered at the firm level.

Profitability outcomes have low persistence: $\hat{\delta} = 0.23$ for return on assets, 0.39 for operating return on assets, and 0.32 for EBITDA over assets. Firm scale is much more persistent, with $\hat{\delta} = 0.74$ for log sales and $\hat{\delta} = 0.78$ for log employment. This discrepancy is an intuitive consequence of the model. The parameter δ reflects the durability of organizational decisions, and

competitive forces erode profit margins faster than they erode firm scale. A CEO who grows or shrinks the firm leaves an imprint on sales that lasts years; a CEO who improves margins through pricing or cost discipline can be undone within a year or two. The high persistence of employment is consistent with the view that reshaping a workforce is among the slowest channels of managerial influence.[7]

Two consistency checks support the firm-level interpretation of δ. Dropping the SIC2×year fixed effect leaves the point estimates almost unchanged: $\hat{\delta} = 0.21$ for ROA versus 0.23 in the headline, and 0.73 for log sales versus 0.74. Industry-wide cyclicality therefore contributes little to the firm-level persistence we estimate. A second check restricts the sample to mover firm-years, in which a new CEO has just taken over. The mover estimates are noisier because the sample is smaller, but the point estimates remain broadly consistent with the stayer values: $\hat{\delta} =$ 0.26 versus 0.23 for ROA, 0.69 versus 0.74 for log sales, 0.93 versus 0.78 for log employment. Appendix B.1 proves identification on mover observations; Table 13 reports the full stayer and mover estimates across all five outcomes. Finally, first-stage F-statistics are large throughout, exceeding 1,800 for every outcome and 58,000 for the scale outcomes, so weak instruments are not a concern.

4.3 Heterogeneity in δ Across Firms

Persistence may vary across firms with the durability of their organizational capital. We document two patterns consistent with our prior intuition: variation across SIC industry sectors, since the durability of organizational decisions varies by sector, and variation across firm size, since larger organizations adjust more slowly.

First, estimating the persistence regression separately within each major SIC division reveals substantial cross-sector dispersion. For log sales, where the headline cross-sector spread is largest, persistence ranges from $\hat{\delta} = 0.61$ in mining to $\hat{\delta} = 0.81$ in wholesale and retail trade, with light and heavy manufacturing close to $\hat{\delta} = 0.74$, services at $\hat{\delta} = 0.69$, and transportation and communications at $\hat{\delta} = 0.71$ (Table 2). The ordering is consistent with the durability of organizational capital across sectors: extractive industries are tied to depletable assets and turn over quickly; distribution networks and customer relationships in retail are slower to dissolve; human-capital-intensive services lie in the middle. Profitability outcomes show the same broad pattern with a smaller cross-sector spread.

[7] Labor adjustment notoriously does not vary continuously with firm performance (cf. Hamermesh, 1989). Guenzel et al. (2023) document that mass layoffs cause adverse health consequences in CEOs, which they attribute to stress; they argue that this is one mechanism through which labor adjustments within firms are slow.

Table 2: Persistence estimates by industry sector, log sales

Sector	$\hat{\delta}$	SE	Obs.
Mining (SIC 10–14)	0.614	(0.048)	1,788
Construction (SIC 15–17)	0.672	(0.038)	649
Manufacturing–light (SIC 20–29)	0.737	(0.049)	6,920
Manufacturing–heavy (SIC 30–39)	0.740	(0.021)	10,699
Transport & Comms (SIC 40–48)	0.707	(0.046)	2,018
Wholesale & Retail (SIC 50–59)	0.806	(0.025)	4,425
Services (SIC 70–89)	0.693	(0.038)	5,873
Pooled	*0.730*	*(0.018)*	*32,485*

Notes: Each row reports the IV estimate of δ in (8) on the stayer subsample within the listed sector. Firm, year, and CEO fixed effects are absorbed; SIC2×year is omitted within sector since it is collinear with the sample restriction. Standard errors are clustered at the firm level.

Persistence also rises with firm size. Splitting the sample at the median of log employment (Table 3), large firms exhibit higher point estimates of δ than small firms across all profitability outcomes. The difference is statistically significant for EBITDA/Assets ($\hat{\delta} = 0.39$ versus 0.24) and directionally consistent for return on assets ($\hat{\delta} = 0.23$ versus 0.16) and log sales ($\hat{\delta} = 0.69$ versus 0.68). Intuitively, larger organizations contain more layers between the CEO and operations, so individual decisions take longer to propagate or be reversed.

Table 3: Persistence estimates by firm size (median split on log employment)

	Small ($\hat{\delta}$)	Large ($\hat{\delta}$)
Return on Assets	0.158	0.234
	(0.085)	(0.073)
EBITDA / Assets	0.237	0.385
	(0.055)	(0.038)
Log Sales	0.677	0.692
	(0.024)	(0.018)

Notes: Stayer-subsample IV estimates with firm, year, and CEO fixed effects. Firms are split at the sample median of log employment. Standard errors clustered at the firm level.

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

How much do leaders matter for firm performance? The standard answer comes from Bertrand and Schoar (2003), who show that individual CEOs have identifiable "styles" that account for a meaningful share of cross-firm variation in performance. Their approach estimates

a two-way fixed effects specification

$$y_{t,f} = \alpha_f + \gamma_t + \theta_{m_{t,f}} + \varepsilon_{t,f}, \tag{9}$$

and reports the standard deviation of the estimated CEO effects $\widehat{\sigma}_\theta := \sqrt{\text{Var}(\hat{\theta}_m)}$ as the answer to the question: "how much does output change when we move one standard deviation up the CEO quality distribution?" This specification implicitly imposes $\delta = 0$ because it assumes that the CEO's full effect materializes immediately upon appointment and vanishes immediately upon departure.

Under our framework, that assumption generates two biases in opposite directions, characterized by Propositions 2 and 3. The static estimate $\widehat{\sigma}_\theta$ overstates the per-period direct effect of a one-standard-deviation increase in CEO quality and understates the long-run effect of permanently shifting CEO quality up by one standard deviation. With expected CEO tenure μ, $\widehat{\sigma}_\theta$ equals $\sigma_\theta/[(1-\delta)+\delta/\mu]$, which lies strictly between σ_θ (the per-period flow) and $\sigma_\theta/(1-\delta)$ (the steady-state value of better CEOs). The incremental R^2 from CEO fixed effects similarly understates the share of output variance attributable to the cumulative effect of past managers, by the factor $(1-\delta^2)/[1-\delta^2(1-1/\mu)^2]$. Both biases arise because, around CEO transitions, the predecessor's accumulated effect on output is unrelated to the incoming CEO's identity and therefore appears as residual variation in the static regression.

We estimate (9) on our CEO–firm panel with firm, year, and CEO fixed effects, and recover both the per-period and the long-run quantities using two complementary approaches. The first is a direct estimation: given $\hat{\delta}$ from Table 1, we form quasi-residuals $q_{t,f} := y_{t,f} - \hat{\delta}\, y_{t-1,f}$, which satisfy $q_{t,f} = \tilde{\alpha}_f + \tilde{\gamma}_t + \theta_{m_{t,f}} + u_{t,f}$. The CEO-level mean of the quasi-residuals (after absorbing firm and year effects) estimates the per-period flow effect θ_m; dividing by $(1-\delta)$ converts it into the long-run effect, and the standard deviation across recovered long-run effects approximates $\sigma_\theta/(1-\delta)$. The second approach applies the closed-form corrections from Propositions 2 and 3 to the static $\widehat{\sigma}_\theta$, using the pooled $\hat{\delta}$ and the average CEO–firm spell length. Table 4 reports both quantities.

Table 4: Dispersion of CEO effects: static vs. persistence-corrected estimates

	Return on Assets	Log Sales	Log Employment
Panel A: Static TWFE (à la Bertrand and Schoar)			
$\widehat{\sigma}_\theta$ (raw TWFE estimate)	0.075	0.456	0.441
CEO adjusted R^2 (incremental)	0.087	0.045	0.043
Panel B: Long-run effect, direct estimation via quasi-residuals			
$\hat{\delta}$ (pooled)	0.222	0.742	0.768
	(0.054)	(0.014)	(0.012)
Long-run SD ($\sigma_\theta/(1-\delta)$)	0.083	0.618	0.787
Panel C: Analytical correction (Props. 2, 3)			
Per-period SD (σ_θ)	0.061	0.179	0.165
Long-run SD ($\sigma_\theta/(1-\delta)$)	0.079	0.694	0.712
Long-run / TWFE multiplier ($1/a$)	[×1.05]	[×1.52]	[×1.61]
CEO adjusted R^2 (long-run)	0.088	0.063	0.064
R^2 multiplier	[×1.02]	[×1.40]	[×1.48]
Avg. CEO tenure (μ, years)	5.5	5.5	5.4

Notes: Panel A reports estimates from the static specification (9) with firm, year, and CEO fixed effects, absorbing all three via iterative demeaning. The raw TWFE estimate $\widehat{\sigma}_\theta$ is the standard deviation of the estimated CEO fixed effects after partialling out firm and year effects, and lies between the per-period and long-run quantities (Proposition 2). The incremental CEO adjusted R^2 is the increase in adjusted R^2 from adding CEO fixed effects to a model with firm and year effects only; adjusted R^2 penalizes for the large number of CEO parameters estimated. For return on assets, the static estimate ($\widehat{\sigma}_\theta = 0.075$) is close to the corresponding estimate in Bertrand and Schoar (2003, Table VI) ($\widehat{\sigma}_\theta = 0.07$), though the comparison is approximate: Bertrand and Schoar define ROA as operating income before depreciation over book assets (`oibdp/at`), whereas our ROA uses net income over lagged total assets. Panel B forms quasi-residuals $q_{t,f} = y_{t,f} - \hat{\delta}\, y_{t-1,f}$ using the pooled persistence estimates, absorbs firm and year fixed effects, and recovers CEO effects as the CEO-level mean divided by $(1-\hat{\delta})$; the SD across recovered effects estimates $\sigma_\theta/(1-\delta)$. Panel C applies the closed-form corrections from Propositions 2 and 3: $\sigma_\theta = \widehat{\sigma}_\theta \cdot [(1-\delta) + \delta/\mu]$ and $\sigma_\theta/(1-\delta) = \widehat{\sigma}_\theta/a$. Bracketed multipliers show the correction factor relative to Panel A. Standard errors for $\hat{\delta}$ are in parentheses, clustered at the firm level.

The magnitude of the long-run correction is increasing in the persistence estimate. For return on assets ($\hat{\delta} \approx 0.22$) the correction is small: the direct estimate raises the SD from 0.075 to 0.083 (11 percent), and the analytical correction to 0.079 (5 percent). For log sales ($\hat{\delta} \approx 0.74$) the corrections are sizeable: the direct estimate is 0.618 (36 percent larger than $\widehat{\sigma}_\theta$) and the analytical correction is 0.694 (52 percent larger). The largest correction is for log employment ($\hat{\delta} \approx 0.77$): the direct estimate is 0.787 (78 percent larger) and the analytical correction is 0.712 (61 percent larger). The direct and analytical corrections are broadly consistent with one another, which gives us confidence that neither result is an artifact of a particular estimation strategy. The small remaining gap between them reflects the stylized geometric tenure distribution we assumed in Proposition 2.[8]

[8]The proposition allows higher-quality CEOs to enjoy longer tenures (good CEOs are not separated for pro-

The per-period direct effect σ_θ is also substantially revised for the scale outcomes. For log sales, $\sigma_\theta \approx 0.18$, well below the static estimate of $\widehat{\sigma}_\theta = 0.46$. Moving one standard deviation up the CEO quality distribution therefore has a much smaller immediate impact than the static regression suggests, but a much larger long-run effect. The static specification, by forcing $\delta = 0$, blends both into a single intermediate number that overstates the short run and understates the long run.

The contrast also has implications for how boards evaluate newly appointed CEOs. At the estimated $\hat{\delta} \approx 0.74$ for log sales, a new CEO who permanently replaces a predecessor realizes only $1 - \hat{\delta} = 26$ percent of the long-run scale differential in the first year, $1 - \hat{\delta}^2 = 45$ percent by year two, and $1 - \hat{\delta}^5 = 78$ percent by year five. For employment, the adjustment is slower still: at $\hat{\delta} \approx 0.77$, 23 percent of the long-run effect appears in the first year and 41 percent by year two. For profitability the adjustment is faster: at $\hat{\delta} \approx 0.22$, 78 percent of the long-run effect appears in the first year and 95 percent by year two. A board that assesses a new CEO on one to two years of sales growth will therefore overvalue CEOs who succeed predecessors who grew the firm, since the inherited scale props up early performance, and undervalue CEOs who succeed predecessors who shrank it. Our persistence estimates quantify the conventional wisdom that new leadership needs time, and show that the wisdom applies most strongly to scale outcomes.

4.5 Which CEO Characteristics Matter: Revisiting Cronqvist–Yu on Daughters and CSR

The variance decomposition measures how much CEOs matter for firm outcomes in aggregate. A related question is how they affect the firm choices that shape outcomes. A large empirical literature addresses this by regressing firm outcomes on CEO traits, and persistence distorts these regressions in the same way it distorts the variance decomposition. As an illustration, we revisit Cronqvist and Yu (2017), who show that firms led by CEOs with daughters increase firms' corporate social responsibility (CSR) ratings, and interpret the result as evidence that having daughters shapes CEOs' social preferences and through them firm policy. CSR is a natural setting for the persistence correction, since diversity policies, employee benefits, environmental programs, and community relations are built up over time and do not reset when a new CEO arrives.

We use the CEO-children data from Cronqvist and Yu (2017), merge it to ExecuComp and Compustat, and construct CSR scores from the KLD data following their definition: strengths

ductive reasons); it only rules out separation in response to an idiosyncratic low-performance year. If manager tenure is quality-dependent in additional ways, the analytical correction, which treats μ as exogenous to manager quality, may be further biased.

minus concerns across community, diversity, employee relations, environment, human rights, and product categories, normalized so that the minimum is zero. The sample covers 1992–2012, matching the original study.[9] Our matched sample contains 2,243 firm-years of non-missing CEO-family and CSR data across 219 firms. A static replication recovers the published result: under Cronqvist and Yu's industry-and-year design we estimate a daughter effect of 0.86, close to their 0.91. Cronqvist and Yu also report a firm-fixed-effects specification (their Table 11) with a daughter effect of 0.94. The static specification with firm and year fixed effects is

$$CSR_{t,f} = \beta D_{m_{t,f}} + X'_{t,f}\Gamma + \alpha_f + \gamma_t + \varepsilon_{t,f}, \tag{10}$$

where $D_{m_{t,f}}$ indicates that the CEO has at least one daughter and $X_{t,f}$ includes CEO age, age squared, CEO gender, and the number of CEO children. The coefficient β is the static effect of replacing a CEO without a daughter by a CEO with a daughter, holding fixed the firm's inherited CSR level only through the firm fixed effect.

We then estimate the dynamic analog

$$CSR_{t,f} = \delta CSR_{t-1,f} + \beta D_{m_{t,f}} + X'_{t,f}\Gamma + \alpha_f + \gamma_t + u_{t,f}, \tag{11}$$

instrumenting $CSR_{t-1,f}$ with $CSR_{t-2,f}$ and requiring the lags to be consecutive calendar years.

Given a data-generating process as in Equation (11), $CSR_{t-1,f}$ will be correlated with the error term $u_{t,f}$, and thus using an identical proof to those in Section 3, if one instruments $CSR_{t-1,f}$ with $CSR_{t-2,f}$, one will recover the true δ, β, and the fixed effects. While other models could be estimated, we add a lagged dependent variable to Cronqvist and Yu's own firm-fixed-effects specification (their Table 11), changing their design only by allowing CSR to depend on its own past. If the true generating process had $\delta = 0$, our model would return $\delta = 0$.

[9] Extended the original study would require additional data on CEO's children, a data set which does not exist commercially and Cronqvist and Yu (2017) hand-collected.

Table 5: CEO daughters and CSR: static vs. persistence-corrected estimates

	Static (no lagged CSR)	Dynamic IV (persistence-corrected)
$\hat{\beta}$ (CEO has daughter)	0.935	0.638
	(0.600)	(0.370)
$\hat{\delta}$ (persistence)	—	0.597
		(0.048)
Long-run effect $\left[\hat{\beta}/(1-\hat{\delta})\right]$	0.935	1.585
		(0.940)
Firm FE	Yes	Yes
Year FE	Yes	Yes
CEO controls	Yes	Yes
Observations	1,772	1,772

Notes: The dependent variable is the KLD CSR score, normalized so its minimum is zero. *CEO has daughter* indicates that the CEO has at least one daughter. Both columns include firm and year fixed effects and control for CEO gender, age, age squared, and the number of CEO children; conditioning on the number of children is the source of as-good-as-random variation in daughter status in Cronqvist and Yu (2017). Both columns are estimated on the common sample of firm-years with two consecutive prior-year CSR observations (1,772 firm-years, 193 firms), so the static and dynamic estimates are directly comparable. The dynamic column estimates (11), instrumenting $CSR_{t-1,f}$ with $CSR_{t-2,f}$ and requiring consecutive calendar-year lags. The long-run effect is $\hat{\beta}/(1-\hat{\delta})$; its standard error is computed via the delta method from the joint variance-covariance matrix of $(\hat{\beta}, \hat{\delta})$. Standard errors clustered at the firm level are in parentheses.

Having a daughter raises CSR by 0.94 in the static specification. Once lagged CSR is included the flow effect falls to 0.64, as Proposition 2 predicts, but CSR is persistent ($\hat{\delta} = 0.60$), so the long-run effect is $0.64/(1-0.60) = 1.59$, about 1.7 times the static estimate on the same sample.[10]

Discussion: The same correction applies to many other influential papers attempting to estimate the effects of CEO characteristics on firm outcomes: Benmelech and Frydman (2015) on military-background CEOs and conservative policies, Malmendier and Tate (2005) on overconfident CEOs and investment, and Bandiera et al. (2020) on "leader" versus "manager" types and productivity. Bandiera et al. (2020) document the persistence directly, finding that the productivity benefit of a leader CEO takes roughly three years to materialize (their Table 5). Re-estimating any of these regressions with a lagged dependent variable would raise their long-run magnitudes above the published static estimates.

Together, the two exercises show that ignoring persistence biases two distinct classes of leader-effect estimates downward. The variance decomposition (how much leaders matter in

[10]Under Cronqvist and Yu's industry-and-year fixed effects, $\hat{\delta} = 0.86$ and the long-run effect is 1.55; under firm-by-industry-year fixed effects, $\hat{\delta} = 0.52$ and the long-run effect is 1.23.

aggregate) understates the long-run dispersion of CEO effects on firm scale by 50–60 percent. The trait regressions (which leader characteristics shape firm choices) understate the long-run effect of CEO daughters on CSR by about 70 percent.

Finally, the CEO analysis shows that ignoring persistence biases the measurement of leader effects, but quantifying the value of a policy that changes how leaders are selected requires exogenous variation in who gets appointed to a large subsample of firms. The public sector provides such variation: the institutional rules governing the appointment of officials change over time and generate quasi-experimental shifts in leader quality and publicly available data. The next section applies our framework to two such reforms.

5 Policy Implications: The Value of Better Leader Selection

Two policy reforms changed how senior leaders were selected in their respective settings: the Warren Fisher Reform in the British Colonial Office, which ended patronage in favor of merit-based selection, and the 2003 introduction of competitive recruitment for senior public-sector positions in Chile, including the CEOs of public hospitals. Xu (2018) estimates that patronage reduced colony revenue by approximately four percent, and Muñoz and Otero (2025) estimates that the Chilean reform reduced public hospital mortality by approximately eight percent. Both estimates come from static specifications that capture only the contemporaneous flow effect of the reform on the outcome. We extend each analysis by incorporating persistence, and show that in both settings the long-run gain from merit-based selection is several times larger than the static estimate suggests.

5.1 Patronage in the British Colonial Service

Xu (2018) estimates that patronage, the allocation of British colonial governorships on the basis of social connections to the Secretary of State for the Colonies, reduced colony revenue by approximately four percent. Therefore, Xu (2018) concludes that Warren Fisher Reform, which ended patronage and replaced it with merit-based selection, resulted in a four-percent increase in revenue. The static specification Xu (2018) uses, however, captures only the contemporaneous flow effect of connectedness on revenue. We extend their analysis by incorporating persistence and show the long-run effect from eliminating patronage is higher than previously estimated.

Data. We apply our framework to the colonial governor dataset assembled by Xu (2018), who digitized over 3,000 volumes of historical personnel and public finance records to con-

struct an individual-level panel of the British Colonial Office. The dataset covers the universe of 456 colonial governors serving across 70 colonies from the establishment of the Colonial Office in 1854 to its dissolution in 1966. The data combine four primary sources: the Colonial Office Lists (1860–1966), which document governor postings, backgrounds, and salaries; the Colonial Blue Books (1821–1949), which compile annual administrative statistics on public finance, demographics, and trade for each colony; biographical data from DeBrett's and the UK *Who's Who*; and genealogical data from the online database *The Peerage*.

In the language of our model, colonies correspond to plants (p) and governors correspond to managers (m). As in Xu (2018), our main outcome variable $y_{t,p}$ is the log of total colony revenue in GBP. Revenue generation is a natural measure of governor performance in this setting: under the revenue imperative of the British Empire, colonies were expected to be financially self-sustaining, and governors exercised direct control over public finances as a core duty of their office. The panel structure of the data, with governors serving terms averaging approximately six years and rotating across colonies, provides the variation in manager assignments needed to separately identify the stayer and mover persistence parameters.

The estimation sample contains 4,687 governor-colony-year observations. Of these, 3,618 are stayer observations, in which the same governor serves in consecutive periods, and 948 are mover observations, in which a new governor takes over from the previous one; the remaining 121 observations are the first year of each colony's panel, for which the stayer/mover classification is undefined. The time dimension of the panel spans over a century, which is what allows the deeper-lag IV strategy to identify δ within each colony. Finally, as described in Section 3, the identification assumptions rely on governor assignment not being correlated with lagged productivity.

Persistence Estimation. We estimate the levels specification (8) on consecutive colony-years where the same governor is in place at $t-1$ and t, with colony, year, and current-governor fixed effects, and an additional fixed effect for the colony's tropics indicator $\times$ year cell. The tropics–year (for parsimony with the SIC-year term in the CEO analysis) terms absorb time-varying shocks specific to tropical or non-tropical colonies, such as commodity-price cycles or weather, that would otherwise show up as serial dependence in colony revenue. We instrument $y_{t-1,p}$ with $y_{t-2,p}$ and cluster standard errors at the colony level. Table 6 reports the headline estimate.

The estimated persistence of log colony revenue is $\hat{\delta} = 0.726$ (SE $= 0.047$), with a first-stage F-statistic of 1,681. About three-quarters of a colony's prior-period revenue level carries over into the next period, with the remaining quarter accounted for by current-period inputs (including the current governor). Without the tropics–year fixed effect, the estimate is $\hat{\delta} =$

0.891 (SE = 0.018); the simpler specification conflates colony-level persistence with time-varying shocks correlated with tropical geography.

As a consistency check we estimate δ separately on stayer and mover subsamples, where mover years are colony-years in which a new governor took over. We use the baseline specification without tropics–year fixed effects for this comparison: with only 771 mover observations, the stricter spec is too saturated for stable mover estimates. The stayer estimate is $\hat{\delta}_S = 0.891$ (SE = 0.018); the mover estimate is $\hat{\delta}_M = 0.921$ (SE = 0.028). The estimates remain close, and we fail to reject equality.

Table 6: Persistence in colony revenue

	Headline (tropics×year FE, stayer)	Stayer/mover consistency check (baseline FE)
$\hat{\delta}$ (stayer)	0.726	0.891
	(0.047)	(0.018)
$\hat{\delta}$ (mover)	—	0.921
		(0.028)
First-stage F (stayer)	1,681	5,721
First-stage F (mover)	—	4,982
H_0: $\delta_S = \delta_M$ (p-val)	—	> 0.10
Observations (IV sample)	2,891	2,891 / 771

Notes: Each column reports IV estimates from the levels specification $y_{t,p} = \delta\, y_{t-1,p} + \alpha_p + \gamma_t + \theta_{m_{t,p}} + u_{t,p}$, with $y_{t-1,p}$ instrumented by $y_{t-2,p}$. The headline column adds tropics–year fixed effects to the baseline colony, year, and governor fixed effects, and is restricted to stayer observations. The consistency-check column reports separate stayer and mover estimates without the tropics–year terms. Standard errors are clustered at the colony level.

Additional Outcome Variables. As in the CEO data, persistence should vary across outcomes that differ in their responsiveness to managerial decisions. We estimate δ for log total expenditures and for social unrest, two outcomes in Xu (2018) with coverage above 80 percent. The expectation is that fiscal outcomes shaped by long-lived spending commitments will exhibit high persistence, while unrest, which responds quickly to enforcement and local shocks, will exhibit low persistence.

Table 7: Persistence estimates: additional outcome variables

	Stayer $\hat{\delta}_S$ (tropics×year FE)	Stayer $\hat{\delta}_S$ (baseline FE)
Log Total Expenditures	0.779	0.878
	(0.036)	(0.020)
Social Unrest	−0.06	0.30
	(0.274)	(0.220)
First-stage F, expenditures	1,820	4,805
First-stage F, unrest	115.1	123.8
Observations (IV, expenditures)	2,840	2,840
Observations (IV, unrest)	3,566	3,566

Notes: Each column reports the stayer-subsample IV estimate of δ in the specification $y_{t,p} = \delta\, y_{t-1,p} + \alpha_p + \gamma_t + \theta_{m_{t,p}} + u_{t,p}$, with $y_{t-1,p}$ instrumented by $y_{t-2,p}$. The first column adds the tropics–year fixed effect; the second omits it. Standard errors are clustered at the colony level.

Expenditures are nearly as persistent as revenue ($\hat{\delta}_S = 0.78$ under the headline tropics-year specification, 0.88 without it), a direct parallel to the high persistence of CEO log employment in Table 1: both reflect a resource footprint that managers build and successors inherit. Social unrest exhibits low and statistically indistinguishable persistence under either specification ($\hat{\delta}_S = -0.06$ with SE 0.27 under tropics-year, 0.30 with SE 0.22 without it). Unrest is shaped by short-run enforcement decisions and local shocks rather than by accumulated administrative capacity, and a governor changeover can reset its trajectory within a year or two.

Revisiting the Cost of Patronage and the Value of the Warren Fisher Reform. A central finding of Xu (2018) is that patronage (the allocation of governorships on the basis of social connections to the Secretary of State for the Colonies) reduced colony revenue by approximately four percent. The same estimate measures the gain from the Warren Fisher Reform, which replaced patronage with merit-based selection: colonies that received a merit-selected rather than a connected governor recorded four percent higher revenue per year. Both estimated are derived from the same static specification

$$y_{ist} = \beta\, c_{it} + \gamma' x_{it} + \nu_{is} + \tau_t + \varepsilon_{ist}, \tag{12}$$

where y_{ist} is log total colony revenue, c_{it} is an indicator for whether governor i is socially connected to the incumbent Secretary of State at time t, ν_{is} is a governor-colony fixed effect, τ_t is a year fixed effect, and x_{it} includes spell-length dummies and the number of colonies served. Here the subscript convention follows Xu (2018): i indexes the governor, s the colony, and t the year. The governor-colony fixed effect restricts identification to within-appointment

variation: the same governor in the same colony performs differently in years when political turnover in London changes their connectedness status. Xu (2018) estimates $\hat{\beta} = -0.040$ (SE $= 0.017$), which we replicate as $\hat{\beta} = -0.040$ (SE $= 0.015$) using the same specification.

This estimate implicitly imposes $\delta = 0$: it assumes that the effect of connectedness on colony revenue is fully and immediately realized in the current period. Under our framework, the governor's incentive distortion enters revenue contemporaneously but the colony's revenue process is itself persistent, so the period-by-period revenue gap accumulates over the governor's tenure. When a governor's incentives deteriorate under patronage, the revenue decline depreciates only slowly even after the governor has been replaced; conversely, when a connected governor loses their connection following a change of Secretary of State, the prior period's compromised revenue still feeds into current revenue at rate δ. A static regression captures only the contemporaneous flow effect and misses the accumulated component.

To quantify this understatement, we augment Xu (2018)'s specification with a lagged dependent variable. Returning to the notation of our model (with p indexing colonies and t indexing years):

$$y_{t,p} = \delta\, y_{t-1,p} + \beta\, c_{it} + \gamma' x_{it} + \alpha_p + \theta_{m_{t,p}} + \tau_t + \zeta_{i(p),t} + u_{t,p}, \tag{13}$$

instrumenting $y_{t-1,p}$ with $y_{t-2,p}$ and including colony, governor, year, spell-length, and tropics–year fixed effects. The additive colony and governor fixed effects replace Xu's interacted governor-colony fixed effect: under the dynamic specification, the governor-colony interaction would absorb much of the within-spell variation that identifies δ, since connectedness changes within governor-colony spells are driven by relatively short-lived Secretary of State turnover.[11] The specification nests Xu's static model as the special case $\delta = 0$, and the long-run effect of connectedness is $\beta/(1-\delta)$.

Table 8 reports the results. The estimated persistence parameter is $\hat{\delta} = 0.799$ (SE $=$ 0.023), close to our headline stayer estimate of $\hat{\delta} = 0.726$ from Table 6 and below the simpler $\hat{\delta}_S = 0.891$ from the baseline specification. The difference is consistent with two features of the cost-of-patronage regression: it restricts to the subsample with complete biographical data (`full` $= 1$ in Xu, reducing observations from 4,687 to 3,510), and it includes the connectedness indicator and spell-length dummies, which absorb some of the variation that the persistence-only specification would attribute to the lagged dependent variable. The contemporaneous flow effect of connectedness is $\hat{\beta} = -0.038$ (SE $= 0.017$), slightly smaller in magnitude than Xu's static estimate because the lagged dependent variable absorbs part of what the

[11] In Appendix D we report results using Xu's original governor-colony fixed effect. The estimated persistence is lower ($\hat{\delta} = 0.33$), consistent with the more saturated specification absorbing within-spell variation; however, the qualitative conclusion that the static estimate understates the long-run cost is unchanged.

static regression attributes to the current connectedness indicator. The implied long-run effect is $\hat{\beta}/(1-\hat{\delta}) = -0.186$ (SE $= 0.080$): the accumulated long-run cost of patronage is roughly five times larger than the static estimate suggests.

Table 8: The cost of patronage: static vs. persistence-corrected estimates

	Static (Xu replication)	Dynamic IV (persistence-corrected)
$\hat{\beta}$ (connected)	−0.040	−0.038
	(0.015)	(0.017)
$\hat{\delta}$ (persistence)	—	0.799
		(0.023)
Long-run effect $\left[\hat{\beta}/(1-\hat{\delta})\right]$	−0.040	−0.186
		(0.080)
Colony FE	Yes	Yes
Governor FE	Gov-colony	Yes
Year FE	Yes	Yes
Spell-length FE	Yes	Yes
Tropics×year FE	—	Yes
Observations	3,510	3,370

Notes: The dependent variable is log total colony revenue in GBP. The sample is restricted to observations with complete biographical data (`full` $= 1$), following Xu (2018). The static column replicates Xu (2018), Table 5, Column 1, with governor-colony, year, and spell-length fixed effects; standard errors are clustered at the governor–secretary-of-state level, as in the original. The dynamic IV column estimates (13) with additive colony, governor, year, spell-length, and tropics–year fixed effects, instrumenting $y_{t-1,p}$ with $y_{t-2,p}$; standard errors are clustered at the colony level. The long-run effect is $\hat{\beta}/(1-\hat{\delta})$; its standard error is computed via the delta method using the joint variance-covariance matrix of $(\hat{\beta}, \hat{\delta})$ from the IV regression.

Two adjustments interact to deliver the fivefold ratio. The dynamic specification attenuates the flow coefficient only slightly, from −0.040 to −0.038, as the lagged dependent variable absorbs part of what the static regression assigns to current connectedness; division by $(1-\hat{\delta}) = 0.201$ then amplifies the flow into a long-run estimate of −0.186. The two factors combine as $(0.038/0.040) \times 1/(1-0.799) \approx 0.95 \times 5.0 \approx 4.7$, matching the estimated ratio $0.186/0.040 = 4.7$.

The long-run estimate also gives a revised valuation of the Warren Fisher Reform, which ended patronage. Under the static coefficient, the reform raises colony revenue by approximately four percent per colony-year, the flow effect of switching from a connected to a merit-selected governor. The long-run effect is roughly nineteen percent. Each year of merit-selected governance raises the revenue flow by 3.8 percent, persistence carries that flow forward, and the steady-state revenue level of a reformed colony lies nearly five times as far above the patronage benchmark as the static estimate implies. The Warren Fisher Reform was therefore

much more valuable than a static reading of the cost of patronage would suggest.

5.2 Public Hospital CEO Selection in Chile

Muñoz and Otero (2025) estimates that the introduction of competitive selection for senior public-sector executives in Chile, including the CEOs of public hospitals, reduced in-hospital mortality at affected hospitals by approximately eight percent. The reform, enacted by the Chilean Congress in 2003 and rolled out gradually starting in 2005, replaced opaque and discretionary appointment with a public, competitive, and transparent merit-based selection process, and raised the base pay of newly appointed CEOs to narrow the differential with the private sector. Like the Warren Fisher Reform, it is a change in how leaders are selected rather than which leaders are appointed in a given period. We extend Muñoz and Otero's analysis by incorporating persistence and show that the long-run reduction in mortality is roughly 2.5 times the static estimate.

Data. We apply our framework to the hospital-quarter panel from Muñoz and Otero (2025), which covers 188 Chilean public hospitals over 2005Q1–2019Q4 (60 quarters, 11,074 hospital-quarter observations). Of these hospitals, 88 were ever treated by the reform, with adoption rolled out over 2005Q4 onwards. The main outcome variable $y_{t,p}$ is the log of the quarterly hospital-level in-hospital death rate, defined as deaths in the hospital divided by total inpatients. In the language of our model, hospitals correspond to plants (p) and CEOs correspond to managers (m). The reform indicator $Reform_{t,p}$ turns on for hospital p at the quarter in which the hospital adopts the new system, and remains on thereafter. The quarterly frequency gives the panel enough depth for the deeper-lag IV to identify δ within each hospital under the same assumptions used in the CEO and colonial settings.

Persistence Estimation. We estimate persistence on consecutive hospital-quarter observations where the same CEO is in place at $t-1$ and t, using the levels specification (8) with hospital, quarter, and current-CEO fixed effects. The estimated persistence of log mortality is $\hat{\delta} = 0.838$ (SE $= 0.035$), with a first-stage F-statistic of 235. About 84 percent of last quarter's deviation from a hospital's mean death rate persists into the current quarter. The high persistence is consistent with the view that the determinants of mortality, including patient mix, staffing, capital stock, organizational routines, and managerial practices, are themselves persistent at the hospital level. Note that this data is at the quarterly level, thus taking $\hat{\delta}^4 \simeq .49$ would result in an apples-to-apples comparison with the previous estimates.

Revisiting the Reform's Effect on Mortality. The headline finding of Muñoz and Otero (2025) is that the reform reduced in-hospital mortality by approximately eight percent. The estimate comes from a stacked event-study specification with hospital and quarter fixed effects and a rich set of patient-mix controls (age bin shares, gender shares, insurance type shares, and 31 Elixhauser comorbidity diagnosis shares), and gives $\hat{\beta} = -0.081$ (SE $= 0.024$).[12] The static specification implicitly imposes $\delta = 0$: it assumes that the change in selection regime alters mortality immediately through the contemporaneously appointed CEO, with no part of the reform's effect carrying forward.

To incorporate persistence we estimate

$$y_{t,p} = \delta\, y_{t-1,p} + \beta\, Reform_{t,p} + \alpha_p + \gamma_t + u_{t,p}, \tag{14}$$

instrumenting $y_{t-1,p}$ with $y_{t-2,p}$ and clustering standard errors at the hospital level. The dynamic specification departs from Muñoz and Otero's static in two ways. First, it operates on the unstacked hospital-quarter panel rather than on a stacked event-study sample, because the duplication of control hospitals across event cells in the stacked design distorts the within-hospital lag structure that identifies δ. Second, it omits Muñoz and Otero's patient-mix controls. These are time-varying compositional measures that themselves co-evolve with mortality; conditional on $y_{t-1,p}$, which is a sufficient statistic for past patient mix in the AR(1) model, including them double-controls and absorbs precisely the persistence signal that the LDV-IV is designed to extract.[13]

Table 9 reports the results. The contemporaneous flow effect is $\hat{\beta} = -0.032$ (SE $= 0.009$), much smaller than the static estimate because the lagged dependent variable absorbs part of what the static regression attributes to the current reform indicator. Meanwhile, the long-run effect, which can be thought of as keeping the competitive appointment process in place, and thus receiving on-average $\hat{\beta}$ higher-quality managers in each period, would result in a long-run effect of $\hat{\beta}/(1-\hat{\delta}) = -0.200$ (SE $= 0.050$). This effect corresponds to a steady-state mortality

[12] The stacked event-study design follows Cengiz et al. (2019), in which the control group for each treated unit is the set of not-yet-treated units within the event window; this avoids the negative-weighting issues of standard two-way fixed effects under staggered adoption.

[13] Appendix E reports the dynamic specification with Muñoz and Otero's full control set. The estimated persistence falls to $\hat{\delta} = 0.71$ (SE $= 0.04$), the contemporaneous flow effect to $\hat{\beta} = -0.009$, and the implied long-run effect to -0.029, consistent with the controls absorbing variation in $y_{t-1,p}$ that the LDV-IV would otherwise use. On the subsample of hospitals with at least sixteen quarters of data on each side of adoption, where the buildup phase is small relative to the steady-state period, the long-run effect even with controls included is -0.060, larger than the static estimate of -0.038 on the same sample. The static estimate on the unstacked panel without patient-mix controls is $\hat{\beta} = -0.181$ (SE $= 0.035$), considerably larger in magnitude than the stacked estimate of -0.081. The gap reflects negative-weighting in standard TWFE under staggered adoption, which the stacking is designed to address. Our long-run estimate of -0.200 is therefore in the same range as the apples-to-apples unstacked static, and substantially larger than Muñoz and Otero's stacked headline.

reduction of approximately eighteen percent from operating under the merit-based selection regime, roughly 2.5 times the static estimate reported by Muñoz and Otero (2025).

Table 9: The effect of merit-based CEO selection on hospital mortality: static vs. persistence-corrected estimates

	Static (Muñoz and Otero replication)	Dynamic IV (persistence-corrected)
$\hat{\beta}$ (Reform)	−0.081	−0.032
	(0.024)	(0.009)
$\hat{\delta}$ (persistence)	—	0.838
		(0.035)
Long-run effect $\left[\hat{\beta}/(1-\hat{\delta})\right]$	−0.081	−0.200
		(0.050)
Hospital FE	Yes	Yes
Quarter FE	Yes	Yes
Patient-mix controls	Yes	No
Sample	Stacked event-study	Unstacked panel
Observations	205,453	9,490

Notes: The dependent variable is the log hospital-quarter in-hospital death rate, weighted by hospital scale (total inpatients in 2005). The static column reproduces Muñoz and Otero (2025) Table 1, Column 1, using the stacked event-study sample and patient-mix controls (age bin shares, gender shares, insurance type shares, and 31 Elixhauser diagnosis shares). The dynamic IV column estimates (14) on the unstacked hospital-quarter panel without patient-mix controls, instrumenting $y_{t-1,p}$ with $y_{t-2,p}$. The long-run effect is $\hat{\beta}/(1-\hat{\delta})$; its standard error is computed via the delta method using the joint variance-covariance matrix of $(\hat{\beta}, \hat{\delta})$ from the IV regression. Standard errors are clustered at the hospital level.

Once persistence is accounted for, the long-run reduction in mortality from operating under merit-based selection is substantially larger than the static estimate suggests. As in the British Colonial Service, the reform was therefore considerably more valuable than a static reading of its effect would imply.

Mechanism: CEO Type and CEO Quality. The long-run effect of the reform is concentrated in the CEOs that merit selection actually delivered. We document this in two complementary decompositions, paralleling Muñoz and Otero's mechanism analysis. The first decomposes the reform's effect by an observable CEO characteristic, management training. Pre-reform, 99 percent of public hospital CEOs were doctors without managerial qualifications; post-reform, about 35 percent of the appointed pool acquired managerial qualifications. Allowing the reform's effect to differ between CEOs with and without management training, we estimate a long-run mortality reduction of −0.225 (SE = 0.059) for management-trained CEOs and −0.183 (SE = 0.059) for those without; both deliver a long-run reduction, with management-

trained CEOs delivering the larger one.

The second decomposition uses estimated AKM CEO fixed effects to measure CEO quality, following Muñoz and Otero's Table 5. We use their empirical-Bayes-shrunk CEO fixed effects, available for the subsample of CEOs in identified connected sets, and split CEOs at the within-connected-set median into high-quality (lower mortality FE) and low-quality (higher mortality FE) groups. The estimated long-run effects are −0.291 (SE = 0.089) for high-quality CEOs and −0.039 (SE = 0.077) for low-quality CEOs. The long-run mortality reduction is therefore concentrated almost entirely in the half of post-reform appointments where merit selection delivered higher-talent managers; hospitals that drew a low-quality CEO under merit selection show no detectable long-run improvement.[14]

Table 10 reports both decompositions.

Table 10: Mechanism: heterogeneity in the long-run effect by CEO characteristics

	$\hat{\beta}$ (flow)	$\hat{\delta}$	Long-run
Panel A: By management training			
Reform × Management-trained CEO	−0.037		−0.225
	(0.011)		(0.059)
Reform × Not management-trained CEO	−0.030	0.838	−0.183
	(0.010)	(0.036)	(0.059)
Observations		9,430	
Panel B: By estimated CEO quality (AKM fixed effects)			
Reform × High-quality CEO	−0.040		−0.291
	(0.019)		(0.089)
Reform × Low-quality CEO	−0.005	0.861	−0.039
	(0.011)	(0.052)	(0.077)
Observations		5,855	

Notes: Each panel reports IV estimates from a single regression that replaces the Reform indicator in (14) with its interactions with the CEO-type or CEO-quality indicator. Panel A uses the same sample as the headline dynamic specification; Panel B restricts to the subsample of CEOs in identified AKM connected sets and absorbs connected-set fixed effects in addition to hospital and quarter fixed effects. The flow effect $\hat{\beta}$ is reported for each interaction; the persistence parameter $\hat{\delta}$ is shared within each panel because it comes from one regression. The long-run effect for each row is $\hat{\beta}/(1-\hat{\delta})$; its standard error is computed via the delta method using the joint variance-covariance matrix from the IV regression. Standard errors are clustered at the hospital level. Quality labels are below versus above the within-connected-set median of Muñoz and Otero's empirical-Bayes-shrunk AKM CEO fixed effects; lower FE corresponds to lower mortality contribution and is labeled "high-quality."

[14] As in Muñoz and Otero's Table 5, the CEO-quality labels are themselves AKM fixed effects estimated on the same mortality outcome, so the comparison across quality groups inherits their identifying choice rather than relying on an independent quality measure.

6 Conclusion

This paper develops an econometric framework for estimating the persistence of productivity in firms and other organizations. Plant output evolves as $y_{t,p} = \delta y_{t-1,p} + \theta_{m_{t,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}$; the persistence parameter δ governs how long past output carries into current output, and the standard two-way fixed effects specification is the special case $\delta = 0$. We show that δ is point-identified using an instrumental variables strategy adapted from Anderson and Hsiao (1982), and that ignoring persistence understates the long-run dispersion of manager effects and the variance share attributable to managers.

The empirical analysis has two parts. In U.S. public-company CEO data, persistence varies sharply across outcomes: profitability measures exhibit low persistence ($\hat{\delta} \approx 0.23$–$0.39$), while firm scale persists at $\hat{\delta} \approx 0.74$ for log sales and $\hat{\delta} \approx 0.78$ for log employment. Ignoring this persistence biases two prominent estimates from the CEO literature. The Bertrand and Schoar (2003) variance decomposition understates the long-run dispersion of CEO effects on firm scale by 50–60 percent, so CEOs matter more for firm performance than the static literature reports. The Cronqvist and Yu (2017) estimate of how CEO daughter status shapes CSR understates the long-run effect by about 70 percent, so observable leader characteristics shape firm choices more than static estimates suggest. In the colonial governor data of Xu (2018), we estimate $\hat{\delta} \approx 0.73$ for log colony revenue in the stayer specification and a long-run cost of patronage nearly five times the static estimate, implying that the Warren Fisher Reform, which replaced patronage with merit-based selection, was much more valuable than the static evaluation suggests. Applying the same framework to the 2003 introduction of merit-based selection for senior public-sector executives in Chile, including the CEOs of public hospitals (Muñoz and Otero, 2025), we estimate $\hat{\delta} \approx 0.84$ for log hospital mortality and a long-run mortality reduction roughly 2.5 times the static estimate. Across these settings, retrospective evaluations of public sector reforms that rely on short-run windows will systematically understate the social returns to acquiring better managers.

Several open questions follow from this analysis. Persistence itself varies across outcomes and institutional settings, and the pattern we document is suggestive: outcomes that reflect accumulated institutional capacity and organizational capital (firm scale, CSR policies, fiscal revenue) exhibit meaningful persistence, while outcomes tied to short-run performance (profitability and social unrest) do not. 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 allow δ to vary with the firm-level environment in which different manager types operate,

separating organizational adjustment speed from manager-style dispersion. With longer time series, one could relax the AR(1) form of the persistence process and estimate richer lag dynamics. The literature has long documented that managers matter; our analysis suggests it has also understated the long-run scale of that 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.

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

Benmelech, Efraim and Carola Frydman, "Military ceos," *Journal of financial Economics*, 2015, *117* (1), 43–59.

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," Technical Report, 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 "The Handbook of the Economics of Corporate Governance," Vol. 1, Elsevier, 2017, pp. 383–539.

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

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.

Guenzel, Marius, Clint Hamilton, and Ulrike Malmendier, "CEO social preferences and layoffs," *Work. Pap*, 2023.

Hamermesh, Daniel, "Labor Demand and the Structure of Adjustment Costs," *American Economic Review*, 1989, *79* (4), 674–89.

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.

Malmendier, Ulrike and Geoffrey Tate, "CEO overconfidence and corporate investment," *The journal of finance*, 2005, *60* (6), 2661–2700.

Minni, Virginia, "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs," *Quarterly Journal of Economics*, 2026. Advance Access publication 23 March 2026.

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

Murphy, Kevin J, "Executive compensation," in "Handbook of Labor Economics," Vol. 3, Elsevier, 1999, pp. 2485–2563.

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

Weidmann, Ben, Joseph Vecci, Farah Said, David J Deming, and Sonia R Bhalotra, "How do you find a good manager?," Technical Report, National Bureau of Economic Research 2024.

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

Appendix A. Model Selection

This appendix reports the diagnostic evidence used to choose the dynamic manager model. The baseline empirical specification is

$$y_{t,p} = \delta y_{t-1,p} + \theta_{m(t,p)} + \alpha_p + \gamma_t + \epsilon_{t,p},$$

where p indexes the plant, firm, or colony; $m(t,p)$ is the manager in charge at date t; α_p are unit fixed effects; γ_t are year fixed effects; and $\epsilon_{t,p}$ is the idiosyncratic innovation. The parameter δ measures persistence in outcomes after accounting for manager identity and fixed effects. In the CEO panel the fixed effects are firm, year, CEO, and SIC2×year; in the CSR panel they are firm and year; in the colonial governor panel they are colony, year, governor, and tropics×year; and in the public-hospital panel they are hospital and quarter. The stock-return placebo uses annual log stock-price returns in the CEO panel with the same fixed effects: firm, year, CEO, and SIC2×year. Further, the baseline identification follows Anderson and Hsiao (1982), and we use the asymptotically equivalent within-group estimator as outlined in Alvarez and Arellano (2003). Further, the main results for this AR(1) specification uses $y_{t-2,p}$ as the instrument.

The diagnostics below ask a series of questions. First, when using non-parametric tests for persistence, does persistence emerge? In summary, yes, all the main specifications in the baseline analysis exhibit persistence. The question then turns to what kind of persistence emerges and whether AR-1 a good econometric model of the persistence that emerges.

Throughout, we note that even if the true data-generating process involves higher-order dynamics, our AR(1) specification represents a strictly weaker restriction than the implicit AR(0) maintained in prior work such as Bertrand and Schoar (2003). Since our diagnostics confirm that persistence is present, ignoring it entirely introduces a more severe misspecification than allowing for first-order dynamics.

Testing for AR(0) Adequacy: Wooldridge Test

The first diagnostic tests whether the static fixed-effect model is adequate. Under the null hypothesis of AR(0) dynamics and iid errors in levels, the residuals from the fixed-effect regression have a mechanical implication after first differencing: if r_{it} denotes the within-transformed residual, then $\text{Corr}(\Delta r_{it}, \Delta r_{i,t-1}) = -1/2$. We therefore regress Δr_{it} on $\Delta r_{i,t-1}$ using cluster-robust standard errors and test whether the slope equals -0.5. Rejection implies that the level innovation is serially correlated under the static model and that persistence remains in the residual. The next subsection discusses how to model that persistence.

Table 11 reports the test. All non-placebo rows use the stayer subsample, matching the sample used to identify δ. We reject AR(0) for all five CEO outcomes, for CSR, for log colonial revenue, for log colonial expenditures, and for log in-hospital mortality at the ten-percent level. Additionally, to demonstrate that out empirical strategy does not mechanically generate spurious persistence, we conduct a placebo test where the outcome variable is stock market returns. Unlike organizational or physical capital, stock returns in efficient markets should reflect information immediately and exhibit no serial correlation. The stock-return placebo correctly fails to reject AR(0), consistent with the efficient market hypothesis. Stock returns are excluded from the main outcome variables because they reflect investor expectations rather than organizational capacity, and any persistence in them operates through different channels than the durable firm-level outcomes the model targets.

Table 11: Wooldridge AR(0) test

Setting	Outcome	$\hat{\rho}$	SE	z vs. -0.5	p-value	Verdict
CEO	Return on Assets	-0.295	0.011	18.780	< 0.001	Reject
CEO	Operating ROA	-0.341	0.012	12.899	< 0.001	Reject
CEO	EBITDA/Assets	-0.174	0.013	24.960	< 0.001	Reject
CEO	Log Sales	0.102	0.021	28.520	< 0.001	Reject
CEO	Log Employment	0.099	0.011	53.502	< 0.001	Reject
CSR	KLD CSR Score	-0.244	0.019	13.235	< 0.001	Reject
Colonial	Log Revenue	-0.237	0.035	7.514	< 0.001	Reject
Colonial	Log Total Expenditures	-0.282	0.072	3.024	0.002	Reject
Hospital	Log Death Rate	-0.474	0.014	1.851	0.064	Reject (10%)
Placebo (CEO FE)	Annual log price return	-0.492	0.006	1.245	0.213	Do not reject

Notes: The table reports the coefficient from a regression of Δr_{it} on $\Delta r_{i,t-1}$, where r_{it} is the residual from the corresponding static fixed-effect model. Standard errors are clustered at the unit level. The null is $\rho = -0.5$, the value implied by iid level errors under AR(0).

Testing for AR(1) Adequacy: Arellano–Bond AR(2) Test

The second diagnostic tests whether the AR(1) specification leaves serial correlation in the level innovation. After estimating the first-differenced AR(1) equation by IV, the differenced residual $\Delta\hat{\epsilon}_{it}$ should exhibit first-order serial correlation mechanically, because it contains $\epsilon_{it} - \epsilon_{i,t-1}$. Under iid level innovations, however, it should not exhibit second-order serial correlation. The Arellano–Bond AR(2) test therefore examines whether $\Delta\hat{\epsilon}_{it}$ is correlated with $\Delta\hat{\epsilon}_{i,t-2}$. Here, rejection indicates that the level innovation remains serially correlated after including one lag of the dependent variable and it also threatens the validity of $y_{t-2,p}$ as an instrument.

Table 12 reports the AR(1) estimates and the AR(2) diagnostic p-values. The headline $\hat{\delta}$ column reports the main-text estimate for each application: 0.231, 0.386, 0.319, 0.744, and

0.784 for the five CEO outcomes; 0.596 for CSR; 0.726 for log colonial revenue; and 0.838 for log in-hospital mortality. First stages clear conventional thresholds in all panels.

The AR(2) test gives a mixed but informative set of pass/fail results. At the five-percent level, it fails to reject for CEO return on assets ($p = 0.974$), CEO operating ROA ($p = 0.082$), CEO log sales ($p = 0.058$), and CSR ($p = 0.486$). It rejects for CEO EBITDA/assets ($p = 0.011$), CEO log employment ($p = 0.036$), log colonial revenue ($p < 0.001$), and log in-hospital mortality ($p < 0.001$). At the one-percent level the CEO rejections no longer hold, while the colonial-revenue and in-hospital mortality rejections remain. Thus the AR(1) model passes the AR(2) diagnostic cleanly for the main CEO profitability outcomes, CEO log sales, and CSR, but the table flags remaining serial-correlation concerns for several highly persistent outcomes.

Table 12: Model-selection diagnostics for autoregressive order

Setting	Outcome	Wooldridge p	AR(1) headline $\hat{\delta}$	AR(1) deeper $\hat{\delta}$	AB-AR(2) p	AR(2) both IV $\hat{\delta}_1+\hat{\delta}_2$	AR(2) predet. $\hat{\delta}_1+\hat{\delta}_2$
CEO	Return on Assets	< 0.001	0.231 (0.056)	0.340 (0.337)	0.974	6.525 (39.935)	-0.170 (2.110)
CEO	Operating ROA	< 0.001	0.386 (0.059)	0.416 (0.233)	0.082	-0.016 (2.184)	0.450 (0.449)
CEO	EBITDA / Assets	< 0.001	0.319 (0.036)	0.325 (0.149)	0.011	1.422 (4.163)	0.294 (0.291)
CEO	Log Sales	< 0.001	0.744 (0.015)	0.810 (0.022)	0.058	0.861 (0.060)	-0.256 (0.948)
CEO	Log Employment	< 0.001	0.784 (0.010)	0.783 (0.014)	0.036	1.032 (0.305)	0.836 (0.252)
CSR	KLD CSR Score	< 0.001	0.596 (0.048)	0.400 (0.062)	0.486	-0.159 (0.206)	3.042 (6.241)
Colonial	Log Revenue	< 0.001	0.726 (0.047)	1.147 (0.124)	< 0.001	1.767 (0.720)	1.187 (0.194)
Hospital	Log Death Rate	0.064	0.838 (0.035)	0.983 (0.067)	< 0.001	0.889 (0.075)	0.968 (0.074)
Placebo	Annual log price return	0.213	0.467 (0.035)	0.138 (0.079)	< 0.001	0.106 (0.127)	0.202 (0.071)

Notes: The table reports the consolidated model-selection diagnostics. Standard errors are in parentheses. The AR(1) headline column reports the main-text estimate of δ for each outcome (single instrument $y_{t-2,p}$, with sample and additional fixed effects as specified in the corresponding empirical section). The remaining columns are computed on a stricter subsample that requires lags up to $y_{t-4,p}$ to be available at consecutive time intervals within each unit: the deeper AR(1) column uses $y_{t-3,p}$ as the single excluded instrument; the Arellano–Bond AR(2) p-value is from the AR(1) IV with $y_{t-2,p}$ and $y_{t-3,p}$ as instruments on that subsample; the AR(2) both-IV column treats both $y_{t-1,p}$ and $y_{t-2,p}$ as endogenous and instruments each with a deeper lag; the AR(2) predetermined column treats $y_{t-2,p}$ as a predetermined control and instruments only $y_{t-1,p}$ with $y_{t-3,p}$.

Robustness: Deeper-Lag Instruments

The headline AR(1) specification uses $y_{t-2,p}$ as the instrument. A stricter check replaces it with $y_{t-3,p}$. This deeper instrument remains valid under an MA(1) level innovation. It therefore relaxes the i.i.d. innovation assumption used by the headline Anderson–Hsiao specification. This diagnostic asks whether the estimated persistence parameter is stable when identification comes only from a deeper lag. Stability would support the AR(1) interpretation and instability indicates either instrument invalidity in the headline specification or that the deeper lag is too noisy to predict outcomes.

The deeper-lag estimates are broadly consistent with the main-text headline for several outcomes. Return on assets shifts from 0.231 to 0.340, EBITDA/assets from 0.319 to 0.325, log sales from 0.744 to 0.810, and log employment from 0.784 to 0.783. Operating ROA rises

modestly from 0.386 to 0.416. The larger standard errors for the profitability outcomes reflect the weaker first stage of the deeper instrument: with persistence as low as $\hat{\delta} \approx 0.23$–$0.39$, the autocorrelation between $y_{t-3,p}$ and $y_{t-1,p}$ is of order δ^2, leaving less predictive power than the one-period-closer $y_{t-2,p}$. CSR shifts from 0.596 to 0.400 under the deeper instrument; the smaller CSR panel offers limited identifying variation at deeper lags. Log colonial revenue rises from 0.726 to 1.147, with a standard error of 0.124, so the deeper-lag check preserves high persistence but does not pin down the exact magnitude tightly. Log in-hospital mortality moves from 0.838 to 0.983, remaining in the upper range. The stock-return placebo moves from 0.467 to 0.138, consistent with the view that the positive headline return coefficient is not a stable persistence estimate.

Robustness: AR(2) Specifications

The final diagnostic asks whether a second autoregressive lag improves the model. We estimate two AR(2) variants. The first treats both $y_{t-1,p}$ and $y_{t-2,p}$ as endogenous and instruments both lagged dependent variables with $y_{t-3,p}$ and $y_{t-4,p}$. The object reported in Table 12 is $\hat{\delta}_1 + \hat{\delta}_2$, the cumulative persistence implied by the two-lag model. The second treats $y_{t-2,p}$ as a predetermined control and instruments only $y_{t-1,p}$ with $y_{t-3,p}$; the table again reports $\hat{\delta}_1 + \hat{\delta}_2$.

The AR(2) evidence does not favor a higher-order specification. For log sales, log employment, and log in-hospital mortality, the AR(2) sums $\hat{\delta}_1 + \hat{\delta}_2$ lie close to the corresponding AR(1) estimates (0.861 versus 0.744, 1.032 versus 0.784, and 0.889 versus 0.838, respectively), so the second lag reallocates persistence rather than adds it. For CEO profitability and CSR, AR(2) standard errors are too wide to discriminate AR(1) from any alternative. For log colonial revenue, AR(2) is precise but unstable across variants: the sum is 1.767 with both lags endogenous and 1.187 with $y_{t-2,p}$ predetermined. The predetermined variant is also unstable in the high-persistence CEO outcomes: log sales collapses to -0.256 with a standard error of 0.948, expected when $y_{t-3,p}$ becomes nearly collinear with $y_{t-2,p}$ once $y_{t-2,p}$ enters as a control. For the outcomes where the AB-AR(2) test rejected AR(1)+iid, moving to AR(2) does not resolve the rejection; the residual serial correlation persists regardless of how many lags appear. AR(2) therefore adds noise without adding identifying power.

Conclusion

In general, the model-selection evidence supports a first-order dynamic specification for the main productivity and governance analysis. AR(0) is too sparse for the primary CEO outcomes, CSR, log colonial revenue, log colonial expenditures, and log in-hospital mortality; the stock-return placebo correctly fails to reject AR(0) under the CEO fixed effects (firm, year,

CEO, and SIC2×year). The headline Anderson–Hsiao AR(1) estimates have strong first stages, and deeper-lag instruments generally preserve the same persistence ranking, though they flag instability for the stock-return placebo. AR(2) specifications add no systematic explanatory power and create severe precision and weak-instrument problems, especially when outcomes are highly persistent. We therefore use the AR(1) Anderson–Hsiao IV specification as the preferred model, while treating diagnostics that fail the deeper-lag or AR(2) checks as outcome-specific validity concerns rather than evidence in favor of a higher autoregressive order.

B Proofs and Additional Results

Proof of Proposition 1. First-differencing (1) between t and $t-1$ eliminates the plant fixed effect:

$$\Delta y_{t,p} = \delta\, \Delta y_{t-1,p} + (\theta_{m_{t,p}} - \theta_{m_{t-1,p}}) + \Delta\gamma_t + \Delta\epsilon_{t,p}. \tag{15}$$

On the subsample where the manager does not change between $t-1$ and t (i.e., $S_{t,p}=1$, so that $m_{t,p} = m_{t-1,p}$), the manager term drops out:

$$\Delta y_{t,p} = \delta\, \Delta y_{t-1,p} + \Delta\gamma_t + \Delta\epsilon_{t,p}, \qquad \text{on } \{S_{t,p}=1\}. \tag{16}$$

Cross-sectional demeaning removes the time effect $\Delta\gamma_t$. For a stayer plant p (i.e., $S_{t,p}=1$),

$$(\Delta y_{t,p})^* = \delta\, (\Delta y_{t-1,p})^* + (\Delta\epsilon_{t,p})^* - \overline{\Delta\theta}_t, \tag{17}$$

where $\overline{\Delta\theta}_t := \frac{1}{P}\sum_{p'=1}^{P}(\theta_{m_{t,p'}} - \theta_{m_{t-1,p'}})$. Under Assumption 1 (E2), manager assignment is conditionally random, so $\mathbb{E}[\theta_{m_{t,p'}}] = \mathbb{E}[\theta_{m_{t-1,p'}}] = 0$ for all p', and hence $\mathbb{E}[\overline{\Delta\theta}_t] = 0$. It follows that $\overline{\Delta\theta}_t$ does not contribute to any of the population moments used below, and we may write

$$(\Delta y_{t,p})^* = \delta\, (\Delta y_{t-1,p})^* + (\Delta\epsilon_{t,p})^* \tag{18}$$

in the sense that all moment conditions involving $Z_{t,p}S_{t,p}$ are identical whether or not $\overline{\Delta\theta}_t$ is included, since $\mathbb{E}[\overline{\Delta\theta}_t \cdot Z_{t,p}S_{t,p}] = \overline{\Delta\theta}_t \cdot \mathbb{E}[Z_{t,p}S_{t,p}]$ and $\mathbb{E}[\overline{\Delta\theta}_t] = 0$.

Exogeneity. We verify (2) for the instrument $Z_{t,p} = y_{t-2,p}$. The differenced error $\Delta\epsilon_{t,p} = \epsilon_{t,p} - \epsilon_{t-1,p}$ depends only on shocks at dates t and $t-1$. Since $y_{t-2,p}$ is $\mathcal{F}_{t-2}$-measurable, Assumption 1 (E1) gives $\mathbb{E}[\epsilon_{t,p} \mid \mathcal{F}_{t-2}] = \mathbb{E}[\epsilon_{t-1,p} \mid \mathcal{F}_{t-2}] = 0$. By Assumption 1 (E2), the stayer indicator $S_{t,p}$ is independent of $\{\epsilon_{t,p}, \epsilon_{t-1,p}\}$ conditional on $\mathcal{F}_{t-2}$. Therefore $\mathbb{E}[\Delta\epsilon_{t,p} \cdot y_{t-2,p} \cdot S_{t,p}] = 0$, and cross-sectional demeaning preserves this orthogonality.

Relevance. We verify (3) for $Z_{t,p} = y_{t-2,p}$. On the subsample $\{S_{t,p}=1\}$, output at $t-1$ satisfies

$y_{t-1,p} = \delta y_{t-2,p} + \theta_{m_{t-1,p}} + \alpha_p + \gamma_{t-1} + \epsilon_{t-1,p}$. The lagged first difference is thus:

$$\Delta y_{t-1,p} = (\delta - 1) y_{t-2,p} + \theta_{m_{t-1,p}} + \alpha_p + \gamma_{t-1} + \epsilon_{t-1,p}. \tag{19}$$

Applying the cross-sectional demeaning operator $(\cdot)^*$ removes γ_{t-1} and yields:

$$(\Delta y_{t-1,p})^* = (\delta - 1) y^*_{t-2,p} + \theta^*_{m_{t-1,p}} + \alpha^*_p + \epsilon^*_{t-1,p}. \tag{20}$$

The relevance condition holds if the covariance $\mathbb{E}[(\Delta y_{t-1,p})^* y_{t-2,p} \mid S_{t,p} = 1]$ is non-zero. Expanding (20):

$$\mathbb{E}[(\Delta y_{t-1,p})^* y_{t-2,p}] = (\delta - 1)\mathrm{Var}(y^*_{t-2,p}) + \mathrm{Cov}(\theta^*_{m_{t-1,p}} + \alpha^*_p, y_{t-2,p}) + \mathbb{E}[\epsilon^*_{t-1,p} y_{t-2,p}].$$

The first term is strictly negative since $\delta < 1$. The second term is strictly positive because $y_{t-2,p}$ depends positively on the persistent plant and manager effects. Finally, $\mathbb{E}[\epsilon^*_{t-1,p} y_{t-2,p}] = 0$ by Assumption (E1).

Identification. Multiplying (18) by $Z_{t,p} \cdot S_{t,p}$ and taking expectations:

$$\mathbb{E}[(\Delta y_{t,p})^* Z_{t,p} S_{t,p}] = \delta\, \mathbb{E}[(\Delta y_{t-1,p})^* Z_{t,p} S_{t,p}] + \underbrace{\mathbb{E}[(\Delta \epsilon_{t,p})^* Z_{t,p} S_{t,p}]}_{=0}.$$

Dividing through by $\mathbb{E}[(\Delta y_{t-1,p})^* Z_{t,p} S_{t,p}] \neq 0$ yields (4). □

Proof of Proposition 2. Iterating (1) backward, we have $y_{t,p} = \sum_{j=0}^{\infty} \delta^j (\theta_{m_{t-j,p}} + \epsilon_{t-j,p})$. By the law of iterated expectations and the properties of the manager process, $\mathbb{E}[\theta_{m_{t-j,p}} \mid \theta_{m_{t,p}} = \theta] = q^j \theta$. Thus:

$$\mathbb{E}[y_{t,p} \mid \theta_{m_{t,p}} = \theta] = \sum_{j=0}^{\infty} \delta^j q^j \theta = \frac{\theta}{1 - \delta q}.$$

The static TWFE model estimates the conditional mean $\hat{\theta}_m = \mathbb{E}[y_{t,p} \mid m_{t,p} = m] = \theta_m/(1-\delta q)$. The estimated standard deviation of manager productivity is $\widehat{\sigma}_\theta = \sqrt{\mathrm{Var}(\hat{\theta}_m)} = \sigma_\theta/(1-\delta q)$. Substituting $q = 1 - 1/\mu$, the denominator is $1 - \delta(1 - 1/\mu) = (1-\delta) + \delta/\mu$. We can then write the estimated standard deviation as:

$$\widehat{\sigma}_\theta = \frac{\sigma_\theta}{(1-\delta) + \delta/\mu} = \frac{1-\delta}{(1-\delta) + \delta/\mu} \cdot \frac{\sigma_\theta}{1-\delta},$$

yielding the expression in the proposition. □

Proof of Proposition 3. Since both measures share the same denominator, it suffices to compare their numerators. From Proposition 2, the numerator of R^2_{TWFE} is $\mathrm{Var}(\mathbb{E}[y_{t,p} \mid m_{t,p}]) =$

$\text{Var}(\theta_{m_{t,p}}/(1-\delta q)) = \sigma_\theta^2/(1-\delta q)^2$.

For the numerator of R^2_{true}, let $V = \text{Var}(\Theta_{t,p})$. Under the stationary manager process, $\text{Cov}(\theta_{m_{t-j,p}}, \theta_{m_{t-k,p}}) = \sigma_\theta^2 q^{|j-k|}$. Thus:

$$V = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \delta^{j+k}\text{Cov}(\theta_{m_{t-j,p}}, \theta_{m_{t-k,p}}) = \sigma_\theta^2 \left[\sum_{j=0}^{\infty}\delta^{2j} + 2\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}\delta^{j+k}q^{k-j}\right]$$
$$= \sigma_\theta^2\left[\frac{1}{1-\delta^2} + \frac{2\delta q}{(1-\delta^2)(1-\delta q)}\right] = \frac{1+\delta q}{(1-\delta^2)(1-\delta q)}\sigma_\theta^2.$$

The ratio of the numerators is:

$$\frac{R^2_{\text{TWFE}}}{R^2_{\text{true}}} = \frac{1/(1-\delta q)^2}{\frac{1+\delta q}{(1-\delta^2)(1-\delta q)}} = \frac{1-\delta^2}{1-(\delta q)^2}.$$

Substituting $q = 1 - 1/\mu$ yields the final result. □

B.1 Heterogeneous Persistence

We now 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 $X_{t,p}$ denote a discrete observable characteristic for plant p at time t, taking values in a finite set $\mathcal{X}$. We modify the output equation from Section 2 so that δ becomes a function of these observables:

$$y_{t,p} = \delta(X_{t,p})\, y_{t-1,p} + \theta_{m_{t,p}} + \alpha_p + \gamma_t + \epsilon_{t,p}, \qquad t \geq 1. \tag{21}$$

To identify the group-specific persistence $\delta(x)$ for each $x \in \mathcal{X}$, we adapt the argument from Proposition 1. 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 of plants where the manager does not change and the observable characteristic remains constant between $t-1$ and t. First-differencing (21) on this subsample eliminates the plant fixed effect, the manager fixed effect, and the heterogeneity in δ:

$$\Delta y_{t,p} = \delta(x)\,\Delta y_{t-1,p} + \Delta\gamma_t + \Delta\epsilon_{t,p}, \qquad \text{on } \{S_{t,p}(x) = 1\}. \tag{22}$$

Applying the cross-sectional demeaning operator $(\cdot)^*$ removes the time effect $\Delta\gamma_t$:

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

Assuming that the sequence of observables $X_{t,p}$ is predetermined with respect to the production shocks, the instrument $Z_{t,p} = y_{t-2,p}$ satisfies the conditional exogeneity condition

$\mathbb{E}[(\Delta\epsilon_{t,p})^* Z_{t,p} \mid S_{t,p}(x) = 1] = 0$. Given the relevance condition $\mathbb{E}[(\Delta y_{t-1,p})^* Z_{t,p} \mid S_{t,p}(x) = 1] \neq 0$, the group-specific persistence $\delta(x)$ is point-identified by the instrumental variables estimand:

$$\delta(x) = \frac{\mathbb{E}\big[(\Delta y_{t,p})^* Z_{t,p} S_{t,p}(x)\big]}{\mathbb{E}\big[(\Delta y_{t-1,p})^* Z_{t,p} S_{t,p}(x)\big]}. \tag{24}$$

For practical implementation, the analogue of the levels specification described in the previous subsection involves estimating the regression separately for each group $x \in \mathscr{X}$. 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 ($m_{t,p} = m_{t-1,p}$). 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. This split-sample approach preserves the asymptotic equivalence to the differenced estimator as $T \to \infty$ and avoids the need to interact the endogenous lagged dependent variable with fixed effects in a pooled specification.

Further, one might wonder whether persistence differs depending on whether the current manager is the same as the predecessor or a new arrival. We now show that 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 identification of δ_S follows from Proposition 1. To identify δ_M, we restrict attention to the subsample of movers where $D_{t,p} = 1$.

For any scalar variable $X_{t,p}$, let $X^{\dagger}_{t,p}$ denote the residual from the population linear projection of $X_{t,p}$ onto the span of time effects γ_t, current-manager effects $\theta_{m_{t,p}}$, and outgoing-manager effects $\theta_{m_{t-1,p}}$, computed on the mover subsample $\{D_{t,p} = 1\}$.

Proposition 4 (Identification of δ_M on switch observations)**.** *Let $y_{t,p}$ satisfy (26) with $\delta_M < 1$. Suppose there exists a scalar variable $Z_{t,p}$ such that, for some $t \geq 2$,*

$$\mathbb{E}\big[(\Delta\epsilon_{t,p})^{\dagger} Z_{t,p} \mid D_{t,p} = 1\big] = 0, \tag{27}$$

$$\mathbb{E}\big[(\Delta y_{t-1,p})^{\dagger} Z_{t,p} \mid D_{t,p} = 1\big] \neq 0. \tag{28}$$

Then δ_M is point-identified and satisfies

$$\delta_M = \frac{\mathbb{E}\big[(\Delta y_{t,p})^\dagger Z_{t,p} \mid D_{t,p} = 1\big]}{\mathbb{E}\big[(\Delta y_{t-1,p})^\dagger Z_{t,p} \mid D_{t,p} = 1\big]}. \tag{29}$$

Moreover, $Z_{t,p} = y_{t-2,p}$ satisfies (27)–(28) under Assumption 1.

Proof of Proposition 4. First-differencing (26) on the mover subsample $\{D_{t,p} = 1\}$ gives:

$$\Delta y_{t,p} = \delta_M \Delta y_{t-1,p} + (\theta_{m_{t,p}} - \theta_{m_{t-1,p}}) + \Delta\gamma_t + \Delta\epsilon_{t,p}. \tag{30}$$

Applying the residualization operator $(\cdot)^\dagger$ removes the time effects $\Delta\gamma_t$ and the manager terms $(\theta_{m_{t,p}} - \theta_{m_{t-1,p}})$, as both the current and outgoing manager identities are partialled out:

$$(\Delta y_{t,p})^\dagger = \delta_M (\Delta y_{t-1,p})^\dagger + (\Delta\epsilon_{t,p})^\dagger. \tag{31}$$

The remainder of the proof follows the same logic as Proposition 1. Under Assumption 1, $y_{t-2,p}$ is predetermined with respect to $\Delta\epsilon_{t,p}$, and the manager identities are independent of the shocks conditional on history. Thus the orthogonality condition holds. Relevance holds because $y_{t-2,p}$ remains correlated with $\Delta y_{t-1,p}$ through the plant fixed effect and the outgoing manager's productivity, both of which are embedded in the lagged levels. □

For practical implementation in levels, 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:

$$y_{t,p} = \delta_M y_{t-1,p} + \alpha_p + \gamma_t + \theta_{m_{t,p}} + \epsilon_{t,p}, \tag{32}$$

instrumenting $y_{t-1,p}$ with $y_{t-2,p}$ and absorbing plant, time, and current-manager fixed effects via iterative demeaning. In the levels formulation, the current manager's productivity is absorbed by the current-manager fixed effect, and unlike in the differenced equation of Proposition 4, there is no need to additionally control for the outgoing manager's identity. This is because the lagged manager's contribution enters the current period's output only through the inherited value $y_{t-1,p}$, which is explicitly controlled for on the right-hand side.

Table 13 reports the empirical stayer and mover estimates for the CEO–firm panel. The specification omits the SIC2×year fixed effect for the mover subsample, for the same reason the colonial consistency check uses the baseline rather than the tropics–year specification: the mover sample is roughly one-seventh the size of the stayer sample, and the more saturated specification yields unstable estimates.

Table 13: Stayer/mover consistency check: U.S. CEO–firm panel

	Stayer ($\hat{\delta}_S$)	Mover ($\hat{\delta}_M$)
Outcome: Return on Assets		
Estimate	0.209	0.262
Clustered SE	(0.061)	(0.180)
First-stage F	1,074	850
Observations (IV sample)	32,255	4,466
Outcome: Operating Return on Assets		
Estimate	0.372	0.350
Clustered SE	(0.066)	(0.197)
First-stage F	884	828
Observations (IV sample)	32,223	4,459
Outcome: EBITDA / Assets		
Estimate	0.305	0.544
Clustered SE	(0.038)	(0.226)
First-stage F	3,291	1,320
Observations (IV sample)	32,200	4,453
Outcome: Log Sales		
Estimate	0.730	0.691
Clustered SE	(0.018)	(0.108)
First-stage F	25,129	4,437
Observations (IV sample)	32,485	4,487
Outcome: Log Employment		
Estimate	0.777	0.925
Clustered SE	(0.012)	(0.033)
First-stage F	27,329	7,737
Observations (IV sample)	32,266	4,458

Notes: Each column reports IV estimates from the levels specification $y_{t,f} = \delta\, y_{t-1,f} + \alpha_f + \gamma_t + \theta_{m_{t,f}} + u_{t,f}$, with $y_{t-1,f}$ instrumented by $y_{t-2,f}$. Firm, year, and CEO fixed effects are absorbed via iterative demeaning; the SIC2×year fixed effect is omitted to preserve stable mover estimates. The stayer column restricts to firm-years in which the same CEO was in place in the prior period; the mover column restricts to firm-years in which a new CEO took over. Standard errors are clustered at the firm level.

C Additional Outcome Variables

The following table reports additional outcome variables for the analysis conducted in Section 5. These additional variables represent the remainder of colonial output data with coverage exceeding 60 percent.

Table 14: Heterogeneous persistence estimates (lower coverage outcomes)

	Stayer ($\hat{\delta}_S$)	Mover ($\hat{\delta}_M$)
Outcome: log customs revenue (GBP)		
Estimate	0.870	0.943
Clustered SE	(0.029)	(0.013)
First-stage F-statistic	5,662.4	6,674.3
Observations (IV sample)	2,139	566
Total subsample observations	3,618	948
Outcome: log internal revenue (GBP)		
Estimate	0.834	1.007
Clustered SE	(0.042)	(0.060)
First-stage F-statistic	1,299.0	848.0
Observations (IV sample)	2,113	558
Total subsample observations	3,618	948
Outcome: log public works expenditures (GBP)		
Estimate	0.743	1.138
Clustered SE	(0.053)	(0.047)
First-stage F-statistic	988.9	1,656.0
Observations (IV sample)	1,952	512
Total subsample observations	3,618	948

Notes: Each column reports IV estimates from the levels specification $y_{t,p} = \delta\, y_{t-1,p} + \alpha_p + \gamma_t + \theta_{m_{t,p}} + u_{t,p}$, with $y_{t-1,p}$ instrumented by $y_{t-2,p}$. The stayer column restricts to observations where the governor does not change; the mover column restricts to observations where a new governor takes over. Standard errors are clustered at the colony level.

D Robustness: Governor-Colony Fixed Effects

In the main text, our persistence-corrected reanalysis of Xu (2018) uses additive colony and governor fixed effects. Here we report results using Xu's original governor-colony fixed effect ν_{is}, which restricts identification to within-appointment variation. The specification is

$$y_{t,p} = \delta\, y_{t-1,p} + \beta\, c_{it} + \gamma' x_{it} + \nu_{is} + \tau_t + u_{t,p},$$

with $y_{t-1,p}$ instrumented by $y_{t-2,p}$ and standard errors clustered at the governor–secretary-of-state level, as in Xu (2018).

Table 15: Persistence-corrected cost of patronage: governor-colony fixed effects

	Static (Xu replication)	Dynamic IV (governor-colony FE)
$\hat{\beta}$ (connected)	−0.040	−0.050
	(0.015)	(0.014)
$\hat{\delta}$ (persistence)	—	0.335
		(0.063)
Long-run effect $\left[\hat{\beta}/(1-\hat{\delta})\right]$	−0.040	−0.075
		(0.022)
Governor-colony FE	Yes	Yes
Year FE	Yes	Yes
Spell-length FE	Yes	Yes
Observations	3,510	3,370

Notes: The dependent variable is log total colony revenue in GBP. The sample is restricted to observations with complete biographical data (`full` = 1), following Xu (2018). Both columns include governor-colony, year, and spell-length fixed effects. The dynamic IV column instruments $y_{t-1,p}$ with $y_{t-2,p}$. Standard errors are clustered at the governor–secretary-of-state level, as in Xu (2018). The long-run effect is $\hat{\beta}/(1-\hat{\delta})$; its standard error is computed via the delta method using the joint variance-covariance matrix of $(\hat{\beta}, \hat{\delta})$ from the IV regression.

The estimated persistence is $\hat{\delta} = 0.335$ (SE = 0.063), well below the $\hat{\delta} = 0.799$ obtained with additive fixed effects: the governor-colony fixed effect absorbs variation across a governor's tenure within a single colony, leaving only high-frequency fluctuations to identify δ. Since connectedness changes within a governor-colony spell are driven by Secretary of State turnover, which occurs on average every three years, the interacted fixed effect competes directly with the lagged dependent variable for the persistent component of within-spell variation.

Despite this attenuation, the qualitative conclusion is unchanged. The implied long-run effect is −0.075 (SE = 0.022), which is 1.9 times larger than the static estimate and statistically significant at the 5% level. Even under the most conservative fixed-effects structure, ignoring persistence leads to a meaningful understatement of the cost of patronage.

E Robustness: Patient-Mix Controls in the Hospital CEO Application

In the main text, our persistence-corrected reanalysis of Muñoz and Otero (2025) omits the time-varying patient-mix controls used in Muñoz and Otero's static specification: age bin shares, gender shares, insurance type shares, and 31 Elixhauser comorbidity diagnosis shares. Here we report the dynamic IV specification with the full control set included, both on the full sample and on a long-window subsample.

With the patient-mix controls included on the full sample, the dynamic IV gives $\hat{\delta} = 0.707$ (SE $= 0.042$), $\hat{\beta} = -0.009$ (SE $= 0.009$), and an implied long-run effect of -0.029 (SE $=$ 0.030). Including the controls shrinks both $\hat{\delta}$ and $\hat{\beta}$ relative to the headline specification, and the implied long-run effect falls below the static estimate. The reason is that the patient-mix controls are time-varying compositional measures that themselves co-evolve with mortality. Conditional on $y_{t-1,p}$, these compositional shares double-control: past patient mix has already entered current mortality through the lagged outcome, so including the shares as separate controls absorbs the persistence signal that the LDV-IV is designed to extract.

The observed pattern coincides with this interpretation. When we restrict to the 43 hospitals with at least sixteen quarters of data on each side of reform adoption, the buildup phase of the reform's effect is small relative to the steady-state period. On this long-window subsample, the dynamic specification with controls gives $\hat{\delta} = 0.621$, $\hat{\beta} = -0.023$, and an implied long-run effect of -0.060, larger than the static estimate of -0.038 on the same sample. With enough panel depth on either side of adoption, the LDV-IV machinery recovers persistence even with the controls included; on shorter panels the compositional controls dominate.

Table 16 reports the full comparison.

Table 16: Robustness: dynamic IV with patient-mix controls

	Headline (full sample)	With controls (full sample)	With controls (long-window sample)
$\hat{\beta}$ (Reform)	−0.032	−0.009	−0.023
	(0.009)	(0.009)	(0.013)
$\hat{\delta}$ (persistence)	0.838	0.707	0.621
	(0.035)	(0.042)	(0.115)
Long-run effect	−0.200	−0.029	−0.060
	(0.050)	(0.030)	(0.034)
Static (same sample, same controls)	−0.181	−0.064	−0.038
Patient-mix controls	No	Yes	Yes
Sample restriction	None	None	$\geq$ 16 pre/post quarters
Observations	9,490	9,490	2,290

Notes: All columns report IV estimates from (14) on the unstacked hospital-quarter panel, with hospital and quarter fixed effects, $y_{t-1,p}$ instrumented by $y_{t-2,p}$, and standard errors clustered at the hospital level. Column 1 is the headline specification from Table 9. Column 2 adds the patient-mix controls used in Muñoz and Otero's static specification: age bin shares, female share, age-bin × female interactions, the 31 Elixhauser diagnosis shares (Sh_DC1–Sh_DC31), and insurance type shares ($Fonasa_A$, $Fonasa_B$, $Fonasa_C$, $Fonasa_D$, $Fonasa_0$). Column 3 restricts the sample to hospitals with at least sixteen quarters of data on each side of reform adoption. The static row reports the unstacked TWFE estimate on the same sample, with the same control choice. The long-run effect is $\hat{\beta}/(1-\hat{\delta})$; its standard error is computed via the delta method.

F Monte Carlo Simulations

This subsection provides Monte Carlo simulations to validate the identification procedure developed in Section 3. Here, we fix values of δ_S and δ_M and generate synthetic data as follows. We assume that manager assignments are simulated through a stochastic rotation process where a manager leaves their current plant with a 20 percent probability. All managers who leave their plants and managers who are currently unassigned to plants are then uniformly at random assigned to plants. This process is done with P plants, M managers, and T time periods. The fixed effects for time and plants are drawn $\mathcal{N}(0,1)$. The manager effects are drawn from a distribution $\mathcal{N}(0,1/(1-\delta_S))$. The reason for this normalization is to ensure the aggregate variance of manager effects remains constant with respect to δ_S.

A qualitative feature throughout all the examples is that when the first-stage F-Statistic is high, the RMSE is quite low. The following table discusses the bias and standard deviation of the error averaged across 100 simulations for $\delta_S = 0.5$ and $\delta_M = 0.3$.

Table 17: Monte Carlo Results: Heterogeneous Persistence Estimation

	Stayer			Mover		
(P,T,M)	Bias	RMSE	F-Stat	Bias	RMSE	F-Stat
(20, 40, 30)	-0.127	0.276	14.3	-0.005	0.107	32.5
(50, 100, 80)	-0.027	0.062	128.9	-0.005	0.027	349.2
(100, 150, 150)	-0.014	0.033	442.9	-0.001	0.014	1206.6
(200, 200, 300)	-0.010	0.021	1220.2	-0.001	0.008	3451.9

We now show what occurs if the true δ values are quite low. Here, we set $\delta_S = \delta_M = 0.1$.

Table 18: Monte Carlo Results: Heterogeneous Persistence Estimation

	Stayer			Mover		
(P,T,M)	Bias	RMSE	F-Stat	Bias	RMSE	F-Stat
(20, 40, 30)	-0.145	1.200	1.3	-0.042	0.244	10.9
(50, 100, 80)	0.061	7.480	1.5	-0.016	0.057	111.0
(100, 150, 150)	-0.537	0.733	4.7	-0.005	0.029	383.7
(200, 200, 300)	-0.345	0.429	14.0	-0.003	0.016	1103.6

A second qualitative feature of both tables is that the bias is negative in almost every row. This is consistent with the asymptotic analysis in Alvarez and Arellano (2003), who show that under joint P, T asymptotics (i.e., if P/T converges to a constant) the within-groups, GMM, and LIML estimators for the AR(1) panel model all exhibit negative asymptotic biases, of orders $(1+\delta)/T$, $(1+\delta)/P$, and $(1+\delta)/(2P-T)$. The biases shrink rapidly as T and P grow, which our simulations also display. In contrast if P/T converges to zero, the sign of the bias

is also reassuring for our empirical conclusions: any finite-sample bias leads us to understate persistence, implying that, if anything, our results are understatements of the true effects.

The divergence in the first-stage F-statistics between the stayer and mover subsamples at low values of δ reflects the differential fixed-effects absorptions. For the stayer subsample, which is restricted to observations where the manager does not change ($m_{t,p} = m_{t-1,p}$), the within transformation absorbs the current manager's fixed effect, $\theta_{m_{t,p}}$. Because the manager identity is constant across these periods, this procedure simultaneously partials out the lagged manager's effect, $\theta_{m_{t-1,p}}$. Consequently, the only remaining systematic link between the endogenous regressor $y_{t-1,p}$ and the instrument $y_{t-2,p}$ is the persistence parameter δ. For low values of δ, this instrument is weak, yielding a low F-statistic. For the mover subsample, restricted to observations where a new manager takes over ($m_{t,p} \neq m_{t-1,p}$), the levels specification absorbs only the incoming manager's fixed effect; there is no need to control for the outgoing manager's identity. The unabsorbed productivity of the outgoing manager, $\theta_{m_{t-1,p}}$, therefore remains embedded in the endogenous lagged dependent variable. When the outgoing manager was in place during the instrumented period, this same effect is also present in $y_{t-2,p}$. This variation leads to a higher first-stage F-statistic.