Research Article

Robertas Zubrickas*

Delegated Contracting, Wage Compression, and Quiet Quitting

https://doi.org/10.1515/bejte-2024-0117
Received October 21, 2024; accepted March 16, 2026; published online April 7, 2026

Abstract: We argue that delegated contracting and transfer pricing can form a set of conditions leading to coarse pay incentives. The allocation mechanism of negotiated transfer pricing creates asymmetry in the intrafirm valuation of employee effort. For a divisional manager, the value of effort is determined by the transfer price but for the center by the contribution to firm profits with the transfer price coming as a cost. Divisional managers' incentives become misaligned with firm profit maximization, which has implications for employees' compensation. We demonstrate the optimality of payroll cost controls imposed on divisional managers despite that managers respond by compressing employees' wages and that employees respond by quiet quitting. We also consider an extension of the model where negotiated transfer pricing can create more distortion in larger firms. Differences in managerial incentives across small and large firms produce the firm-size wage effects: a higher average wage but less wage variation in larger firms.

Keywords: delegated contracting; wage compression; quiet quitting; cost controls; performance evaluation; firm-size wage effects

JEL Classification: D21; J30; C70

I thank the anonymous referees and the editor for their constructive comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The author reports there are no competing interests to declare.

***Corresponding author: Robertas Zubrickas**, Department of Economics, University of Bath, Claverton Down, Bath BA2 7AY, UK; and Faculty of Economics and Business Administration, Vilnius University, Sauletekio av. 9, Vilnius 10222, Lithuania, E-mail: r.zubrickas@bath.ac.uk.
https://orcid.org/0000-0002-6940-5175

1 Introduction

In economic analysis, there is hardly a wider gap between theory and practice than between the theory of incentives and the practice of incentives. Contrary to theoretical predictions based on merit pay, employees' productivity has little explanatory power of their wages which typically exhibit much compression. See Bishop (1987) and Frazis and Loewenstein (2006) for empirical evidence and Prendergast (1999) for a comprehensive overview of wage compression. Consequently, wage compression and underlying coarse incentive schemes are often viewed as inconsistent with standard incentive theory and instead explained by alternative approaches based on some form of utility interdependence.[1] However, employees' workplace behavior is a far cry from being driven by social concerns, reciprocity, or group identity but is rather characterized by a minimal fulfillment of job requirements, which is popularly known as *quiet quitting*.[2] At the same time, such behavior remains consistent with standard incentive theory, as it is optimal under coarse incentive schemes. This leaves the question open why firms provide coarse incentives given that employees appear to be individualistic.

In this paper, we argue that wage compression can arise from organizational frictions within firms rather than from worker preferences or behavioral considerations. The common organizational features of delegated contracting and transfer pricing can create asymmetry in the valuation of employee effort within the firm. For a divisional manager the value of employee effort is determined by the transfer price but for the center by the contribution to firm profits with the transfer price coming as a cost. Then, within a standard incentive framework we demonstrate for an environment with employee heterogeneity and without utility interdependence that wage compression or coarse employee screening can be a solution to the intrafirm problem of incentive design.

Our modeling approach builds on practical aspects of economic organization. We study the multi-divisional firm as a three-tier economic organization with delegated contracting and transfer pricing similarly to the *corrupted M-form* in Holmstrom and Tirole (1991). The three tiers are the center (also the residual claimant), divisional managers, and employees. In the firm, production is divided across separate divisions, each of which is headed by a divisional manager, and every

1 The leading examples of alternative theories are the gift exchange theory, fair wage-effort hypothesis, and group identity theory proposed by Akerlof (1982), Akerlof and Yellen (1990), and Akerlof and Kranton (2000), respectively.

2 The workplace survey of Gallup (2022) finds that over 50 % of US workers are quiet quitters with other 20 % being even actively disengaged from their work. Similar proportions have been observed at least since 2,000. For a popular account of the problem see, e.g., O'Connor (2022).

division supplies an input toward the final product. The organizational and production workings are as follows. First, a divisional manager and the center negotiate over the price at which the division supplies its input. Second, the manager is delegated to contract with the division's employees about their production effort subject to the cost controls imposed by the center, if any. In particular, the center has the authority to impose an upper bound on employee payroll expenses. The compensation of a divisional manager is set to be proportional to the division's internal profits. The informational setting that we study is where an employee's productivity is his private information.

In the model, we take the structure of the firm and its allocation mechanism of negotiated transfer pricing as given. Here, we outline our main assumptions about the organizational structure and their justifications. The assumption that it is optimal for the firm to produce inputs internally rather than procure them externally can be motivated by the transaction cost theory of the firm (Coase 1937) or the asset specificity theory (Williamson 1985). Transfer pricing is the principal method of intrafirm resource allocation, practiced by 90 % of multi-divisional firms (Merchant and Van der Stede 2007, p. 280). This practice can be explained by the impossibility or only partial applicability of market mechanisms for intrafirm transactions (Hirshleifer 1956; Arrow 1974; Amershi and Cheng 1990). In our model, the choice of negotiated transfer pricing can also encompass two other most common types of transfer pricing, market-based and cost-based, as they typically involve negotiated price adjustments to account for firm specificity.[3] The central assumption of delegated contracting is discussed in the *Related Literature* section, drawing on Mookherjee (2006). Delegation entails a transfer of "real authority" to managers and, consequently, the emergence of managerial bargaining power (Aghion and Tirole 1997). The usage of divisional profits rather than firm profits in determining managerial compensation can be explained by incentive reasons (Oyer and Schaefer 2005). Lastly, we allow the center to impose payroll cost controls, e.g. performance appraisal systems, which, as discussed later, are a common feature in most firms.

Our results are driven by asymmetry in the intrafirm valuation of employee effort embedded in delegated contracting and transfer pricing. First, we show that it is optimal for the center to impose a cap on employees' payroll expense aimed at correcting for misaligned managerial incentives, which otherwise steer employees

3 See Dutta and Reichelstein (2010) for an example of full cost transfer pricing with negotiated adjustments. Theoretical foundations of negotiated transfer pricing can be found in Edlin and Reichelstein (1995), Vaysman (1998), and Baldenius et al. (1999).

toward overproduction.[4] Our second result is the compression of employees' pay which emerges as a side effect of imposed cost controls. We show that when capped in available rewards, a divisional manager chooses an incentive scheme that pools an interval of productivity types at the highest wage available rather than screening them, as would occur without cost controls. In fact, an incentive scheme where the same wage is set for all productivity types can also emerge in equilibrium.

The results of the optimality of payroll cost controls and compressed pay offer a new perspective on the usage of job performance appraisal systems and their ubiquitous outcome of compression of appraisal ratings (Murphy and Cleveland 1995; Prendergast 1999). Job performance appraisal systems are typically put in place for the purpose of salary administration and effectively serve as payroll cost controls. Under such a system, line managers use appraisal ratings to evaluate and reward their employees' performance as ratings are money equivalent for employees. The range and monetary values of ratings are determined by the center and are used to control the payroll expense that managers can incur. The bounded scale of ratings makes a performance appraisal system equivalent to the pay-cap controls modeled here. Given our findings, we argue that the usage of performance appraisal systems is optimal from the center's perspective, whereas the compression of appraisal ratings is an equilibrium outcome of the resultant manager-employee relationship with cost controls.

Lastly, we show that our proposed mechanism of delegated contracting, negotiated transfer pricing, and cost controls can also address a separate, longstanding puzzle in labor economics: the firm-size wage effects. Empirically, larger firms tend to (i) pay higher average wages but (ii) exhibit lower wage dispersion.[5] We extend the main model with firm size to show how the same organizational frictions can help explain the firm-size wage effects. Specifically, we assume that the degree of incentive misalignment between the center and divisional managers increases with firm size, consistent with the view that larger organizations face greater internal control problems.[6] As a result, the constraints imposed by cost controls on

---

4 The view that negotiated transfer pricing can lead to managers' choosing suboptimal levels of production from the firm's perspective is well recognized in the management literature, see Kaplan and Atkinson (1998, p. 461).

5 The firm-size wage effects were first documented in the early work of Moore (1911); also see Stigler (1962), Brown and Medoff (1989), and Oi and Idson (1999). The open question is what factors that vary systematically across firms of different size can explain these effects (Troske (1999), Fox (2009), Cobb and Lin (2017)).

6 In the main text, we justify this extension by referencing the attenuation of incentives in larger organizations (Williamson (1985), Aghion and Tirole (1997)) and the bargaining literature (Binmore et al. (1986)), which suggests that bargaining power may be weakened in slower or more wage-setting, when combined with internal bargaining frictions, can lead to more compressed wage structures in larger firms.

The remainder of the paper is organized as follows. In Section 2, we present the framework of the nested agency model. In Section 3, we solve the model and present our main findings on wages and cost controls. In Section 4, we extend the analysis to explore the implications for firm-size wage effects. In Section 5, we review related literature. The last section concludes the study. All omitted proofs are in the Appendix.

2 Framework

Consider a single-product firm owned by "the center". In the firm, there are $N \geq 1$ production divisions, each supplying an input toward the final product. The input of division i, for $i = 1, \ldots, N$, can be furnished of various quality $q_i > 0$. The quality of the final product is given by production function $Q(q_1, \ldots, q_N)$, which is shaped by technological constraints and institutional factors external to the firm. The selling price of the final product per unit of quality is normalized to 1.

In every division, there are one employee and one manager, and it is the employee who produces the input. The production of the input of quality q costs the employee a disutility of $C(q)/\theta$, where θ is his privately known productivity level. The function $C(\cdot)$ is increasing, twice differentiable, and strictly convex with $C(0) = 0$ and $C'(0) = 0$. Productivity θ is distributed on interval $[\theta_L, \theta_H]$, $0 < \theta_L < \theta_H$, according to publicly known distribution $F(\cdot)$ with density $f(\cdot) > 0$ that satisfies the non-decreasing hazard rate condition. If paid r, the employee obtains a net utility of

$$U = r - \frac{C(q)}{\theta}. \tag{1}$$

The employee's reservation utility is 0. The manager is delegated to contract with the employee, which entails designing an incentive scheme as a schedule of pay-for-quality allocations $\{(r, q)\}$ subject to the payroll cost controls imposed by the center, if any. These controls take the form of a pay cap $\bar{r}$ on employee pay r, which is equivalent to the introduction of a performance appraisal system with the bounded rating scale of $[0, \bar{r}]$.

Let t denote a transfer price per unit of quality q produced in the division, which is negotiated by the divisional manager and the center. We follow Vaysman

complex decision-making environments. Empirical evidence consistent with a positive relationship between firm size and delegation is provided by Colombo and Delmastro (2004), who show that more complex organizational structures are associated with greater decentralization of authority.

(1998) in that transfer price negotiations take place after the center and the manager agree on the form of managerial compensation denoted by w. Assuming that the manager and the center are risk neutral, we can consider without loss of generality linear managerial compensation schemes

$$w = \phi + \beta(tq - r), \tag{2}$$

where ϕ is a constant and $\beta \in (0,1)$ is the manager's share of internal profits $tq - r$.[7] As our results on employee compensation do not qualitatively depend on managerial compensation scheme (ϕ, β), we take the form of managerial compensation in (2) as part of the firm's exogenously given structure. For ease of exposition, we set that $\phi = 0$, by which we also assume that the center finds it optimal to offer the manager high-powered incentives in his contracting with the employee or, put differently, the fixed-wage component ϕ of the managerial scheme is only of the second order compared to the incentive component β.[8]

The realized profits of the firm are given by

$$\pi = Q(q_1, \dots, q_N) - \sum_{i=1}^{N} w_i - \sum_{i=1}^{N} r_i, \tag{3}$$

where q_i, w_i, and r_i are division i's quality produced, managerial compensation, and employee pay, respectively. Throughout we assume that the non-negativity constraint on the profits is never binding, i.e. $\pi \geq 0$.

Finally, in every division the agency relationship develops as follows. First, the manager and the center negotiate over transfer price t. Second, the center can impose cost controls $\bar{r}$ on employee pay. Third, the manager designs employee contract $\{(r, q)\}$. With this timing structure we also capture the scenario when the center has discretion in revising its cost controls policy at any moment prior to managers designing incentive schemes for employees. In the paper we also discuss the robustness of our results to alternative timing, in particular, when transfer price and cost controls are simultaneously negotiated.

7 For agency problems with risk-neutral parties, Laffont and Tirole (1986) and Bhattacharyya and Lafontaine (1995) demonstrate that the restriction to linear compensation schemes is without loss of generality. See Stoughton and Talmor (1999) for an application of this result in the study of managerial bargaining power and determination of compensation contracts. Also see Melumad et al. (1992) who for a similar three-tier hierarchy with asymmetric information and risk neutrality demonstrate the optimality of linear compensation schemes between the center and the delegated manager.

8 This assumption ensures a role for transfer price negotiations. Generally, the determination of the optimal managerial compensation scheme (ϕ, β) would require the introduction of monitoring and contracting costs for managers. See, e.g., Holmstrom and Milgrom (1991).

3 Wages and Cost Controls

In this section, we solve the nested agency model by backward induction starting with the manager's problem of designing an incentive scheme for the employee.

3.1 Employee Incentive Scheme

The manager looks for the pay-for-quality incentive scheme $\{(r, q)\}$ that maximizes his expected compensation w given transfer price t and payroll cost controls $\bar{r}$. As productivity is private information, designing incentives for the employee is a monopolistic screening problem with a pay-cap constraint.

Applying the Revelation Principle, the manager looks for the direct incentive scheme $\{(r(\theta), q(\theta))\}_{\theta\in[\theta_L,\theta_H]}$ that maximizes own expected compensation

$$w(\bar{r}, t) = \max_{(r(\theta),q(\theta))} \int_{\theta_L}^{\theta_H} \beta\left(tq(\theta) - r(\theta)\right) f(\theta)\mathrm{d}\theta \tag{4}$$

s.t.

$$r(\theta) - \frac{C(q(\theta))}{\theta} \geq r(\widehat{\theta}) - \frac{C(q(\widehat{\theta}))}{\theta}, \tag{5}$$

$$r(\theta) - \frac{C(q(\theta))}{\theta} \geq 0, \tag{6}$$

$$r(\theta) \leq \bar{r} \quad \text{for every } \theta, \widehat{\theta} \in \left[\theta_L, \theta_H\right]. \tag{7}$$

Constraint (5) is the employee's incentive-compatibility constraint to report own productivity type truthfully, (6) is his individual-rationality constraint, and (7) is the pay-cap constraint imposed by the center.

We note that the manager's problem and its solution are analytically similar to the problem studied by Besanko et al. (1987), in which a monopolist seller designs a menu of price-quantity bundles subject to a price ceiling. The main property of the solution, presented in Proposition 1 below, is that if the pay-cap constraint is binding, then the manager pools an interval of types at the top of the type space for a uniform allocation. In other words, when constrained in employee rewards, the manager does not screen the most efficient types, in contrast to the case with unconstrained rewards. The reason is that with a binding pay cap, the gain from perfectly screening types cannot outweigh the opportunity cost of ensuring incentive compatibility.

Proposition 1. *Denote the starting point of the pooling interval by θ^+ and the quality level $q(\theta)$ for $\theta \in [\theta^+, \theta_H]$ by $\overline{q}$. The optimal incentive scheme $\{(r, q)\}$ is characterized by the following three conditions. Firstly, we have*

$$\theta^+ - \frac{1 - F(\theta^+)}{f(\theta^+)} = \frac{C'(\overline{q})}{t}, \tag{8}$$

secondly, for $\theta \leq \theta^+$

$$\left(t - \frac{C'(q(\theta))}{\theta}\right) f(\theta) - \frac{C'(q(\theta))}{\theta^2}\left(F(\theta^+) - F(\theta) + \frac{t}{C'(\overline{q})/\theta^+}(1 - F(\theta^+))\right) = 0, \tag{9}$$

and, thirdly,

$$r(\theta) = \frac{C(q(\theta))}{\theta} + \int_{\theta_L}^{\theta} \frac{C(q(\widetilde{\theta}))}{\widetilde{\theta}^2} \mathrm{d}\widetilde{\theta}. \tag{10}$$

In the proposition, equation (8) is the condition for the starting pooling point expressed by its “virtual type”, (9) specifies the optimal quality levels $q(\theta)$ for $\theta < \theta^+$, and (10) – the pay levels $r(\theta)$. The highest quality level $\overline{q}$ is found from (10) at $\theta = \theta^+$. When the pay-cap constraint is not binding, the solution is given by (9) and (10) with $\theta^+ = \theta_H$. Finally, should equation (9) have $q(\theta) < 0$ for some θ, then in the solution we set $q(\theta) = 0$ for those values of θ.

3.2 Cost Controls

Next, we examine whether the center finds it profitable to impose payroll cost controls $\overline{r}$. Intuitively, the center may wish to restrict the manager’s discretion in contracting with the employee if the center’s and the manager’s incentives are misaligned to the disadvantage of the center. Such disadvantageous misalignment can occur if the transfer price exceeds the marginal return from the quality supplied. As we see next, this intuition proves correct.

Let $p_i(\overline{r})$ denote the expected marginal return of input quality q_i at the managers’ optimal profile of qualities $\{q_1(\theta), \ldots, q_N(\theta)\}$ given the pay cap $\overline{r}$. It is defined as

$$p_i(\overline{r}) = \left.\frac{\mathrm{d}\,\mathbb{E}\,Q(q_1, \ldots, q_i + \varepsilon, \ldots, q_N)}{\mathrm{d}\,\varepsilon}\right|_{\varepsilon=0} \tag{11}$$

where $\mathbb{E}\,Q(q_1, \ldots, q_N)$ is the expected quality of the final product obtained by taking the expectation with respect to the distribution of input qualities $(q_1, \ldots, q_N)$ resulting from the optimal profile $\{q_1(\theta), \ldots, q_N(\theta)\}$. As before, we shall omit subscript i when dealing with some arbitrary division.

The next proposition establishes our first main result.

Proposition 2. *Let $\overline{p}$ be the expected marginal return of input quality when no cost controls are imposed. The center sets binding cost controls if and only if transfer price $t > \overline{p}$.*

As we demonstrate in the proof, if transfer price t is larger than the expected marginal return $\overline{p}$ of input quality without cost controls, the center finds it optimal to impose binding cost controls $\overline{r}$ that satisfy the following condition

$$p(\overline{r}) - \beta t = (1-\beta)\frac{C'(\overline{q}(\overline{r}))}{\theta^{+}(\overline{r})}. \tag{12}$$

In (12), $\theta^{+}(\overline{r})$ is the mapping from $\overline{r}$ into the starting point θ^{+} of the pooling interval $[\theta^{+}, \theta_H]$ found from the manager's maximization problem (4)–(7), and $\overline{q}(\overline{r})$ is the mapping from $\overline{r}$ into the highest contracted quality $\overline{q}$, respectively.

Intuitively, for $t \le \overline{p}$ the manager's incentives are either perfectly aligned with the profit maximization of the firm ($t = \overline{p}$) or misaligned to his disadvantage ($t < \overline{p}$). In either case, there is no benefit for the center in restricting the manager's discretion by imposing binding cost controls. However, with transfer price $t > \overline{p}$, the center finds it optimal to restrict the manager's discretion as now their incentives are misaligned to the center's disadvantage. Condition (12) has a natural interpretation. It requires imposing the pay cap $\overline{r}$ that equates the center's marginal revenue $p(\overline{r}) - \beta t$ from the highest contracted quality $\overline{q}$ with its marginal cost of $(1-\beta)C'(\overline{q})/\theta^{+}$ borne by the center.

3.3 Transfer Pricing

As established in Proposition 2, the center finds it optimal to restrict employee incentives when managerial objectives are misaligned with the maximization of firm profits. This misalignment can arise from various sources, one of which is negotiated transfer pricing.[9]

In many organizations the transfer price t at which a division supplies its input, and which in our framework determines the manager's expected compensation, is the outcome of bargaining between the center and the divisional manager. Managerial bargaining power can stem from several sources. One is production complementarities across divisions, as captured by a supermodular production function $Q(q_1, \ldots, q_N)$ (Milgrom and Roberts 1990). A dissatisfied manager can, for example, respond to an unfavorable transfer price by offering the employee a low-powered

9 Qualitatively similar results obtain if we replace negotiated transfer pricing with cost-based or market-based transfer prices supplemented by negotiated adjustments, as in Dutta and Reichelstein (2010).

incentive scheme, thereby reducing input quality. This reduction may be justified on the basis of low employee productivity, but due to complementarities it also lowers the marginal productivity of other divisions and the quality of the final product. Anticipating such distortions, the center may prefer to negotiate transfer prices rather than impose them unilaterally. Another source of bargaining power is organizational decentralization: as argued by Aghion and Tirole (1997), delegating decision-making authority transfers "real authority" to managers, reducing direct control by the center and increasing managerial influence over key parameters such as transfer prices.

To capture bargaining power in reduced form, we introduce a parameter $\alpha \in (0,1)$ representing the manager's share in a Nash bargaining program. This abstraction allows us to focus on the implications of bargaining outcomes without modeling all determinants of α, which may include decentralization, information asymmetries, institutional arrangements, or production complexities. In Section 4 we examine one possible mapping from α to firm size.

Assuming both parties correctly anticipate the evolution of their agency relationship following negotiations, we can express the center's expected profit per division as $\widetilde{\pi}(t)$ and the manager's expected compensation as $\widetilde{w}(t)$. Following Vaysman (1998), suppose that proposals over t are made in alternating order. Binmore et al. (1986) show that this alternating-offers bargaining game can be approximated by the Nash bargaining solution:

$$t = \underset{t' \in \mathbb{R}_+}{\operatorname{argmax}} \; (\widetilde{w}(t') - w_d)^{\alpha} \, (\widetilde{\pi}(t') - \pi_d)^{1-\alpha}, \tag{13}$$

where (w_d, π_d) is the disagreement outcome and α and $1 - \alpha$ denote the bargaining powers of the manager and the center, respectively.

The next proposition shows that negotiated transfer pricing can lead to managerial preferences for a transfer price exceeding the firm's profit-maximizing benchmark $\overline{p}$, even when cost controls are imposed.

Proposition 3. *Let the manager's expected compensation $\widetilde{w}(t)$ be strictly positive in the neighborhood of $t = \overline{p}$. If the manager has sufficiently strong bargaining power, then $t > \overline{p}$.*

Propositions 2 and 3 jointly imply that if the manager's bargaining power is sufficiently strong, the center will impose binding payroll cost controls to curtail managerial discretion. This outcome is robust to a change in the timing of the model in which both transfer price and cost controls are negotiated simultaneously. Specifically, let $(\overline{r}^*, t^*)$ solve

$$(\bar{r}^*, t^*) = \underset{(\bar{r},t')\in\mathbb{R}_+^2}{\operatorname{argmax}} \left(w(\bar{r}, t') - w_d \right)^\alpha (\tilde{\pi}(\bar{r}, t') - \pi_d)^{1-\alpha}, \tag{14}$$

where $w(\bar{r}, t')$ is the manager's expected compensation and $\tilde{\pi}(\bar{r}, t')$ is the center's expected profit per division.

If the Nash program in (13) yields $t > \overline{p}$, then it must also be that $t^* > \overline{p}$ in the new program as otherwise $t > \overline{p}$ would not be the solution to (13). To see this, we first note that the maximum in program (13) falls within the range of program (14) and, thus, the solution to (14) must lead to a higher maximal value. Next, we observe that if $t^* \leq \overline{p}$, then no binding controls are imposed. But this outcome is also feasible under the program in (13), which implies that $t^* \leq \overline{p}$ cannot be the solution to (14). Lastly, if $t^* > \overline{p}$, then in the solution to (14) the pay cap r^* is binding unless the center has no bargaining power, i.e. $1 - \alpha = 0$ but this would contradict our assumption $\alpha \in (0, 1)$. Thus, we have proved

Proposition 3′. *Let $t > \overline{p}$ if transfer price t is the only object of negotiations. Then $t > \overline{p}$ also holds if both t and cost controls $\bar{r}$ are negotiated.*

3.4 Wage Compression

The implication of Propositions 1–3 is wage compression or shallow differentiation of employee types. In the case of negotiated transfer price $t > \overline{p}$, the condition for the starting point θ^+ of the pooling interval for equal pay can be expressed as

$$\frac{1 - F(\theta^+)}{f(\theta^+)} = \theta^+ \frac{t - p(\bar{r})}{(1 - \beta)t}, \tag{15}$$

which is obtained by plugging cost controls condition (12) into initial pooling condition (8). In words, when transfer price negotiations lead to the outcome of $t > \overline{p}$, the center finds it optimal to impose payroll cost controls in the attempt to correct for misaligned managerial incentives with the firm's profit maximization. The consequence of cost controls is the manager's designing threshold employee incentives. Specifically, the manager pools most efficient employee types $\left[\theta^+, \theta_H\right]$ for the highest pay available. Thus, despite variation in employee ability we obtain compressed pay in equilibrium. From a different perspective, any performance level equal to or above the performance threshold $\overline{q}$ is rewarded with the same pay, which, in turn, can be related to the quiet quitting behavior of employees.

Our model demonstrates that the practice of threshold incentives and the resultant outcome of fixed wages can be rationalized within a standard incentive framework. Consider the example studied by Akerlof (1982), where cash posters at the Eastern Utilities Co. were paid the same wage provided they recorded at least 300 postings per hour, with no bonuses or promotions for exceeding the limit. As cash posters evidently differed in productivity, this threshold incentive scheme may appear puzzling from the perspective of the basic agency model, in which pay is typically linked to marginal productivity. This apparent puzzle has motivated explanations based on behavioral considerations such as the theory of gift exchange (Akerlof 1982) or the fair wage-effort hypothesis (Akerlof and Yellen 1990).[10] In the context of the Eastern Utilities Co. example, our model provides an alternative explanation: if the misalignment of managerial incentives leads to tight cost controls, the manager's optimal response can be to pool the entire ability space, thereby producing the observed threshold incentive scheme without invoking behavioral or psychological assumptions.

Next, given our result about the optimality of cost controls, we argue that the usage of job performance appraisal systems is optimal due to their performing the function of cost controls. In effect, appraisal systems constrain payroll expenses by imposing a bounded scale of ratings that managers can use in evaluating their employees' performance. Then, the empirical regularity of compressed ratings can be explained by our result of equilibrium pay pooling. Furthermore, as discussed in more detail below, our model also predicts that cost controls are of a lesser use in smaller firms due to better aligned managerial incentives. In connection to performance appraisal, this prediction is in line with the empirical evidence that small firms employ performance appraisal systems less frequently (Murphy and Cleveland 1995, p. 4).[11]

4 Firm-Size Wage Effects

In this section, we extend the model to explore its implications for the firm-size wage effects. These effects refer to the empirical regularity that larger firms tend to offer higher average wages but exhibit lower within-firm wage dispersion. We argue that the same institutional features such as delegated contracting, negotiated transfer pricing, and cost controls can help account for these patterns when applied to larger, more complex firms. The main assumption of our analysis is that in larger

10 At the same time, the broader incentive literature, including models of multitasking and multiple objectives (e.g., Holmstrom and Milgrom (1991)) and models with subjective evaluation (e.g., MacLeod (2003), Levin (2003)), can also account for such schemes.

11 Generally, one of the factors behind varying appraisal standards in firms is organization size. Landy and Farr (1983, pp. 104–105) describe the practice of smaller organizations to hold supervisor conferences to evaluate and, accordingly, reward the performance of each employee in turn, which is not feasible in large organizations. Murphy and Cleveland (1995, p. 355) see decentralization as a way to increase the efficiency of performance appraisal practice in organizations, because it would allow performance appraisal standards to be tailored for every functional unit. See Section 5 for further discussion.

firms, divisional managers possess greater bargaining power vis-à-vis the center, thus, exacerbating the incentive misalignment.

Assumption 1. The managerial bargaining power α increases in firm size N.

The argument that firm size affects the distribution of bargaining powers can be motivated by the theory of loss of control (Williamson 1967; Calvo and Wellisz 1978; McAfee and McMillan 1995). As firms grow more complex, the center's control over divisions weakens owing to limited monitoring resources and greater informational asymmetries. Managers possess firm-specific knowledge and privileged access to information channels about production and supply chains that the center lacks (Arrow 1974; Williamson 1985), making it costlier for the center to replace internally produced inputs with those procured externally. This informational advantage "holds up" the center and strengthens the manager's bargaining position in transfer price negotiations. This mechanism aligns with Aghion and Tirole (1997), who argue that delegation transfers "real authority" to managers, inherently reducing the center's direct control. It is also supported by Colombo and Delmastro (2004), who empirically show that greater organizational complexity and a larger number of subordinates are associated with increased delegation of authority, though only in firms without advanced network technologies that facilitate centralization.

The reasoning in the previous paragraph can be formalized analytically using Binmore et al. (1986). Consider the bargaining process between the center and the manager, who make proposals about the transfer price t in alternating order. Let Δ_C be the length of the time interval that elapses between the center's reaction to the manager's proposal, and similarly Δ_M the interval between the manager's reaction to the center's proposal.[12] In the Nash approximation (as in (13)) of the bargaining problem with asymmetric time intervals, Binmore et al. (1986, pp. 186–187) show that the manager's bargaining power is given by $\alpha = \Delta_C/(\Delta_M + \Delta_C)$. It follows that α increases with Δ_C, or in words, the longer it takes the center to react to the manager's proposal, the stronger the manager's bargaining position. If Δ_C increases with firm size N – for instance, because a larger number of divisional managers makes the center's expected response time longer – then managerial bargaining power increases with firm size, as assumed.

Assumption 1 has a straightforward implication on the negotiated transfer price t. As the firm expands, divisional managers bargain for higher transfer prices.

12 The time-consuming nature of transfer price negotiations has been recognized in the management literature, see Kaplan and Atkinson (1998, p. 461).

Lemma 1. *The negotiated transfer price t increases in firm size N.*

Proof. At the negotiated transfer price t determined by (13) the center's payoff has $\pi'(t) \leq 0$ and the manager's compensation has $w'(t) \geq 0$ (see the proof of Proposition 3 for details). Then, by the Implicit Function Theorem applied to (13) we obtain $\mathrm{d}t/\mathrm{d}\alpha \geq 0$. Thus, by Assumption 1, the transfer price t increases in firm size N. □

The effects of firm size on employee wages are thus channeled through transfer prices and cost controls. First, we show that wage dispersion, measured by the range of the pay schedule $\bar{r} - r(\theta_L)$, decreases in firm size. Second, we provide a condition under which larger firms pay higher average wages.

Proposition 4. *Let $t > \bar{p}$ and $\bar{q} > q(\theta_L)$. The range of the equilibrium pay schedule, $\bar{r} - r(\theta_L)$, decreases in firm size N.*

Intuitively, when managerial incentives are less aligned with the firm's profit maximization, the center tightens cost controls to limit payroll expenses. A lower pay cap induces managers to set coarser incentive schemes that elicit more effort from lower employee types at the expense of distortions for the most efficient types. Empirically, Proposition 4 suggests that in smaller firms the wage schedule has a higher coefficient on employee skills, a finding supported by Garen (1985), Evans and Leighton (1989), and Zenger (1994), who show that larger firms reward abilities and acquired skills at a lower rate.

In the next proposition, we show how the large-firm wage premium arises in our model. If a larger firm's equilibrium incentive scheme demands a higher expected level of quality from employees, the expected employee pay will also be higher. Let t^i, r^i, q^i, and w^i denote the negotiated transfer price, expected equilibrium employee pay, quality, and managerial compensation, respectively, when the firm's size is N^i.

Proposition 5. *Let the firm size increase from N^1 to N^2, and let $t^1 > \bar{p}$. If the expected quality increases, $q^1 < q^2$, then the expected employee pay also increases, $r^1 < r^2$.*

Proof. By Lemma 1, $N^1 < N^2$ implies $t^1 \leq t^2$. By the revealed preference argument, as the manager with transfer price t^1 chooses the contract with expected quality q^1 and payroll expense r^1 rather than that with q^2 and r^2 (which, by Proposition 4, is also feasible due to a higher pay cap), it must be that

$$t^1q^1 - r^1 \geq t^1q^2 - r^2. \tag{16}$$

Rearranging this inequality yields

$$r^2 - r^1 \geq t^1(q^2 - q^1), \tag{17}$$

which proves the proposition. □

The finding of Proposition 5 offers a different interpretation of the empirical evidence of Idson and Oi (1999) used to support the productivity hypothesis. This hypothesis draws on the degree of complementarity of physical capital and human skills increasing in firm size due to scale effects (also see Hamermesh 1980). Under this hypothesis, the large-firm wage premium arises because in larger firms employees, being better equipped, are more productive and, therefore, command higher wages (or alternatively, as argued by Troske 1999, large employers hire better employees). Our findings demonstrate the possibility of a different causal relationship in the empirically observed link of large firm size, more output per employee, and higher wages. As we argue, a large firm can produce more output per employee because of more lenient incentive schemes set up by its managers, which, however, may not be in the best interest of the firm because of lower profits. This conclusion is supported by the empirical evidence of higher profitability levels in smaller firms as documented in Fama and French (1992) and Asness et al. (2018).

5 Related Literature

A central element of our analysis is the assumption of delegated contracting, whereby contracting and communication are restricted to adjacent hierarchical layers. The rationale for this assumption connects to the broader debate on centralized versus decentralized organization of economic activity. As reviewed comprehensively by Mookherjee (2006), delegation can be optimal when it economizes on information-processing costs (Radner 1993), reduces costly communication (Melumad et al. 1992), mitigates contract complexity (Laffont and Martimort 1998), or addresses problems of incomplete commitment and renegotiation (Dessein 2002). Conversely, delegation can create incentive costs arising from misaligned objectives between managers and the organization. Regarding such misalignment, Holmstrom (1984) argues that an agent (a divisional manager in our model) will be granted more discretion or laxer cost controls when their preferences are better aligned with those of the principal, but does not explore implications for employees' wages.

The implications of organizational structure for employee pay are examined by McAfee and McMillan (1995). In a delegated agency framework, they also show a shallower differentiation of employee types and performance levels in larger firms. An important distinction between McAfee and McMillan (1995) and the present paper is that in their model firm size is given by the number of hierarchical layers.

Their result about the shallower differentiation of employee types stems from more severe production inefficiency in larger firms due to the multi-layered problem of asymmetric information. Based on their findings, they recommend a flatter organizational structure in order to limit the scope of organizational diseconomies of scale. In contrast, the present paper shows that even with a flat organizational structure, where the firm grows in the horizontal dimension with the hierarchical levels fixed, diseconomies of scale can still arise with similar implications for employee compensation.

In the economics literature, wage compression can also arise in agency models with subjective evaluation, such as MacLeod (2003) and Levin (2003). The distinctive feature of these models is that performance outcomes are not verifiable, so the agent may choose to discontinue the agency relationship should they disagree with the principal's evaluation of their performance. The optimal relational contract that minimizes the likelihood of costly conflicting situations prescribes a shallow differentiation of performance among the most efficient agent types. In our paper, this result is achieved with verifiable performance outcomes and in a one-stage setting. Finally, there is a large body of literature that addresses non-differentiation policies from various behavioral perspectives and complementarities of effort and perceived ability. For instance, Fang and Moscarini (2005) show that if morale effects are salient, owing to employees' overconfidence, no wage differentiation is a profit-maximizing policy. Relatedly, Crutzen et al. (2013) obtain non-differentiation as an optimal outcome of confidence management among multiple agents. In a three-tier agency problem with the delegated manager potentially biased toward the employee, Letina et al. (2020) show that the principal's optimal mechanism contains a cap on employee effort.

We also relate our findings to the practice and outcomes of job performance appraisal. According to surveys of business organizations (Murphy and Cleveland 1995, p. 4), most public and private companies, between 74 % and 89 % of those surveyed in the United States, with large companies more prevalent, practice a formal job performance appraisal system. The main purpose of such systems is employee salary administration. Line managers rate various aspects of their employees' performance on a pre-specified scale, and each employee is then paid in accordance with the overall rating received. The practice of performance appraisals, however, has fallen short of the expectations about their utility. The distribution of ratings typically exhibits a shallow differentiation of good from bad performance, a phenomenon labeled "compression of ratings" in the psychological literature (Landy and Farr 1983; Murphy and Cleveland 1995). Economists see this phenomenon as one of the causes of the dominance of fixed wages in company payrolls (Prendergast 1999) and raise the question of why job performance appraisal systems are inefficient in creating stronger economic incentives for employees (see Bruns

1992). As an explanation, Ockenfels et al. (2015) relate the compression of ratings to the fair wage-effort hypothesis of Akerlof and Yellen (1990) in a framework with reference-dependent preferences.

Industrial and organizational psychologists have traditionally viewed the compression of ratings phenomenon as a measurement problem. They distinguish three measurement biases: the "halo effect," a tendency to rate the same on all dimensions; "centrality bias," an overreliance on the middle of the rating scale; and "leniency bias," a tendency to give extreme ratings, which is the main focus of this paper. Psychologists have found no evidence that personal characteristics of raters or ratees have explanatory power for the systematic patterns observed in performance appraisal (Landy and Farr 1983). Instead, they have concluded that performance appraisal cannot be adequately understood outside its organizational context, and in particular, a rater's "goal-oriented rating behavior" (Murphy and Cleveland 1995), or in economic terms, the agency relationship as examined here.

In the literature dealing with the large-firm wage effects, the present paper is closest to the labor economics strand that assumes the costs of monitoring employee performance increase in firm size. Stigler (1962) explains the inverse relationship between wage dispersion and firm size by the ability of the center of a small company to better judge the quality of employee performance. Garen (1985) theoretically and empirically shows that, despite paying on average lower wages, small firms reward their employees' abilities and acquired skills, such as experience, at a greater rate than do large firms. His theoretical argument relies on the idea that the cost of screening employees' abilities rises in firm size. Another strand of literature, represented by Shapiro and Stiglitz (1984), Bulow and Summers (1986), and Oi (1990), takes the view that higher wages in large companies are "efficiency wages" that raise employees' costs of opportunistic conduct. In our model, we derive the firm-size wage effects without relying on repeated-play arguments.

6 Conclusions

In this paper, we argue that the organizational attributes of delegated contracting and transfer pricing can create conditions for the optimality of coarse pay incentives. The main idea is that transfer pricing can create asymmetry in the intrafirm valuation of employee effort. As a result, the incentives of divisional managers become misaligned with firm profit maximization. The center's attempts to correct for misaligned incentives by restricting (over-)production via cost controls result in managers' designing coarse incentive schemes for employees with implications for wage compression and quiet quitting. We also demonstrate that transfer pricing can lead to the firm-size wage effects. We derive these compensation patterns within a standard incentive-theoretic framework and without utility interdependence. The model also generates empirical predictions that may help distinguish its mechanism from alternative explanations of wage compression and declining effort. These outcomes should be more pronounced in firms with greater intra-firm incentive misalignment, reflected in decentralized budgeting, negotiated transfer pricing, or increased managerial autonomy. Variation in internal bargaining structures, even among similarly sized firms, should correlate with differences in wage dispersion and employee motivation. While testing these predictions is beyond the scope of this paper, they provide a direction for future research.

Appendix

Proof of Proposition 1

Because of the single-crossing property, the incentive-compatibility constraint (5) can be equivalently expressed by the local constraints

$$r'(\theta) - \frac{C'(q(\theta))}{\theta} q'(\theta) = 0 \tag{18}$$

and

$$q'(\theta) \geq 0. \tag{19}$$

Replace the pay variable $r(\theta)$ from the expression for the net utility $U(\theta) = r(\theta) - C(q(\theta))/\theta$. The local constraint (18) becomes

$$U'(\theta) = \frac{C(q(\theta))}{\theta^2}. \tag{20}$$

From (20), the individual rationality constraint $U(\theta) \geq 0$ simplifies to $U(\theta_L) \geq 0$, which has to be binding in the manager's optimum.

The manager chooses $q(\theta)$ and $U(\theta)$ to maximize

$$\int_{\theta_L}^{\theta_H} \beta \left(tq(\theta) - \frac{C(q(\theta))}{\theta} - U(\theta) \right) f(\theta) \mathrm{d}\theta \tag{21}$$

subject to (19), (20), $U(\theta_L) = 0$, and $U(\theta) + C(q(\theta))/\theta \leq \bar{r}$.

Without the constraint (19), the Hamiltonian of the relaxed problem is

$$H(q, U, \mu, t, \theta) = \beta \left(tq - \frac{C(q)}{\theta} - U \right) f(\theta) + \mu \frac{C(q)}{\theta^2} + \lambda \left(\bar{r} - U - \frac{C(q)}{\theta} \right), \tag{22}$$

where $\mu(\theta)$ denotes the co-state variable, $U(\theta)$ the state variable, and $\lambda(\theta)$ the multiplier of the upper bound constraint. The necessary conditions imply that

$$\beta\left(t - \frac{C'(q(\theta))}{\theta}\right) f(\theta) + \mu(\theta)\frac{C'(q(\theta))}{\theta^2} - \lambda(\theta)\frac{C'(q(\theta))}{\theta} = 0, \tag{23}$$

$$-\beta f(\theta) - \lambda(\theta) = -\mu'(\theta), \tag{24}$$

$$\lambda(\theta)\left(\bar{r} - U - \frac{C(q)}{\theta}\right) = 0, \lambda(\theta) \geq 0, \tag{25}$$

$$\mu(\theta_H) = 0. \tag{26}$$

In what follows, we consider the case where the upper bound constraint is binding (otherwise we have a standard solution with $\lambda(\theta) = 0$ and $\mu(\theta) = F(\theta) - 1$).

Let θ^+ denote the lowest type for whom the upper bound constraint is binding. Denote the allocation for θ in $\left[\theta^+, \theta_H\right]$ by $(\bar{q}, \bar{U})$. We have $\theta^+ < \theta_H$ as otherwise the optimal allocations $q(\theta)$ for $\theta < \theta_H$ are the same to the ones from the unconstrained problem, not feasible with the binding upper bound constraint. Combining (23) and (24) for $\theta \geq \theta^+$ and further simplifying render

$$\beta t f(\theta) + \mu(\theta)\frac{C'(\bar{q})}{\theta^2} - \mu'(\theta)\frac{C'(\bar{q})}{\theta} = 0. \tag{27}$$

Solving for $\mu(\theta)$ with $\mu(\theta_H) = 0$ we obtain

$$\mu(\theta) = -\beta t\frac{1 - F(\theta)}{C'(\bar{q})/\theta}. \tag{28}$$

The condition (24) for $\theta \leq \theta^+$ can be rewritten as

$$\mu(\theta) = \beta F(\theta) + \mu(\theta_L). \tag{29}$$

From (28) and (29), we obtain

$$\mu(\theta_L) = -\beta\left(F(\theta^+) + t\frac{1 - F(\theta^+)}{C'(\bar{q})/\theta^+}\right). \tag{30}$$

Rewrite (23) for $\theta \leq \theta^+$ as

$$\beta\left(t - \frac{C'(q(\theta))}{\theta}\right) f(\theta) + \left(\beta F(\theta) + \mu(\theta_L)\right)\frac{C'(q(\theta))}{\theta^2} = 0. \tag{31}$$

At $\theta = \theta^+$ we get, after rearranging,

$$\frac{1 - F(\theta^+)}{f(\theta^+)} = \theta^+ - \frac{C'(\bar{q})}{t}. \tag{32}$$

In the relaxed problem, the optimal allocation $q(\theta)$ is given by (30)–(32), where $\overline{q}$ is obtained from the binding upper bound constraint that can be expressed from (20) as

$$\overline{U} = \bar{r} - \frac{C(\overline{q})}{\theta^+} = \int_{\theta_L}^{\theta^+} \frac{C(q(\theta))}{\theta^2} \mathrm{d}\theta. \tag{33}$$

The optimal net utility $U(\theta)$ or, correspondingly, the optimal pay level $r(\theta)$ is found from (20).

See Besanko et al. (1987) for the sufficient conditions and that the omitted constraint (19) holds.

Proof of Proposition 2

Given the optimal employee pay scheme $\{(r, q)\}$ and symmetry across the divisions, the center's expected profit function of pay cap $\bar{r}$ and transfer price t can be expressed as

$$\pi(\bar{r}, t) = \mathbb{E}\, Q(q_1, \ldots, q_N) - N\big(w(\bar{r}, t) - \mathbb{E}\, r\big). \tag{34}$$

In (34), $\mathbb{E}\, Q(q_1, \ldots, q_N)$ is the expected quality of the final product obtained by taking the expectation with respect to the distribution of input qualities $(q_1, \ldots, q_N)$ resulting from the optimal employee pay scheme, $w(\bar{r}, t)$ is the expected managerial compensation from (4), and $\mathbb{E}\, r$ is the expected employee pay. First, we show that the derivative of the expected profit function $\pi(\bar{r}, t)$ with respect to $\bar{r}$ is given by

$$\frac{\mathrm{d}\pi(\bar{r}, t)}{\mathrm{d}\bar{r}} = N(1 - F(\theta^+(\bar{r})))\left[(p(\bar{r}) - \beta t)\frac{\theta^+(\bar{r})}{C'(\overline{q}(\bar{r}))} - (1 - \beta)\right], \tag{35}$$

where $p(\bar{r})$ is the expected marginal return of input quality defined in (11), $\theta^+(\bar{r})$ is the mapping from $\bar{r}$ into the starting point θ^+ of the pooling interval $[\theta^+, \theta_H]$, and $\overline{q}(\bar{r})$ is the mapping from $\bar{r}$ into the highest contracted quality $\overline{q}$, respectively.

Applying the dynamic envelope theorem of LaFrance and Barney (1991, Theorem 1) to the managerial compensation $w(\bar{r}, t)$, we have

$$\frac{\mathrm{d}w(\bar{r}, t)}{\mathrm{d}\bar{r}} = \int_{\theta_L}^{\theta_H} \lambda(\theta) \mathrm{d}\theta, \tag{36}$$

where $\lambda(\theta)$ is the multiplier on the upper bound constraint (7), $r(\theta) \le \bar{r}$. Because $\lambda(\theta) = 0$ for $\theta \le \theta^+$, we obtain from (24), (26), and (28)

$$\frac{\mathrm{d}w(\bar{r}, t)}{\mathrm{d}\bar{r}} = \beta(1 - F(\theta^+(\bar{r})))\left(t\frac{\theta^+(\bar{r})}{C'(\overline{q}(\bar{r}))} - 1\right). \tag{37}$$

In words, pay cap $\bar{r}$ matters for the expected managerial compensation only through its direct effect on the highest contracted levels of pay and quality. Observing that

$\frac{\mathrm{d}\,\mathbb{E}r}{\mathrm{d}\,\overline{r}} = 1 - F(\theta^+)$ and recalling that $w(\overline{r}, t) = \beta(t\mathbb{E}\, q - \mathbb{E}\, r)$, we find that the change in expected quality is $\frac{\mathrm{d}\,\mathbb{E}\,q}{\mathrm{d}\,\overline{r}} = (1 - F(\theta^+))\frac{\theta^+}{C'(\overline{q})}$. Then, using the derivatives obtained, the symmetry of production function $Q(\cdot)$, and the symmetry across the divisions, we obtain the derivative of expected profit function $\pi(\overline{r}, t)$ with respect to $\overline{r}$ as given in (35).

Next, denote by r^m the smallest lower bound on non-binding pay caps so that setting $\overline{r} = r^m$ corresponds to no payroll controls imposed. Let $t \leq \overline{p}$. With a binding pay cap $\overline{r} < r^m$ and, thus, $\theta^+ < \theta_H$, the derivative in (35) is strictly positive because $\theta^+/C'(\overline{q}) > 1/t \geq 1/\overline{p}$ which follows from (8) and $t \leq \overline{p}$. Hence, for $t \leq \overline{p}$ it must be that $\theta^+ = \theta_H$ which implies no payroll controls imposed.

Let $t > \overline{p}$. Now suppose that the center does not impose payroll controls, i.e. sets $\overline{r} = r^m$, which implies $\theta^+ = \theta_H$. The expression in the square brackets of the derivative $\mathrm{d}\pi/\mathrm{d}\overline{r}$ in (35) is strictly negative because from (8) we have $\theta_H/C'(\overline{q}) = 1/t$ which is less than $1/\overline{p}$ due to $t > \overline{p}$. Then, by continuity, it must be that the derivative $\mathrm{d}\pi/\mathrm{d}\overline{r}$ is negative for at least some values of $\overline{r}$ to the left from r^m which implies that r^m cannot be a maximizer of the profit function π for $t > \overline{p}$. Hence, for $t > \overline{p}$ the center imposes the binding pay cap found from the first-order condition

$$p(\overline{r}) - \beta t = (1 - \beta)\frac{C'(\overline{q}(\overline{r}))}{\theta^+(\overline{r})}.$$

Proof of Proposition 3

To prove the proposition, it is sufficient to establish that the manager's most preferred transfer price has $t > \overline{p}$. If it is the case, then with $1 - \alpha$ approaching 1 the Nash program in (13) yields $t > \overline{p}$.

The manager's expected compensation is given by

$$\widetilde{w}(t) = \beta \int_{\theta_L}^{\theta_H} \left(tq_t(\theta) - r_t(\theta)\right) f(\theta)\mathrm{d}\theta,$$

where q_t and r_t are the manager's optimal quality and pay schedules designed for the employee at transfer price t as determined by (8)–(10).

First, we consider the case of $t \leq \overline{p}$, which is the case when no binding cost controls are imposed (Proposition 2). Let $t' < t \leq \overline{p}$, then we have

$$\widetilde{w}(t) = \int (tq_t(\theta) - r_t(\theta))f(\theta) \geq \int (tq_{t'}(\theta) - r_{t'}(\theta))f(\theta) > \int \big(t' q_{t'}(\theta) \\ - r_{t'}(\theta)\big) f(\theta) = \widetilde{w}(t'),$$

where the first inequality follows from the revealed preference argument as $\left\{\left(q_{t'}, r_{t'}\right)\right\}$ are feasible under t as no cost controls are imposed. This proves that $\widetilde{w}(t)$ increases in $t < \overline{p}$.

The revealed preference argument is not applicable when $t > \overline{p}$ because of binding cost controls. For $t \geq \overline{p}$ the manager's expected compensation can be expressed as

$$\widetilde{w}(t) = \beta\left[\int_{\theta_L}^{\theta^+} \big(tq(\theta) - r(\theta)\big) f(\theta)\mathrm{d}\theta + (1 - F(\theta^+))(t\overline{q} - \overline{r})\right].$$

Its derivative $\widetilde{w}'(t)$ is equal to

$$\beta\left[\int_{\theta_L}^{\theta^+} \left(q(\theta) + t\frac{\mathrm{d}q(\theta)}{\mathrm{d}t} - \frac{\mathrm{d}r(\theta)}{\mathrm{d}t}\right) f(\theta)\mathrm{d}\theta + (1 - F(\theta^+))\left(\overline{q} + t\frac{\mathrm{d}\overline{q}}{\mathrm{d}t} - \frac{\mathrm{d}\overline{r}}{\mathrm{d}t}\right)\right]. \quad (38)$$

The internal derivative $\mathrm{d}q(\theta)/\mathrm{d}t$, found from (9),

$$\frac{\mathrm{d}q(\theta)}{\mathrm{d}t} = -\frac{f(\theta) + \frac{C'(q(\theta))}{\theta^2}\left(\frac{(1-F(\theta^+))(1-\beta)\,p(\overline{r})}{(p(\overline{r})-\beta t)^2} - \frac{(t-p(\overline{r}))f(\theta^+)\mathrm{d}\theta^+/\mathrm{d}t}{p(\overline{r})-\beta t}\right)}{-\frac{C''(q(\theta))}{\theta}f(\theta) - \frac{C''(q(\theta))}{\theta^2}\left[\frac{(1-F(\theta^+))(t-p(\overline{r}))}{p(\overline{r})-\beta t} + 1 - F(\theta)\right]} \quad (39)$$

is positive because of the functional assumptions of the model and $\mathrm{d}\theta^+/\mathrm{d}t < 0$ which follows from (15). In deriving (39), we also used (15) and that $p(\overline{r}) - \beta t$ is positive, see (12). Now consider a transfer price t close to $p(\overline{r})$ so that the second term in (38) is negligible (with $t \to \overline{p}$, we have $\theta^+ \to \theta_H$ as the center's and the manager's incentives become more closely aligned). If $\widetilde{w}'(t) < 0$, it means that the expected change in employee pay $\mathrm{d}r(\theta)/\mathrm{d}t$ exceeds the total expected output created by the employee. But this implies a negative expected compensation for the manager, which contradicts the assumption of positive compensation in the neighborhood of $\overline{p}$. Thus, we have that $\widetilde{w}'(t) > 0$ for at least some $t \in [1, \tau)$, $\tau > 1$, which implies that the manager's most preferred transfer price has $t > \overline{p}$.

Proof of Proposition 4

A change in transfer price t has a direct and indirect effect on the highest contracted quality level $\overline{q}$. The direct effect, denoted by $\partial\overline{q}/\partial t$, comes from the manager's optimization problem, whereas the indirect effect comes from the center's setting pay cap $\overline{r}$. The total effect can be formally expressed as the total derivative

$$\frac{\mathrm{d}\overline{q}(\overline{r}, t)}{\mathrm{d}t} = \frac{\partial\overline{q}}{\partial t} + \frac{\partial\overline{q}}{\partial\overline{r}}\frac{\mathrm{d}\overline{r}}{\mathrm{d}t} \quad (40)$$

The sign of the total derivative $\mathrm{d}\overline{q}/\mathrm{d}t$ is found by totally differentiating the first-order condition (12):

$$-\beta\mathrm{d}t = (1 - \beta)\left(\frac{C''(\overline{q})}{\theta^+}\mathrm{d}\overline{q} - \frac{C''(\overline{q})}{(\theta^+)^2}\mathrm{d}\theta^+\right) \quad (41)$$

or, after reorganization,

$$-\frac{\beta}{1-\beta} = \frac{C''(\overline{q})}{\theta^+}\frac{\mathrm{d}\overline{q}}{\mathrm{d}t} - \frac{C'(\overline{q})}{(\theta^+)^2}\frac{\mathrm{d}\theta^+}{\mathrm{d}t}. \tag{42}$$

Since $\mathrm{d}\theta^+/\mathrm{d}t < 0$ from (15), the last expression yields that total derivative $\mathrm{d}\overline{q}/\mathrm{d}t < 0$. The internal differentiation $\mathrm{d}q(\theta)/\mathrm{d}t$ in (39) from the proof of Proposition 3 renders that the direct effect is positive, $\partial\overline{q}/\partial t > 0$. As the manager contracts more quality at a higher pay cap, $\mathrm{d}\overline{q}/\mathrm{d}\overline{r} = \theta^+/C'(\overline{q}) > 0$ from (10), it follows from (40) that $\mathrm{d}\overline{r}/\mathrm{d}t < 0$.

As for the lowest contracted quality, there is only a direct effect, $\partial q(\theta_L)/\partial t$, provided that $q(\theta_L) < \overline{q}$. This effect is again found from the internal differentiation in (39) for θ_L, from which we obtain $\partial q(\theta_L)/\partial t > 0$. Given the binding individual-rationality constraint, the reward paid to the lowest type, $r(\theta_L)$, has to increase in t.

As by Lemma 1 the transfer price increases in firm size N, we obtain that the wage dispersion $\overline{r} - r(\theta_L)$ decreases in N.

References

Aghion, P., and J. Tirole. 1997. “Formal and Real Authority in Organizations.” *Journal of Political Economy* 105: 1–29.

Akerlof, G. A. 1982. “Labor Contracts as Partial Gift Exchange.” *Quarterly Journal of Economics* 97: 543–69.

Akerlof, G. A., and R. E. Kranton. 2000. “Economics and Identity.” *Quarterly Journal of Economics* 115: 715–53.

Akerlof, G. A., and J. L. Yellen. 1990. “The Fair Wage-Effort Hypothesis and Unemployment.” *Quarterly Journal of Economics* 105: 255–83.

Amershi, A. H., and P. Cheng. 1990. “Intrafirm Resource Allocation: The Economics of Transfer Pricing and Cost Allocations in Accounting.” *Contemporary Accounting Research* 7: 61–99.

Arrow, K. J. 1974. *The Limits of Organization*. New York: Norton.

Asness, C., A. Frazzini, R. Israel, T. J. Moskowitz, and L. H. Pedersen. 2018. “Size Matters, if You Control Your Junk.” *Journal of Financial Economics* 129: 479–509.

Baldenius, T., S. Reichelstein, and S. A. Sahay. 1999. “Negotiated Versus Cost-Based Transfer Pricing.” *Review of Accounting Studies* 4: 67–91. https://doi.org/10.1023/a:1009638001487.

Besanko, D., S. Donnenfeld, and L. J. White. 1987. “Monopoly and Quality Distortion: Effects and Remedies.” *Quarterly Journal of Economics* 102: 743–67.

Bhattacharyya, S., and F. Lafontaine. 1995. “Double-Sided Moral Hazard and the Nature of Share Contracts.” *Rand Journal of Economics* 26: 761–81.

Binmore, K., A. Rubinstein, and A. Wolinsky. 1986. “The Nash Bargaining Solution in Economic Modelling.” *Rand Journal of Economics* 17: 176–88.

Bishop, J. 1987. “The Recognition and Reward of Employee Performance.” *Journal of Labor Economics* 4: S36–S56.

Brown, C., and J. Medoff. 1989. “The Employer Size-Wage Effect.” *Journal of Political Economy* 97: 1027–59.

Bruns, Jr., W. J., eds. 1992. *Performance Measurement, Evaluation and Incentives*. Boston, MA: Harvard Business School Press.

Bulow, J. I., and L. H. Summers. 1986. “A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment.” *Journal of Labor Economics* 4: 376–414.

Calvo, G. A., and S. Wellisz. 1978. “Supervision, Loss of Control, and the Optimum Size of the Firm.” *Journal of Political Economy* 86: 943–52.

Coase, R. 1937. “The Nature of the Firm.” *Economica* 4: 386–405.

Cobb, J. A., and K.-H. Lin. 2017. “Growing Apart: The Changing Firm-Size Wage Premium and its Inequality Consequences.” *Organization Science* 28: 429–46.

Colombo, M. G., and M. Delmastro. 2004. “Delegation of Authority in Business Organizations: An Empirical Test.” *Journal of Industrial Economics* 52: 53–80.

Crutzen, B. S. Y., O. H. Swank, and B. Visser. 2013. “Confidence Management: On Interpersonal Comparisons in Teams.” *Journal of Economics and Management Strategy* 22: 744–67.

Dessein, W. 2002. “Authority and Communication in Organizations.” *Review of Economic Studies* 69: 811–38.

Dutta, S., and S. Reichelstein. 2010. “Decentralized Capacity Management and Internal Pricing.” *Review of Accounting Studies* 15: 442–78.

Edlin, A., and S. Reichelstein. 1995. “Specific Investment Under Negotiated Transfer Pricing: An Efficiency Result.” *Accounting Review* 70: 275–91.

Evans, D. S., and L. S. Leighton. 1989. “Why Do Smaller Firms Pay Less?” *Journal of Human Resources* 24: 299–318.

Fama, E. F., and K. R. French. 1992. “The Cross-Section of Expected Stock Returns.” *Journal of Finance* 47: 427–65.

Fang, H., and G. Moscarini. 2005. “Morale Hazard.” *Journal of Monetary Economics* 52: 749–77.

Fox, J. T. 2009. “Firm-Size Wage Gaps, Job Responsibility, and Hierarchical Matching.” *Journal of Labor Economics* 27: 83–126.

Frazis, H., and M. A. Loewenstein. 2006. *Wage Compression and the Division of Returns to Productivity Growth: Evidence from EOPP*. U.S. Bureau of Labor Statistics. Working Paper No 398.

Gallup. 2022. “State of the Global Workplace: 2022 Report.” https://www.gallup.com (accessed May 17, 2023).

Garen, J. E. 1985. “Worker Heterogeneity, Job Screening, and Firm Size.” *Journal of Political Economy* 93: 715–39.

Hamermesh, D. S. 1980. “Commentary.” In *The Economics of Firm Size, Market Structure, and Social Performance*, edited by J. J. Siegfried, 383–8. Washington, D.C.: Federal Trade Commission.

Hirshleifer, J. 1956. “On the Economics of Transfer Pricing.” *Journal of Business* 29: 172–84.

Holmstrom, B. 1984. “On the Theory of Delegation.” In *Bayesian Models in Economic Theory*, edited by M. Boyer, and R. E. Kihlstrom, 113–41. New York: North-Holland.

Holmstrom, B., and P. Milgrom. 1991. “Multitask Principal Agent Analyses: Incentive Contracts, Asset Ownership and Job Design.” *Journal of Law, Economics, and Organization* 7: 24–52.

Holmstrom, B., and J. Tirole. 1991. “Transfer Pricing and Organizational Form.” *Journal of Law, Economics, and Organization* 7: 201–28.

Idson, T. L., and W. Y. Oi. 1999. “Workers Are More Productive in Large Firms.” *American Economic Review* 89: 104–8.

Kaplan, R., and A. Atkinson. 1998. *Advanced Management Accounting*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.

Laffont, J. J., and D. Martimort. 1998. “Collusion and Delegation.” *Rand Journal of Economics* 29: 280–305.

Laffont, J. J., and J. Tirole. 1986. "Using Cost Observation to Regulate Firms." *Journal of Political Economy* 95: 619–41.
LaFrance, J. T., and L. D. Barney. 1991. "The Envelope Theorem in Dynamic Optimization." *Journal of Economic Dynamics and Control* 15: 355–85.
Landy, F. J., and J. L. Farr. 1983. *The Measurement of Work Performance: Methods, Theory, and Applications*. New York: Academic Press.
Letina, I., S. Liu, and N. Netzer. 2020. "Delegating Performance Evaluation." *Theoretical Economics* 15: 477–509.
Levin, J. 2003. "Relational Incentive Contracts." *American Economic Review* 93: 835–57.
MacLeod, B. W. 2003. "Optimal Contracting with Subjective Evaluation." *American Economic Review* 93: 216–40.
McAfee, R. P., and J. McMillan. 1995. "Organizational Diseconomies of Scale." *Journal of Economics and Management Strategy* 4: 399–426.
Melumad, N., D. Mookherjee, and S. Reichelstein. 1992. "A Theory of Responsibility Centers." *Journal of Accounting and Economics* 15: 445–84.
Merchant, K. A., and W. A. Van der Stede. 2007. *Management Control Systems: Performance Measurement, Evaluation, and Incentives*. London: Prentice Hall.
Milgrom, P., and J. Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review* 80: 511–28.
Mookherjee, D. 2006. "Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective." *Journal of Economic Literature* 44: 367–90.
Moore, H. L. 1911. *Laws of Wages*. New York (Reprinted 1967, Augustus M. Kelley, New York): Macmillan.
Murphy, K. R., and J. Cleveland. 1995. *Understanding Performance Appraisal: Social, Organizational, and Goal-based Perspectives*. SAGE.
Ockenfels, A., D. Sliwka, and P. Werner. 2015. "Bonus Payments and Reference Point Violations." *Management Science* 61: 1496–513.
O'Connor, S. 2022. "The Term 'Quiet Quitting' is Worse than Nonsense." *Financial Times*. 13 September [Online] https://www.ft.com (accessed May 17, 2023).
Oi, W. Y. 1990. "Employment Relations in Dual Labor Markets ("it's Nice Work if You Can Get It")." *Journal of Labor Economics* 8: 124–49.
Oi, W. Y., and T. L. Idson. 1999. "Firm Size and Wages." In *Handbook of Labor Economics*, Vol. 3, edited by O. Ashenfelter, and D. Card, 2165–214. Elsevier.
Oyer, P., and S. Schaefer. 2005. "Why Do Some Firms Give Stock Options to all Employees? An Empirical Examination of Alternative Theories." *Journal of Financial Economics* 76: 99–133.
Prendergast, C. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37: 7–63.
Radner, R. 1993. "The Organization of Decentralized Information Processing." *Econometrica* 61: 1109–46.
Shapiro, C., and J. E. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74: 433–44.
Stigler, G. J. 1962. "Information in the Labor Market." *Journal of Political Economy* 70: 94–105.
Stoughton, N. M., and E. Talmor. 1999. "Managerial Bargaining Power in the Determination of Compensation Contracts and Corporate Investment." *International Economic Review* 40: 69–93.
Troske, K. R. 1999. "Evidence on the Employer Size-Wage Premium from Worker-Establishment Matched Data." *Review of Economics and Statistics* 81: 15–26.
Vaysman, I. 1998. "A Model of Negotiated Transfer Pricing." *Journal of Accounting and Economics* 25: 349–84.

Williamson, O. E. 1967. "Hierarchical Control and Optimum Firm Size." *Journal of Political Economy* 75: 123–38.
Williamson, O. E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.
Zenger, T. R. 1994. "Explaining Organizational Diseconomies of Scale in R & D: Agency Problems and the Allocation of Engineering Talent, Ideas, and Effort by Firm Size." *Management Science* 40: 708–29.