When Should a Firm Employ its Workers? A Relational Contracting Approach

Anja Schöttner* and Harvey Upton†‡

13th December 2023

Abstract

A principal hiring an agent chooses between employment and self-employment. Under employment, the principal can control the manner in which the agent performs the assignment (the 'work design'), whereas this is legally forbidden under self-employment. Due to incomplete contracts, the relationship is governed by relational agreements, which are influenced by the allocation of control. We show that employment typically results in an over-demanding work design, compared to an under-demanding work design under self-employment, and that employment relationships are more rigid and better suited to favorable production environments. We further examine the impact of taxation, minimum wages, and formal performance pay.

Keywords: employment; self-employment; theory of the firm; relational contracts.

JEL classification: D86; M55; L23; L24.

*Humboldt-Universität zu Berlin; anja.schoettner@hu-berlin.de

†Humboldt-Universität zu Berlin; harvey.upton@hu-berlin.de

‡We gratefully acknowledge comments and suggestions by Jörg Budde, Hoa Ho, Michael Raith, participants of the GEABA 2022 Symposium, the TRR 266 Annual Conference 2022, the Colloquium on Personnel Economics (COPE) 2023, the Annual Meeting of the Committee for Organizational Economics 2023 and seminar audiences at Goethe-Universität Frankfurt am Main and Humboldt-Universität zu Berlin. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 403041268 – TRR 266 Accounting for Transparency.

1 Introduction

Many firms, particularly those that offer delivery and transport services on digital labor platforms, face a choice between hiring their workers as employees, or instead contracting with solo self-employed workers.[1] The employment status of platform workers has long been an issue of substantial controversy, in particular since workers classified as self-employed are typically not eligible for additional benefits that often accompany employment, such as minimum wages, sick pay and parental leave or pension contributions (Boeri et al., 2020). While this is often cited as a key motivator for outsourcing work to contractors, employment status also significantly influences the organization of the work process, especially with respect to the degree of autonomy accorded to the worker (Eurofound, 2018). Self-employed workers often have the discretion to choose their working hours, accept or decline tasks and control the manner in which work is performed, in a way that employed workers cannot.

From a legal perspective, this difference in the allocation of control is a defining feature of a worker's employment status. Black's Law Dictionary (page 471, 5th ed. 1979) defines an "employee" as "a person in the service of another under any contract of hire, express or implied, oral or written, where the employer has the power or right to control and direct the employee in the material details of how the work is to be performed." When a court decides whether a worker is an employee or an independent contractor, it typically evaluates the "totality of the circumstances surrounding a worker's employment, with a focus on who has the right—the employer or the employee—to control the work process" (Muhl, 2002). Recently, the European Commission published a proposal for a directive on the regulation of digital labor platforms according to which an employment relationship exists, for example, if the platform restricts the freedom to organize one's work, in particular the discretion to choose one's working hours or periods of absence, or to accept or refuse tasks.[2]

[1]The solo self-employed are self-employed workers who do not have dependent workers on their payroll. See Boeri et al. (2020) for a recent overview on the increasing relevance of solo self-employment in OECD countries over the last two decades.

[2]See https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6605. Similarly, in the US, the IRS state that anyone who performs services is an employee "if [the employer] can

In this paper, we study how contracting between a firm and a worker is affected by differences in the extent to which parties are permitted to control the production process, and the ensuing implications for the firm's optimal choice between employment and self-employment. For that purpose, we study a principal-agent framework with multiple periods. The principal (firm, she) needs to contract with an agent (worker, he) to perform an assignment in each period. The non-verifiable outcome can be either high or low, depending on the non-verifiable manner in which the assignment is performed, which we refer to as the "work design". This captures the aspects of the work process where the allocation of control is dictated by the organizational form due to the legal system, in line with the foregoing discussion. A more demanding work design increases both the probability of the high outcome and the agent's costs, creating a conflict of interest between the two parties. The agent's costs also depend on the state of the world, which is observable by both parties but non-verifiable and captures the favorability of the production environment, in terms of potential gains from trade. The state may vary between periods and is realized after the contracting stage in each period.

As discussed in the foregoing, the sole difference between the two organizational forms lies in the allocation of control. Under employment, the principal is permitted to control the work design and therefore directly chooses it herself. In contrast, this is legally forbidden under self-employment, so that the agent has full discretion over the manner in which the work is performed. The principal can then only indirectly affect the agent's choice via the contract design.

The work design, though non-verifiable, is assumed to be observable by both the principal and the agent. It follows that the labor relationship can benefit from the use of relational contracts, which differ between organizational forms due to the disparate allocations of control. Under employment, the principal stipulates a work design associated with each state of the world and promises that, following the realization of the state, she will adhere to this schedule. Since the principal can always deviate from this agreement by selecting the most

control what will be done and how it will be done"; see https://www.irs.gov/businesses/small-businesses-self-employed/employee-common-law-employee.

demanding work design—which maximizes the probability of a high outcome but also the agent's costs—the schedule must be designed in such a way that her promise is self-enforcing. In a given period, the agent must complete the assignment under the work design stipulated by the principal and incur the associated costs, as the agent's loss from prematurely terminating the relationship are prohibitively high. The agent can, however, decline to work for the principal in the subsequent period.

Under self-employment, since the principal cannot directly control the production process, she instead uses monetary incentives to induce the appropriate work design. Specifically, she outlines a work design schedule, which she asks the agent to implement, and associated discretionary bonuses, which she promises to pay conditional on the agent's obedience. It follows that in the case of self-employment, both parties have the opportunity to renege on the relational agreement. The agent can ignore the principal's request by implementing the least demanding work design and thus minimizing his costs, while foregoing the discretionary bonus; the principal, in turn, can refuse to pay the bonus even after observing that the agent has obeyed her request. As in the case of employment, the potential for opportunistic behavior implies that the parties' agreement must be self-enforcing.[3]

When parties place sufficiently high value on the future, the problems arising from contractual incompleteness can be resolved entirely by relational agreements under both organizational forms, so that the first-best work design can be implemented in each state. However, for lower discount factors, this outcome is unattainable and the work design schedules must be adjusted in order to remain feasible. In particular, we show that employment and self-employment lead to distinct second-best work design schedules. Since the principal's optimal deviation under employment is to implement the most demanding work design, maximizing expected output, the principal is most tempted to cheat on the agreement in unfavorable

[3]For example, consider a food delivery platform. An employed rider typically must accept each assigned delivery, while a self-employed rider may decide to work only in nice weather and accept only less exhausting or time-consuming rides. Under a relational agreement, the self-employed worker could promise to accept each delivery that the app's algorithm assigns to him. Under employment, the platform could promise to make only reasonable requests and program its algorithm accordingly.

states of the world, when the first-best work design is relatively undemanding. In these states, in order to curtail the principal's deviation incentives, the implemented work design must be adjusted such that it is over-demanding relative to the first-best.

In contrast, the principal's optimal deviation under self-employment is to renege on payment of the discretionary bonus. Since incentive compatibility implies that the agent must be paid a higher bonus to induce a more demanding work design, the principal is most tempted to cheat in favorable states of the world. In these states, to guarantee a self-enforcing agreement the implemented work design must be less demanding relative to the first-best—thus lowering the associated discretionary bonus—so that the principal's temptation to withhold payment is reduced. It then follows that inefficiencies arise under self-employment from an under-demanding work design in favorable states; the opposite is true under employment, where they arise from an over-demanding work design in unfavorable states. These distinct outcomes yield several predictions regarding a firm's optimal selection of labor relationship.

First, self-employment is likely to be optimal in unfavorable production environments, where the agent's costs are typically high or where prospective output is small; this follows since, by the foregoing arguments, employment is associated with large inefficiencies in such cases. Second, we predict that the work design is typically more rigid under employment, in the sense that it responds less to changes in the production environment. In the absence of contractual incompleteness, parties would implement the first-best work design in each period, which varies with the state of the world. However, when the relationship is governed by relational agreements, under employment the principal may implement an identical work design for many different realizations of the state; in fact, for a sufficiently low discount factor, this is the case in all states and the relationship is perfectly rigid. In contrast, under self-employment each state is always associated with a distinct work design; accordingly, our findings are consistent with empirical observations that self-employed relationships tend to be more flexible (Hurst and Pugsley, 2011; Boeri et al., 2020).

Third, our model yields predictions regarding differences in the size and structure of worker remuneration between the two organizational forms. Under employment, the principal directly controls the production process and has no need to create incentives, so that she pays a fixed wage in each period that compensates the agent's expected costs. In contrast, the discretionary bonuses paid under self-employment depend on both the implemented work design and the state of the world, and therefore vary from period-to-period. In addition, since the agent's costs are typically lower under self-employment—due to the over-demanding vs. under-demanding work designs implemented under employment and self-employment respectively—our findings suggest that self-employed workers should receive relatively lower compensation on average, in line with empirical evidence (Boeri et al., 2020).

The remainder of the paper considers three applications of the model that yield various further predictions. First, we study the implications of income taxation on the agent. We find that taxes distort the contract design problem to a greater extent under self-employment, since incentives must be created through monetary transfers in this case. Specifically, taxation increases the size of the discretionary bonus that must be paid to induce a particular work design, which, in turn, increases the principal's incentive to renege on the agreement by withholding payment. In contrast, under employment, the principal selects the work design herself; since this involves no monetary transaction, her incentive to deviate is unaffected.

However, there is an additional effect of taxation. Since taxes increase the principal's costs of compensating the agent, the introduction of a tax, in effect, increases the agent's costs. This reduces the favorability of the production environment and makes the (tax-conditional) efficient work design less demanding, which, in line with the paper's main findings, favors self-employment. We show that this effect often dominates, so that an increase in taxes tends to raise the relative profitability of self-employment; in particular, we show that for sufficiently high levels of income taxation, an organizational design of self-employment must become strictly optimal.

Second, we investigate the introduction of a minimum wage under employment, showing that this changes the implemented work design and reduces the efficiency of the relationship, so that self-employment becomes relatively more attractive.[4] Specifically, we show that the principal responds to a binding minimum wage constraint by implementing an even more demanding work design, exacerbating the problem that occurs due to contractual incompleteness. At first, the increase in the agent's wage is entirely offset by higher costs, so that his utility remains unchanged and the labor relationship becomes strictly less (Pareto) efficient. However, for sufficiently large values of the minimum wage, when the most demanding work design is already in place, the agent begins to extract a strictly positive rent.

It follows that implementation of a high minimum wage can successfully achieve a transfer of wealth from the principal to the agent, albeit at the cost of a reduction in efficiency. Our model predicts that this efficiency loss will be smallest in favorable production environments, where the optimal work design is already quite demanding and hence cannot be increased much further. Moreover, in such cases, firms find it difficult to avoid the minimum wage by switching to self-employment, as this is not well-suited to favorable production environments. Similarly, the efficiency loss is small in situations where a low discount factor implies very inefficient contracts with an over-demanding work design even in the absence of a wage floor.

Third, we extend our analysis to an environment where the principal has the ability to create formal incentives by conditioning payments on a verifiable assignment outcome. We show that even with outcome-contingent pay, relational agreements are often unable to achieve the first-best, so that the key trade-offs underlying the principal's choice of organizational form persist. Specifically, in such cases, the implemented work designs under employment and self-employment remain over- and under-demanding relative to the first-best, respectively, as in our main analysis. However, we also show that formal performance pay complements relational agreements and allows the principal to implement more efficient work designs, relative to the case where the task outcome is non-verifiable. Under self-employment, formal incentives allow the principal to reduce the size of discretionary bonus payments. Under employment, despite the absence of a need to create incentives, rewarding

[4] The finding that a minimum wage reduces the attractiveness of employment relative to self-employment is in line with recent empirical evidence (Glasner, 2023; Ganserer et al., 2022).

the agent with a proportion of output following a successful assignment reduces the principal's benefit from deviating by adjusting the work design, as she no longer receives the full output. In both cases, the principal's incentive to renege on the relational agreement is reduced, relaxing feasibility constraints and increasing efficiency. We further argue that in the presence of formal performance pay, a higher outside option for the agent (e.g., due to increased competitiveness of the labor market) will typically favor self-employment.

Altogether, the paper contributes to the existing literature in several different ways. First, we provide a novel explanation for the factors that drive the decision between employment and self-employment, based on the allocation of control imposed by the legal system. Second, our model yields various empirically testable predictions that are broadly in line with existing evidence and can be used to inform future research. Third, we provide a tractable framework for analyzing the distinct effects of various policies on different types of labor relationship. In the paper, we focus on the cases of taxation and minimum wages, allowing us to derive several additional insights.

The paper proceeds as follows. After reviewing the related literature in Section 2, the model setup is presented in Section 3. Section 4 solves for the optimal relational contract under employment and self-employment and discusses the implications for the firm's optimal choice between the two. Section 5 considers applications of the model, while Section 6 concludes. All proofs are relegated to the appendix.

2 Related Literature

Our paper is related to the literature that compares employment with market transactions (Wernerfelt, 1997, 2004, 2015; Hart and Moore, 2008; Levin and Tadelis, 2010). In particular, we contribute to the strand of that literature that emphasizes the firm's control rights under employment, which dates back to Coase (1937) and Simon (1951) and typically highlights adaptation to an ex ante uncertain production environment as the key advantage of employment. In our model, however, both employment and self-employment allow the contracting parties to continually adapt to the production environment; the two organizational forms differ only in the allocation of control over adaptation of the work design.[5]

Baker et al. (2011) analyze the optimal allocation of control to facilitate relational adaptation; however, they consider an abstract setting with an arbitrary number of parties and decisions. While control over these decisions is contractible ex ante, the optimal allocation of decision rights depends on the ex post realization of the state of the world. They show that the optimal allocation in the repeated game minimizes the parties' aggregate reneging temptation and hence typically differs from the optimal allocation in a spot setting. In our model, the work design is also adapted to the state of the world in each period, but the allocation of control is legally determined by the organizational form, which is assumed to be fixed; accordingly, control is non-transferable between periods, in contrast to Baker et al. (2011). Furthermore, the additional structure of the model allows us to make detailed predictions regarding how the optimal organizational form responds to changes in the economic environment, the structure of optimal contracts and worker compensation, as well as the impact of labor market policies such as minimum wages.

Raith (2023) compares employment and self-employment in an environment with complete contracts but asymmetric information, assuming that the choice between organizational forms determines real authority over the worker's action. In his paper, employment corresponds to a relational agreement for the worker to undertake productive actions in each period, while self-employment corresponds to sequential spot trade.[6] Inefficiencies under the former organizational form occur because trade may occur even when it is socially wasteful and under the latter because trade may not occur even when efficient. In our

[5]In fact, the main insights of our paper also apply when there is no uncertainty about the production environment, which is a special case of our model. This is because the work design is not verifiable, so that post contractual opportunistic behavior is always possible.

[6]In Raith (2023) as well as in Van den Steen (2010), Marino et al. (2010), and Rantakari (2023), the firm's control rights need to be self-enforcing under employment, since the worker retains formal authority. In contrast, we assume that the firm has the legal right to instruct an employed worker on how to execute the assignment. We discuss this assumption in the model section.

framework, trade between parties is always efficient (and always occurs), but contracting outcomes may nonetheless be inefficient due to contractual incompleteness; in particular, the non-verifiability of the work design implies that relational contracts play an important role under both organizational forms, in contrast to Raith's approach.

More broadly, our paper contributes to the literature on the theory of the firm. Unlike most of these papers, we focus on the relationship between a firm and a single worker (instead of two firms) and take into account that different organizational forms may come with different contracting environments due to labor market regulation. In our model, the defining feature of the organizational form is not asset ownership but the allocation of control in accordance with the legal system. Moreover, ex-ante relationship-specific investments and ex-post bargaining do not play a role in our model, which distinguishes our paper from the property-rights theory of the firm (Grossman and Hart, 1986; Hart and Moore, 1990).

Nonetheless, our paper is related to studies that consider how asset ownership and the associated allocation of decision rights impacts relational contracts. Adopting a property-rights setting, Baker et al. (2002) study optimal firm boundaries when an upstream and a downstream party cannot write formal contracts. They emphasize that different organizational forms entail different possibilities to renege on relational contracts, which is also the case in our paper. In their setting, however, optimal firm boundaries crucially depend on the characteristics of the spot market for the service provided by the downstream party and how these market characteristics affect price bargaining between the two parties under outsourcing. In our setting, there is no spot market and the parties cannot bargain.

Baker et al. (1999) study the delegation of decision-making authority in a principal-agent setting where the agent exerts non-observable effort to screen potential projects. They primarily focus on the case of informal delegation as a repeated-game equilibrium, but in an extension briefly consider formal delegation of authority by means of divestiture: when the principal sells the asset needed for production to the agent, the agent obtains full control over project implementation. They argue that different forms of asset ownership affect reneging temptations and thus impact relational agreements.

Finally, in an extension, we analyze the interaction between formal performance pay and relational contracts under employment and self-employment. The optimal combination of formal and relational contracts has been studied in various settings; seminal contributions include Baker et al. (1994), Schmidt and Schnitzer (1995), and Pearce and Stacchetti (1998). We are not aware of any paper that analyzes a model resembling the case of employment in our paper. Demougin and Fabel (2019) study a setting similar to the case of self-employment but focus on other aspects.

3 Model Setup

A principal contracts with an agent to repeatedly perform an assignment, which we formalize as an infinitely-repeated game in discrete time. Both parties are risk-neutral and discount the future at a common rate $\delta \in (0,1)$. In each period, the assignment's non-verifiable outcome $x \in \{0,1\}$ is either high ($x=1$), generating output $\pi > 0$ for the principal, or low ($x=0$), in which case output is zero. The probability of a high outcome is $s \in [0,1]$, where each value of s is associated with a specific work design. The agent's associated costs are $C(s,\theta)$, where $\theta \in [\underline{\theta}, \overline{\theta}] \subset (0,\infty)$ is a random variable representing the state of the world, which is distributed independently between periods according to the function $G(\cdot)$ with $G'(\cdot) = g(\cdot)$.[7] We assume that $C_s(s,\theta) > 0$ and $C_{ss}(s,\theta) > 0$ for all $s \in (0,1]$, with $C(0,\theta) = C_s(0,\theta) = 0$. We also assume that $C_\theta(s,\theta) < 0$ and $C_{s\theta}(s,\theta) < 0$, so that higher values of θ imply a more favorable state of the world, in the sense that each s is associated with lower costs and lower marginal costs for the agent.

The work design, s, captures the aspects of work discussed in the introduction, namely the manner or way in which the assignment is performed. Depending on the assignment, this could include elements such as the allocation of tasks; working hours and patterns such as the

[7]We allow for $\underline{\theta} = \overline{\theta}$, i.e., G can be degenerate. Our environment under self-employment with a degenerate G is a special case of the frameworks of MacLeod and Malcomson (1989) and Levin (2003).

duration and scheduling of shifts; the place of work, or equipment used. It could also include the extent to which the worker is supervised, monitored or required to communicate with the firm. More generally, s can include any aspect of the work process or work environment for which (i) parties have conflicting interests and (ii) control is legally determined by the organizational form. The model's assumptions imply that a higher s corresponds to a more demanding work design, which makes success more likely but entails higher costs for the agent. We assume that, regardless of the organizational form, s is observable to both parties.

The state of the world θ represents common ex ante uncertainty over production conditions that cannot be influenced by either party, which may arise from various sources. For the example of a food delivery platform considered in the introduction, θ could reflect idiosyncratic shocks to the working environment such as the weather. θ is realized in each period and observed by both the principal and the agent, but only **after** the contract has been signed. We assume that neither the work design s nor the state θ can be part of a formal contract, due to either the nature of the information or the prohibitive costs involved.

The agent is assumed to be wealth constrained, so that transfers from the agent to the principal are infeasible in all periods. This rules out the possibility of the principal selling the firm to the agent, but otherwise does not affect our main analysis. We normalize the reservation utilities of both parties to zero; extending the model to allow for non-zero outside options is straightforward, but does not yield interesting results.[8]

Given a particular realization of the state, the first-best work design $s^*(\theta)$ solves:

$$\max_{s\in[0,1]} s\pi - C(s,\theta) \tag{1}$$

[8]In Section 5.3, we allow the agent's reservation utility to be strictly positive in order to investigate the implications for formal performance pay, where variations in the outside option can have interesting effects. In this section, the assumption that the agent is wealth constrained plays an important role.

Our restrictions on $C(s, \theta)$ imply that $s^*(\theta)$ satisfies:

$$\pi = C_s(s, \theta) \tag{2}$$

so that $s^*(\theta)$ is strictly increasing in θ. We further impose that $s^*(\overline{\theta}) < 1$, implying the first-best work design is always at the interior of $[0, 1]$, and $\frac{d}{d\theta}[C(s^*(\theta), \theta)] > 0$, so that, when implementing $s^*(\theta)$, a greater realization of θ results in both a more demanding work design and higher associated costs for the agent.[9] The relationship's first-best surplus is:

$$S^* = \int_{\underline{\theta}}^{\overline{\theta}} [s^*(\theta)\pi - C(s^*(\theta), \theta)] \cdot g(\theta) \cdot d\theta > 0 \tag{3}$$

The focus of the paper is the principal's choice between employment and self-employment. The sole difference between the two organizational forms concerns which party has control over the work design s. Under employment, the principal retains control; that is, after observing the state θ, the principal is free to implement any $s \in [0, 1]$, with the associated costs being borne by the agent. We assume that the principal has no credible means of restricting her choice of s at the contracting stage and, as discussed below, that the agent cannot escape the costs associated with the chosen work design. In contrast, under self-employment, the agent retains control over s.

Since s, θ and x are all non-verifiable to third parties, the formal part of any contract under either organizational form simply specifies a non-negative transfer w from the principal to the agent. However, since these variables are all observable to both parties, contracting can be enhanced by self-enforcing relational agreements. We restrict our analysis to stationary contracts, i.e., the principal offers the same contract in every period as long as the contracting relationship continues, with no possibility of renegotiation.[10] Following any

[9] An example of a function that satisfies all necessary conditions when combined with an appropriate upper bound on π is $C(s, \theta) = \frac{1}{2\theta}s^2$, which we use throughout the paper when considering numerical examples. The restrictions are also satisfied by most power functions with appropriately chosen parameters.

[10] Since the agent's wealth constraints play no role in our main analysis, our environment under self-employment is a special case of Levin's (2003) framework, where stationary relational contracts are shown to be optimal. Slightly adjusting his arguments, one can show that, for given fallback positions, stationary

deviation from the informal agreement, relational contracting breaks down. In this case, the parties can either continue to interact on the terms of the formal part of the contract alone, or instead walk away from the relationship. The parties' fallback payoffs are thus determined endogenously by the economic environment in conjunction with the chosen organizational structure and contract.[11] Finally, we assume that the principal has all of the bargaining power and in each period makes a take-it-or-leave-it offer to the agent.[12]

The timing of the game is as follows. First, at time $t = 0$, the principal chooses the organizational form, employment or self-employment, and outlines a stationary contract, consisting of formal and relational elements. Both the organizational form and the contract are then fixed and cannot be changed in later periods. Then, in each subsequent period we have the following sequence of events:

1. The principal decides whether to offer the contract to the agent or not. If the contract is offered, the agent then decides whether to accept or reject. If the contract is offered and accepted then the game proceeds. In all other cases, the game ends and both parties obtain their reservation utilities.

2. The state θ is realized and observed by the principal and the agent.

3. The work design s is chosen by the principal (under employment) or by the agent (under self-employment).

4. The outcome $x \in \{0, 1\}$ is realized and transfers are made.

Before solving the model, we briefly discuss our assumptions. First, we assume the principal is free to implement any work design under employment, even when this choice represents

contracts are also optimal in our environment under employment.

[11]As we later show, the model's assumptions imply that, under either organizational form, parties prefer to terminate the relationship after a breakdown of the relational agreement, so that fallback payoffs are zero.

[12]For our main analysis, allocating the entirety of the bargaining power to the agent would not change the implemented work designs; only the distribution of the associated surplus is affected.

a deviation from a relational agreement. This control stems from her legal rights as an employer.[13] We also assume that the agent is unable to immediately exit the relationship to avoid a highly demanding work design, but can do so at the end of the period. Our assertion is that bringing an employment relationship to an immediate rather than a gradual end is often associated with additional costs. These include forfeiting wage payments through refusal to work, further income losses associated with quitting before lining up a new position, reputational damages or diminished career opportunities due to a poor reference or a breakdown in professional relationships, or other undesirable outcomes associated with a spell of unemployment.[14] In addition, in many legal systems an employee who quits without notice is in breach of a legal contract, which may lead to further costs: in the UK, government websites warn workers that they can be sued for damages following such breaches.[15]

A rational agent would only immediately quit if these costs were outweighed by the associated benefits. Our specification is therefore equivalent to an implicit assumption that these costs are sufficiently high such that the agent always prefers to comply with the principal's instruction in the short term, waiting until the end of the period to terminate the relationship, rather than break the formal contract and incur such damages. In particular, a model that allows the agent to ignore the principal's instruction by incurring costs that are greater than the agent's highest possible costs, $C(1, \underline{\theta})$, yields an identical analysis.

Second, we assume that contract renegotiation is infeasible. One particular implication of this assumption is that, after deviations from relational agreements, parties must trade under the formal terms of the existing contract. As we show in the ensuing analysis, this is not profitable under either organizational form, so that the relationship breaks down and parties receive their outside options. This is consistent with Levin (2003), who also assumes

[13]For example, in Germany, the employer has a legal right to determine the content, place and time of the performance of work, within reasonable limits (§106 of the German Industrial Code ("Gewerbeordnung"), `http://www.gesetze-im-internet.de/gewo/`).

[14]See Pohlan (2019) on the various negative impacts of unemployment, including social exclusion. Faberman et al. (2022) provide empirical evidence that job search by employed workers is significantly more effective and results in higher quality offers, relative to unemployed workers.

[15]See e.g. `https://www.nidirect.gov.uk/articles/breach-employment-contract`

that parties cease to trade following a breakdown in the relational agreement, but in contrast to other approaches (e.g., Baker et al., 1994), where trade continues after renegotiation of the contract. Under self-employment, spot contracting is not profitable, as the principal cannot create incentives, so that the analysis would be unaffected by the ability to renegotiate. Under employment, spot contracting after renegotiation may or may not be profitable, depending on the model's parameters. In the case where it is profitable, the optimal form of the relational contract and the key underlying trade-offs between employment and self-employment remain unchanged; however, some comparative static results may be affected in this case, since changes in the economic environment also impact fallback positions.

4 Contracting

4.1 Employment

We first analyze the case of employment. As outlined in the previous section, the contracting environment imposes that the formal part of any contract can consist only of a fixed transfer $w \geq 0$ from the principal to the agent. The remainder of the contract must be sustained by a relational agreement between the two parties. Specifically, the principal specifies a work design $s(\theta)$ associated with each realization of θ and promises to implement this schedule. As she has control over s, she always has the possibility to deviate from this promise once θ has been realized and instead impose an alternative work design. In particular, her optimal deviation (following any realization of θ) is to implement $s = 1$, maximizing expected output. The principal also has the possibility to incorporate a relational transfer $b(\theta)$ into the agreement, which either party has the option to refuse.[16] The contractual elements $\{w, b(\theta), s(\theta)\}$ must be designed in such a way that the parties' promises are self-enforcing, with any deviation from the agreement resulting in the principal and agent receiving their

[16]Since s is directly observable, the parties cannot benefit from conditioning the relational agreement on the assignment outcome x.

respective fallback payoffs, $f_P^E \geq 0$ and $f_A^E \geq 0$, for the remainder of the game. As discussed in the introduction, these payoffs are determined endogenously and depend on whether the parties continue to interact after a breakdown in the relational agreement.

Formally, the principal's problem is as follows:

$$\max_{s(\theta)\in[0,1],w,b(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\cdot\pi - b(\theta)]\cdot g(\theta)\cdot d\theta - w \tag{4}$$

subject to:

$$w, w+b(\theta') \geq 0 \quad \forall\, \theta' \in \left[\underline{\theta}, \overline{\theta}\right] \tag{5}$$

$$w + \int_{\underline{\theta}}^{\overline{\theta}} [b(\theta) - C(s(\theta),\theta)]\cdot g(\theta)\cdot d\theta \geq 0 \tag{6}$$

$$s(\theta')\cdot\pi - b(\theta') - w + \frac{\delta}{1-\delta}\left[\int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\cdot\pi - b(\theta)]\cdot g(\theta)\cdot d\theta - w\right] \geq \pi - w + \frac{\delta}{1-\delta} f_P^E \quad \forall\, \theta' \in \left[\underline{\theta}, \overline{\theta}\right] \tag{7}$$

$$w + b(\theta') - C(s(\theta'),\theta') + \frac{\delta}{1-\delta}\left[w + \int_{\underline{\theta}}^{\overline{\theta}} [b(\theta) - C(s(\theta),\theta)]\cdot g(\theta)\cdot d\theta\right] \geq w - C(s(\theta'),\theta') + \frac{\delta}{1-\delta} f_A^E \quad \forall\, \theta' \in \left[\underline{\theta}, \overline{\theta}\right] \tag{8}$$

The objective function (4) is simply expected per-period output net of payments to the agent. The constraints (5) and (6) respectively ensure that these payments are always non-negative and are sufficiently large to guarantee the agent's participation. The principal's feasibility constraint (7) ensures that she will not renege on the relational agreement by setting $s = 1$, in which case no discretionary transfers occur. The left-hand side of (7) consists of the sum of the present-period payoff and the discounted future utility stream from sticking to the agreement; the right-hand side is the sum of the present-period payoff from reneging and the future utility from the fallback payoff, f_P^E. The agent's feasibility constraint (8) can be explained similarly and ensures that he will not renege on payment of the discretionary transfer in cases where $b(\theta) < 0$. Both (7) and (8) must hold for all realizations of θ. The following lemma simplifies the principal's problem.

Lemma 1. *The solution to the problem:*

$$\max_{s(\theta)\in[0,1]} \int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\pi - C(s(\theta),\theta)] \cdot g(\theta) \cdot d\theta \tag{9}$$

subject to:

$$\frac{\delta}{1-\delta}\left[\int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\pi - C(s(\theta),\theta)] \cdot g(\theta) \cdot d\theta\right] \geq \pi\left[1 - s(\theta')\right] \quad \forall\, \theta' \in \left[\underline{\theta}, \overline{\theta}\right] \tag{10}$$

along with $w = \int_{\underline{\theta}}^{\overline{\theta}} C(s(\theta),\theta) \cdot g(\theta) \cdot d\theta$ *and* $b(\theta) = 0$ *for all* θ *also solves the problem (4)-(8).*

The lemma states that the optimal work design schedule maximizes the surplus, subject to an 'aggregate' feasibility constraint which ensures the discounted future surplus from a continuation of the relationship exceeds the principal's gain from her optimal deviation. Moreover, there are no relational transfers, and the fixed payment compensates the agent for his expected costs. Intuitively, the transfers between the parties, w and $b(\theta)$, have no impact on the aggregate feasibility constraint (10); changes in these payments loosen one feasibility constraint, but tighten the other. This implies that the principal can reduce payments until the agent's participation constraint binds, so that her goal is to maximize the surplus from the relationship. In turn, a binding participation constraint implies that the agent's total compensation is equal to his expected costs. This is insufficient to cover the agent's costs in the absence of a relational agreement, since in this case the principal sets $s = 1$ for any realization of θ. Accordingly, following a deviation by either party, the agent prefers to reject the contract in all future periods, so that the relationship breaks down and $f_P^E = f_A^E = 0$.

Since the agent's continuation utility is zero whether he sticks to the relational agreement or not, reneging on any payment $b(\theta) < 0$ yields a short-term benefit with no corresponding long-term loss. Accordingly, negative discretionary transfers are not feasible. Moreover, positive discretionary payments $b(\theta) > 0$ are not useful to the principal: since she has no need to create incentives under employment, the role of transfers is simply to remunerate the agent for his costs. While this can be done through either formal or relational transfers, the former is clearly more advantageous for the feasibility of the contract since the principal cannot renege on formal payments. It follows that discretionary transfers play no role under employment and $b(\theta) = 0$ for all θ. The proof of Lemma 1 verifies that, in this case, (10) is sufficient for both the individual feasibility constraints, (7) and (8), to hold.

Accordingly, the principal designs the work design schedule to maximize the surplus from the relationship, (9), given the constraint (10) that $s(\theta)$ must ensure adherence to the relational agreement. The following proposition solves this problem.

Proposition 1. *Suppose there exists a weakly positive attainable surplus under employment. Then the optimal contract implements a work design schedule that takes the form:*

$$s^E(\theta) = \begin{cases} \overline{s} & \underline{\theta} \leq \theta < \theta^E \\ s^*(\theta) & \theta^E \leq \theta \leq \overline{\theta} \end{cases} \tag{11}$$

where $\overline{s}$ is independent of θ with $\overline{s} > s^*(\theta)$ over the interval $\left[\underline{\theta}, \theta^E\right)$ and $\theta^E \in \left[\underline{\theta}, \overline{\theta}\right]$. The discretionary transfer $b^E(\theta)$ is zero for all θ and:

$$w^E = \int_{\underline{\theta}}^{\overline{\theta}} C(s^E(\theta), \theta) \cdot g(\theta) \cdot d\theta \tag{12}$$

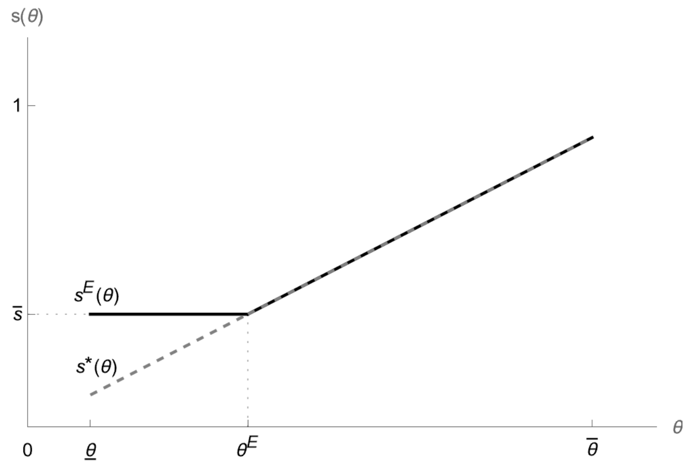

Figure 1: An example of the optimal work design schedule under employment, $s^E(\theta)$.

An example of the optimal work design schedule (11) is illustrated by Figure 1.[17] In the absence of the feasibility constraint (10), the principal would always maximize the relationship's surplus by choosing the first-best work design schedule, $s^*(\theta)$. For high values of the discount factor δ, this solution remains feasible and we have $\theta^E = \underline{\theta}$. However, for sufficiently low values of δ, setting $s(\theta) = s^*(\theta)$ for all θ is no longer possible. Intuitively, for low realizations of θ, implementing the first-best work design requires a relatively low value of $s(\theta)$. In such cases, the principal gains significantly by deviating from the agreement and setting $s = 1$. Accordingly, to facilitate a relational agreement the difference between $s(\theta)$ and 1 cannot be too large for any θ. The principal thus sets $s(\theta)$ to be as close to $s^*(\theta)$ as possible, subject to this restriction, which requires adjusting the work design schedule upwards for lower realizations of θ, so that the lower portion of $s^E(\theta)$ becomes flat. However, for larger realizations of θ, the first-best work design is already relatively demanding and thus the principal's temptation to deviate is small; accordingly, on the upper portion of $s^E(\theta)$ the prescribed work design remains equal to the first-best.

The extent to which the work design schedule must be adjusted depends, in particular, on the size of the discount factor. As δ becomes smaller, the left-hand side of (10) decreases, so that $\bar{s}$ must increase for the agreement to remain feasible. This also implies that θ^E must

[17]The figure illustrates the case where $C(s, \theta) = \frac{1}{2\theta}s^2$, $\theta \sim U(0.1, 0.9)$, with $\pi = 1$ and $\delta = 0.74$.

increase, so that $s^E(\theta)$ is distorted away from the first-best for a larger range of realizations of the state. As δ becomes very small, $s^*(\theta)$ cannot be implemented for any θ, in which case we have $\theta^E = \overline{\theta}$. The work design $\overline{s}$ is then constant over the entire schedule and greater than $s^*(\overline{\theta})$. In the limit, as $\delta \to 0$, (10) implies that implementing any $s < 1$ becomes infeasible.[18]

4.2 Self-Employment

Under self-employment, the agent retains control over the work design. As in the case of employment, formal contracting consists solely of a fixed transfer w from the principal to the agent, so that incentives for $s > 0$ must be created through a relational agreement. Specifically, the principal requests a work design $s(\theta)$ associated with each realization of θ and the agent promises to comply with this request. In turn, the principal promises, conditional on observing the agent's compliance, to pay the corresponding discretionary bonus $b(\theta)$. Both parties can renege on this agreement. The agent's optimal deviation is to implement the least costly work design by setting $s = 0$, thus incurring zero costs.[19] In this case, the principal pays no discretionary bonus and the parties receive their respective fallback payoffs, $f_P^{SE} \geq 0$ and $f_A^{SE} \geq 0$, for the remainder of the game. If the agent does not deviate and instead implements the promised work design, then the principal has the option to renege on her promise by refusing to pay the discretionary bonus. Similarly, parties then receive their fallback payoffs in each period thereafter. The contractual elements $\{w, b(\theta), s(\theta)\}$ must therefore be chosen such that both promises are self-enforcing.

Formally, the principal's problem is:

[18] If the model's parameters are such that $\pi \leq \int_{\underline{\theta}}^{\overline{\theta}} C(1, \theta) \cdot g(\theta) \cdot d\theta$, then a contract that implements $s = 1$ in each state cannot generate a non-negative surplus. In this case, for δ sufficiently small, a feasible relational agreement that yields a weakly positive profit to the principal may fail to exist. In such cases, spot contracting also fails to generate a weakly positive surplus, so that the parties prefer not to interact.

[19] In principle, for cases where $b(\theta) < 0$, the agent can deviate at two separate points in time: first, when implementing the work design and second, when required to pay $b(\theta)$. Clearly, the former of these is the more profitable deviation, since he benefits from setting $s = 0$ *and* from subsequently refusing to pay the discretionary transfer. In practice, however, we shall see that $b(\theta) > 0$ is required to incentivize $s > 0$, so that this distinction is unnecessary.

$$\max_{s(\theta)\in[0,1],w,b(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\cdot\pi - b(\theta)]\cdot g(\theta)\cdot d\theta - w \tag{13}$$

subject to:

$$w, w + b(\theta') \geq 0 \quad \forall \theta' \in \left[\underline{\theta}, \overline{\theta}\right] \tag{14}$$

$$w + \int_{\underline{\theta}}^{\overline{\theta}} [b(\theta) - C(s(\theta),\theta)]\cdot g(\theta)\cdot d\theta \geq 0 \tag{15}$$

$$s(\theta')\pi - b(\theta') - w + \frac{\delta}{1-\delta}\left[\int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\cdot\pi - b(\theta)]\cdot g(\theta)\cdot d\theta - w\right] \geq s(\theta')\pi - w + \frac{\delta}{1-\delta} f_P^{SE} \quad \forall \theta' \in \left[\underline{\theta}, \overline{\theta}\right] \tag{16}$$

$$w + b(\theta') - C(s(\theta'),\theta') + \frac{\delta}{1-\delta}\left[w + \int_{\underline{\theta}}^{\overline{\theta}} [b(\theta) - C(s(\theta),\theta)]\cdot g(\theta)\cdot d\theta\right] \geq w + \frac{\delta}{1-\delta} f_A^{SE} \quad \forall \theta' \in \left[\underline{\theta}, \overline{\theta}\right] \tag{17}$$

The objective function (13) and the constraints (14) and (15) are identical to the case of employment. The principal's feasibility constraint (16) guarantees that she will not renege on payment of the relational transfer $b(\theta)$, while the agent's feasibility constraint (17) ensures that he will not deviate by implementing $s = 0$, avoiding any costs but also forfeiting the discretionary payment. As before, both constraints must hold for all realizations of θ. The following lemma simplifies the principal's problem.

Lemma 2. *The solution to the problem:*

$$\max_{s(\theta)\in[0,1]} \int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\cdot\pi - C(s(\theta),\theta)]\cdot g(\theta)\cdot d\theta \tag{18}$$

subject to:

$$\frac{\delta}{1-\delta}\left[\int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\cdot\pi - C(s(\theta),\theta)]\cdot g(\theta)\cdot d\theta\right] \geq C(s(\theta'),\theta') \quad \forall\,\theta' \in \left[\underline{\theta},\overline{\theta}\right] \tag{19}$$

in combination with $w = 0$ *and* $b(\theta) = C(s(\theta),\theta)$ *also solves the problem (13)-(17).*

As in the case of employment, transfers between the parties do not affect the 'aggregate' feasibility constraint (19), which implies that the fixed payment w can be reduced until the agent's participation constraint binds. Accordingly, the principal's goal is to maximize the surplus from the relationship. Moreover, since the transfers affect neither the surplus nor the aggregate feasibility constraint, they can be chosen freely in such a way that both the individual feasibility constraints and the financial constraints are satisfied. Specifically, the proof of Lemma 2 shows that the principal can set $w = 0$ and $b(\theta) = C(s(\theta),\theta)$; in this case (19) becomes both necessary and sufficient for (16) and (17) to be satisfied and hence the work design schedule to be implementable. Finally, note that in the absence of a relational agreement, the principal has no means of creating incentives for the agent; accordingly, $f_P^{SE} = f_A^{SE} = 0$.

Proposition 2. *The optimal relational contract under self-employment is characterized by a fixed payment* $w^{SE} = 0$ *and a bonus schedule:*

$$b^{SE}(\theta) = \begin{cases} C(s^*(\theta),\theta) & \underline{\theta} \leq \theta \leq \theta^{SE} \\ \overline{b} & \theta^{SE} < \theta \leq \overline{\theta} \end{cases} \tag{20}$$

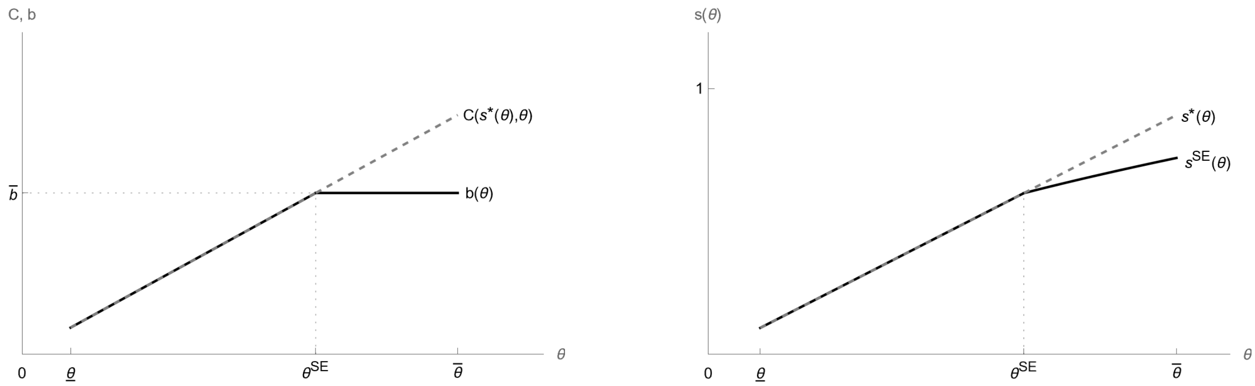

Figure 2: An example of the optimal bonus and work design schedules under self-employment, $b^{SE}(\theta)$ and $s^{SE}(\theta)$.

where $\overline{b}$ is independent of θ and $\theta^{SE} \in [\underline{\theta}, \bar{\theta}]$. The associated work design schedule is then:

$$s^{SE}(\theta) = \begin{cases} s^*(\theta) & \underline{\theta} \leq \theta \leq \theta^{SE} \\ \tilde{s}(\theta, \overline{b}) & \theta^{SE} < \theta \leq \overline{\theta} \end{cases} \tag{21}$$

where $\tilde{s}(\theta, \overline{b})$ is implicitly defined by:

$$C(\tilde{s}, \theta) = \overline{b} \tag{22}$$

for each realization of θ. Moreover, we have $\tilde{s}(\theta, \overline{b}) < s^*(\theta)$ over the region $(\theta^{SE}, \overline{\theta}]$.

An example of the optimal bonus and work design schedules is illustrated by Figure 2.[20] Absent the aggregate feasibility constraint (19), the principal would again maximize surplus by implementing the first-best work design schedule, $s^*(\theta)$. If δ is sufficiently high, this continues to be feasible and we have $\theta^{SE} = \overline{\theta}$. However, for lower values of δ the aggregate feasibility constraint binds for some realizations of θ. Intuitively, both $s^*(\theta)$ and the associated costs to the agent $C(s^*(\theta), \theta)$ are increasing in θ. For high realizations of θ, providing sufficient incentives to the agent would therefore require paying a large discretionary bonus; in this case, the principal would prefer to renege on the agreement by refusing to make the

[20] The figure illustrates the case where $C(s, \theta) = \frac{1}{2\theta} s^2$, $\theta \sim U(0.1, 0.9)$, with $\pi = 1$ and $\delta = 0.55$.

transfer. Accordingly, in order to facilitate the relational contract, discretionary transfers cannot be too large for any realization of θ. It follows that for high realizations of θ, the principal makes the bonus as high as possible, subject to this restriction, so that the bonus schedule $b^{SE}(\theta)$ becomes flat on its upper portion. In turn, this implies that for these realizations of θ, the demanded work designs must also be reduced, so $s^{SE}(\theta)$ becomes flatter over this section. For lower realizations of θ however, the first-best continues to be implemented since the associated costs—and hence bonus payments—are sufficiently low.

The size of the maximum permissible bonus payment $\bar{b}$ depends, in particular, on the discount factor. As δ becomes smaller, the left-hand side of (19) decreases, which implies that the maximal costs borne by the agent must be reduced, so that bonuses remain sufficiently low for the agreement to be feasible. This decrease in $\bar{b}$ also implies that the work design schedule is distorted away from the first-best for a greater range of realizations of θ; that is, θ^{SE} is also reduced. As δ becomes very small, $s^*(\theta)$ is not implementable for even the lowest realizations of θ. In this case, $\theta^{SE} = \underline{\theta}$ and the discretionary transfer $b^{SE}(\theta) = \bar{b}$ becomes constant over the whole schedule. In contrast, while $s^{SE}(\theta)$ becomes flatter, disparate work designs continue to be implemented for each realization of θ. In the limit as $\delta \to 0$, the principal would renege on any positive discretionary transfer, so that implementing any $s > 0$ is infeasible.

4.3 Comparison of the Two Organizational Forms

Our analysis has shown that, under both organizational forms, the principal chooses the work design that maximizes the relationship's surplus, subject to a feasibility constraint. The left-hand sides of the two feasibility constraints—(10) and (19)—are identical. However, they differ on the right-hand side. In the case of employment, the discounted future surplus from the relationship must be greater than the principal's benefit from deviating by implementing $s = 1$. When δ is small and the constraint binds, this implies that $s(\theta)$ must be increased above the first-best for low realizations of θ for the contract to remain credible. In contrast, the discounted future surplus under self-employment must be greater than the principal's temptation to renege on payment of the highest relational bonus. Since more costly work designs require larger bonuses, a binding constraint implies under-demanding work designs for high realizations of θ in order to curtail this temptation.

These observations have important implications for the principal's optimal choice between the two organizational forms. When δ is sufficiently low, both employment and self-employment necessarily require implementing inefficient work designs for some realizations of θ. However, these inefficiencies differ in both their nature (under-demanding vs. over-demanding work designs) and when they occur (low realizations vs. high realizations of θ). The principal therefore chooses the organizational form best able to minimize inefficiencies for the particular production environment. In situations where there are small gains to trade—where π is small, or the distribution function G is such that low realizations of θ are common—the surplus maximizing work design in each period is unlikely to be very demanding; in this case, self-employment is the optimal organizational form, since the principal's incentive to renege on the agreement is typically low because only a small bonus is required to provide appropriate incentives. In contrast, the principal's incentive to renege under employment is often high, as she can considerably gain by implementing a more costly work design. Conversely, the model predicts that employment becomes optimal in environments with large expected gains to trade. The following proposition formalizes these ideas.

Proposition 3. *For sufficiently low π, self-employment is the strictly optimal organizational form.*

Along similar lines, Figure 3 illustrates the impact of a shift in the distribution of θ. Specifically, the example assumes that $\theta \sim U(\tilde{\theta} - 0.2, \tilde{\theta} + 0.2)$ and considers profit under the two organizational forms as $\tilde{\theta}$ increases.[21] The dotted line plots the first-best surplus from the relationship, which increases as the production environment becomes more favorable. The darker solid line shows the profit associated with employment, which is also increasing and

[21] In the example, we have $C(s, \theta) = \frac{1}{2\theta} s^2$, $\pi = 1$, $\delta = 0.45$ and let $\tilde{\theta}$ vary between 0.68 and 0.78.

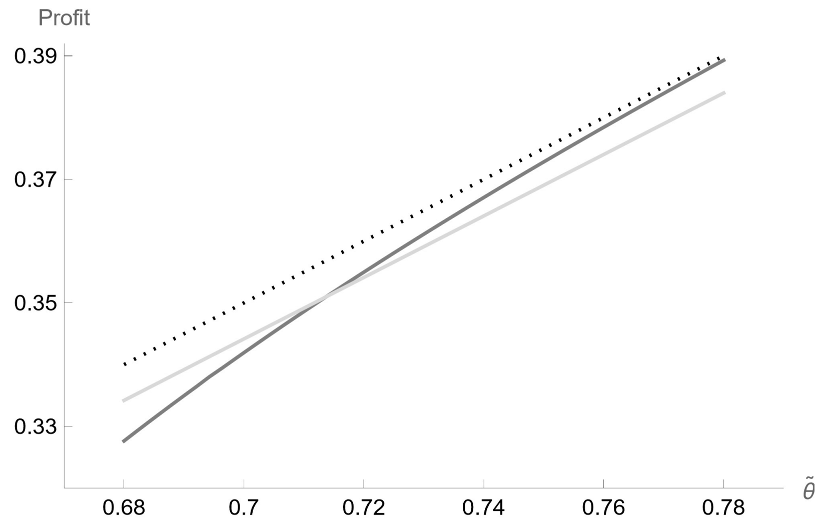

Figure 3: Profit under employment and self-employment.

approaches the first-best as $\tilde{\theta}$ grows large. The lighter solid line plots the profit associated with self-employment. In line with the foregoing discussions, for relatively low values of $\tilde{\theta}$ self-employment is optimal; however, as the production environment becomes more favorable, employment eventually becomes dominant.[22]

Another implication of our findings is that the choice of organizational form can impact the degree to which the relationship responds to changes in the production environment. The following two propositions formalize this idea.

Proposition 4. *For δ sufficiently small, there is a strictly positive probability that the work design under employment does not change between periods. This is never the case under self-employment.*

Intuitively, under employment, when δ is sufficiently low, $s^E(\theta)$ becomes flat in sections, so that the agent's work design does not respond to variations in the environment; indeed, for very small values of δ, s is constant for all realizations of θ. In contrast, under self-

[22] As noted by Mas and Pallais (2020), empirical evidence for what drives a firm's decision whether to employ their workers is scarce. Abraham and Taylor (1996) and Slaughter and Ang (1996) suggest that firms tend to employ workers in areas that are key to their business, while using contractors in more peripheral roles. Oyer (2020) notes that the majority of labor buyers on the online freelancing marketplace Upwork are small firms with less than 10 employees. The result is also consistent with empirical evidence from the broader literature that studies the boundaries of the firm; for instance, Alfaro et al. (2023) find that final good producers are more likely to integrate suppliers of more valuable inputs.

employment the work design always strictly increases in θ, so that two separate realizations cannot be associated with the same s.

Proposition 5. *Let $W^E(\theta)$ and $W^{SE}(\theta)$ denote the agent's total compensation under employment and self-employment, respectively. Then:*

(i) $\text{Var}\left[W^{SE}(\theta)\right] \geq \text{Var}\left[W^E(\theta)\right] = 0.$

(ii) $\int_{\underline{\theta}}^{\overline{\theta}} W^E(\theta) \cdot g(\theta) \cdot d\theta \geq \int_{\underline{\theta}}^{\overline{\theta}} W^{SE}(\theta) \cdot g(\theta) \cdot d\theta.$

Under both organizational forms, the participation constraint binds, so that the agent's expected total compensation is equal to his expected costs. However, there are differences in the compensation structure. Under employment, since there is no need to create incentives for the worker, he is remunerated entirely through a fixed wage that does not change between periods. In contrast, a self-employed worker is compensated through a discretionary bonus that is equal to his realized costs and therefore typically varies from period-to-period; accordingly, pay under self-employment is more variable than under employment. Altogether, the results presented in Proposition 4 and part (i) of Proposition 5 predict that employment will typically represent a more rigid form of labor relationship than self-employment—both in terms of work design and compensation—consistent with empirical evidence (Hurst and Pugsley, 2011; Boeri et al., 2020).

Finally, part (ii) of Proposition 5 states that there are also differences in the level of worker compensation. From Propositions 1 and 2, we have $s^{SE}(\theta) \leq s^*(\theta) \leq s^E(\theta)$ for all θ, which immediately implies that expected costs are greater under employment. Since the worker is compensated his costs, the model then predicts that employed workers can expect higher total compensation relative to self-employment, due to the more demanding work design implemented in this case. This result is also consistent with recent empirical findings (Boeri et al., 2020).

To conclude the section, note that the model's key trade-offs do not depend on variations in the state. When G is degenerate so that θ is constant, the same work design is implemented in each period. Under employment, s must be chosen such that the principal will not renege by instead implementing the most demanding job design; under self-employment, it must be chosen such that the associated bonus is sufficiently low. It follows that (i) employment is better suited to favorable production environments, where s^* is high, and (ii) total compensation is higher under employment, analogous to Propositions 3 and 5 above. In the remainder of the paper, to keep the analysis concise, we assume G is degenerate.

5 Applications of the Model

The previous section presented a general model that highlighted differences in the outcomes of relational contracting under employment and self-employment, as well as the ensuing implications for firms' choices of labor relationships. In this section, we apply the model to investigate the effects of income taxation and minimum wages; we also consider the implications of formal performance pay for our results.

5.1 Income Taxation

In most countries, workers are subject to income taxation, which has the potential to create a distortionary effect on the labor relationship. Income taxation typically applies to both employed and self-employed workers, though often to a differing extent. In this subsection, we assume for simplicity that under both employment and self-employment the agent's income is taxed at the (identical) proportional rate $\tau \in [0, 1)$. This allows us to compare the distortionary effect under the two organizational forms. We first solve the principal's problem in each case, before discussing the associated implications of income taxation.

Following the analysis in Section 4, with a degenerate distribution function the principal's problem under **employment** becomes:

$$\max_{w \geq 0,\ s \in [0,1]} \quad s\pi - w \tag{23}$$

subject to:

$$(1-\tau)\,w - C(s,\theta) \geq 0 \tag{24}$$

$$s\pi - w + \frac{\delta}{1-\delta}\left[s\pi - w\right] \geq \pi - w \tag{25}$$

where, as before, discretionary bonus payments cannot be optimal and are therefore excluded from the analysis. As in Section 4, the participation constraint (24) binds, which in turn implies $w \geq 0$. The problem can then be rewritten as:

$$\max_{s\in[0,1]} \; s\pi - \frac{C(s,\theta)}{1-\tau} \tag{26}$$

subject to:

$$\frac{\delta}{1-\delta}\left[s\pi - \frac{C(s,\theta)}{1-\tau}\right] \geq \pi\left[1-s\right] \tag{27}$$

The solution to this problem depends on whether the feasibility constraint (27) binds. If not, the optimal work design is implicitly defined by:

$$(1-\tau)\,\pi = C_s(s,\theta) \tag{28}$$

If (26)-(27) has a solution and the constraint binds, the optimal work design is the smallest $s \in [0,1]$ that satisfies:

$$\pi\,(1-\tau)\,(s+\delta-1) - \delta C(s,\theta) = 0 \tag{29}$$

so long as the associated profit, $s\pi - \frac{C(s,\theta)}{1-\tau}$, is strictly positive. Otherwise, the principal will not choose employment.

The principal's problem under **self-employment** is:

$$\max_{w,b,s\in[0,1]} \quad s\pi - b - w \tag{30}$$

subject to:

$$w, w+b \geq 0 \tag{31}$$

$$(1-\tau)(w+b) - C(s,\theta) \geq 0 \tag{32}$$

$$(1-\tau)(w+b) - C(s,\theta) + \frac{\delta}{1-\delta}\left[(1-\tau)(w+b) - C(s,\theta)\right] \geq (1-\tau)w \tag{33}$$

$$s\pi - b - w + \frac{\delta}{1-\delta}\left[s\pi - b - w\right] \geq s\pi - w \tag{34}$$

Following the same approach as in Section 4, one can show that the optimal contract sets $w = 0$ and $b = \frac{C(s,\theta)}{(1-\tau)}$. This immediately implies that both non-negativity constraints (31) are satisfied and that the participation constraint (32) and agent's feasibility constraint (33) hold with equality. It then follows that the principal's problem can be rewritten as:

$$\max_{s\in[0,1]} \quad s\pi - \frac{C(s,\theta)}{1-\tau} \tag{35}$$

subject to:

$$\frac{\delta}{1-\delta}\left[s\pi - \frac{C(s,\theta)}{1-\tau}\right] \geq \frac{C(s,\theta)}{1-\tau} \tag{36}$$

The objective functions (26) and (35) are identical. Accordingly, in the absence of binding feasibility constraints, the principal implements the same work design—implicitly defined by (28)— under both employment and self-employment. Otherwise, when (36) binds, the optimal work design under self-employment is the largest $s \in [0,1]$ which satisfies:

$$\delta s\pi(1-\tau) - C(s,\theta) = 0 \tag{37}$$

Such an s always exists and yields a non-negative surplus. The following lemma shows how the optimal work designs under the two organizational forms change as the tax rate increases.

Lemma 3. *If the optimal work designs under employment, s^E, and self-employment, s^{SE}, are implicitly defined by* (29) *and* (37) *respectively, then (i) s^E is strictly increasing in τ and (ii) s^{SE} is strictly decreasing in τ.*

Under either organizational form, an increase in τ reduces the value of the future relationship, since the principal must increase the agent's compensation to ensure participation. Under employment, this tightens the feasibility constraint (27), so that the implemented work design must be increased and becomes even more over-demanding. By the same logic, the reduced surplus tightens (36) under self-employment. However, in this case there is an additional effect on the right-hand side of the feasibility constraint that is absent under employment. Higher taxes also imply that bonus payments must be increased, further raising the principal's incentive to deviate and tightening (36). Accordingly, both effects work in the same direction, leading to a lower s^{SE}. In this respect, our results suggest that taxation distorts the outcome of relational contracting to a greater extent under self-employment, as incentives must be provided through monetary payments—and are thus directly impacted by taxes—whereas this is not the case under employment.

However, taxation also affects the (identical) objective functions (26) and (35). Since the costs of implementing work designs are increased by taxation, a higher τ implies the (tax conditional) optimal work design—implicitly defined by (28)—is reduced. From the previous section, self-employment is well-suited to implementing small values of s, whereas this is particularly difficult under employment. It follows that this effect tends to favor self-employment as an organizational form. The following result shows that, with a sufficiently

Figure 4: An example of taxation, for cases where $\tau = 0.1$ (left-hand panel) and $\tau = 0.3$ (right-hand panel).

high tax rate, the principal will always choose self-employment.

Proposition 6. *There exists a critical level of taxation, $\hat{\tau} < 1$, such that for all $\tau \geq \hat{\tau}$, self-employment is the strictly optimal organizational form.*

Intuitively, as τ becomes large, the maximum attainable surplus necessarily becomes very small. In this case, from (27), employment is only compatible with s close to 1. Such work designs are very inefficient with high taxation, since the principal must remunerate the agent with extremely large compensation for his costs. For sufficiently high τ, it therefore becomes impossible for the relationship to remain profitable under employment. Under self-employment, from (36), a small surplus implies that the principal can only implement s close to 0. However, since the costs associated with such work designs are small, this is relatively efficient with high taxes and the principal can still choose an s that yields a strictly positive profit.

Figure 4 illustrates a numerical example.[23] The thick black curve shows the principal's profit, for s between zero and one. The thin black curves illustrate the feasibility constraints under self-employment (on the LHS of each graphic) and employment (on the RHS). They depict the difference between the LHS and the RHS of (36) and (27), respectively. The intervals associated with the shaded regions below these curves then show the set of work

[23]The figure considers the case where $C(s, \theta) = \frac{1}{2\theta}s^2$, with $\pi = 1.4$, $\theta = 0.55$ and $\delta = 0.4$.

designs—$[0, s^{SE}]$ and $[s^E, 1]$—that are implementable under the two organizational forms.

The left-hand panel considers a low tax environment, $\tau = 0.1$, in which neither organizational form can implement the tax-conditional surplus maximizing work design, s^*, defined by (28). However, both forms can induce relatively efficient work designs, with profit slightly higher under employment. The right-hand panel shows the impact of a raise in the tax rate to $\tau = 0.3$. Since the increase in taxes reduces the relationship's surplus, work designs under both organizational forms become less efficient: for employment, we have an increase in s^E, while there is a decrease in s^{SE} under self-employment. However, there is also a change in the principal's profit curve, since each work design becomes more costly to implement, so that the tax-conditional optimal work design is reduced. In this case, it is clear from the graphic that self-employment yields a significantly higher profit than employment.[24]

5.2 Minimum Wages & Non-Pecuniary Benefits

Labor relationships between firms and employees are typically subject to various regulations regarding the worker's remuneration package. These include minimum wage laws, as well as legislation that guarantees employees additional benefits (e.g. pension contributions, statutory holiday and sick pay, parental leave). From the firm's perspective, such regulation represents an additional cost of employment relative to self-employment.[25] This section analyzes the impact of such policies on the outcome of contracting under employment, by considering the introduction of a minimum wage. In particular, we show that minimum wages change the implemented work design and reduce the relative profitability of employment. More importantly, our model also yields predictions regarding the circumstances under which minimum wages are likely to be most successful at achieving policy goals.

Technically, we require that all payments from the principal to the agent must be greater

[24]There is an empirical literature that studies the interaction between taxes and the employment vs. contractor decision (e.g. Gentry and Hubbard, 2000). However, these papers tend to consider the supply-side, rather than the demand-side, and often focus on issues such as differences in marginal tax rates or tax progressivity, rather than the implications of taxation for the design of labor relationships.

[25]Even if self-employed workers are eligible for such benefits, in many cases they are significantly less generous; see the discussion in Boeri et al. (2020).

than $\underline{w} \geq 0$. In the case of a degenerate distribution function, the principal's problem can then be expressed as (23)–(25) for the case where $\tau = 0$ and where we additionally require that $w \geq \underline{w}$. Clearly, for sufficiently large values of $\underline{w}$, the principal is unable to simultaneously satisfy the minimum wage constraint while still achieving a positive profit. We ignore such cases and restrict attention to situations in which $\underline{w}$ is sufficiently small (in combination with the remaining parameters of the model) such that there exists a feasible contract in which the principal's profit is non-negative. The following lemma then solves the principal's problem.

Lemma 4. *The optimal work design under employment, s^E, is weakly increasing in the minimum wage $\underline{w}$. The agent earns a rent if and only if $\underline{w} > C(1, \theta)$, in which case $s^E = 1$ so that the most demanding work design is implemented.*

The proof of Lemma 4 shows that the impact of a minimum wage constraint can be separated into three distinct intervals. From Section 4, under employment the principal pays a strictly positive fixed wage that compensates the agent's costs; accordingly, for low values of $\underline{w}$, the minimum wage constraint does not bind and the additional requirement has no impact on the outcome of contracting. For middling values of $\underline{w}$, the minimum wage constraint begins to bind. In this case, the principal responds to the requirement of increased compensation by demanding a more costly work design from the agent, compared to the case with no minimum wage, so that the participation constraint continues to bind and the agent's improved income is entirely offset by higher costs. Due to the increased work design, the feasibility constraint becomes slack. It follows that over this region, an increase in $\underline{w}$ has no impact on the agent's expected utility, while strictly decreasing the efficiency of the relationship and the principal's expected payoff; in other words, the introduction of a minimum wage is associated with a Pareto loss. Finally, for sufficiently high $\underline{w}$ the principal sets $s = 1$ as she is unable to respond to any further increases in the minimum wage by demanding a more costly work design; in this case, the sole impact of further increases in $\underline{w}$ is to raise the agent's wage and utility.

These findings yield two key predictions regarding the effectiveness of minimum wages under employment. First, while a binding minimum wage transfers wealth from the principal to the agent, this does not necessarily improve the agent's utility, due to the associated increase in his costs. It is only for large minimum wages that the agent's increased compensation translates to improved well-being. Second, this improved well-being is necessarily associated with a reduction in efficiency. However, the size of this reduction will vary between industries. In particular, for favorable environments where the work design is already strongly aligned with the principal's interests (i.e. s^E close to 1), the efficiency costs of a high minimum wage are relatively small. In addition, our model predicts that employment will typically be optimal in such situations, so that firms find it difficult to avoid the minimum wage by switching to self-employment. In contrast, the model predicts that minimum wages will be less effective in less favorable production environments: either because of the large associated reduction in efficiency or because the principal will prefer to switch to an organizational design of self-employment. Along similar lines, the efficiency costs of a minimum wage will also be small in situations where δ is low, since in such cases the implemented work design under employment is very over-demanding, and thus inefficient, even in the absence of regulation.

5.3 Formal Performance Pay

We now suppose that the assignment outcome x is verifiable and can be formally contracted on by the parties. Specifically, we assume that following a successful assignment, the agent receives a proportion $r \in [0, 1]$ of the output π, with the remaining $(1 - r)\,\pi$ accruing to the principal. As x is verifiable, these transfers can be enforced by courts and do not form a relational agreement between the parties. Nonetheless, formal contracting alone is unable to achieve the first-best outcome, due to the agent's constrained wealth, so that relational agreements continue to play an important role.

The purpose of this section is twofold. First, we show that while formal performance pay complements relational contracts under both organizational forms, allowing the principal to increase her profits, the fundamental trade-off of the model persists. In particular, the parties continue to implement over- and under-demanding work designs under employment and self-employment respectively, relative to the first-best. Second, we allow the agent's outside option to be positive, $\underline{u} \geq 0$, and show that changes in this parameter have an important impact on the principal's choice of organizational form when formal performance pay is available. Throughout the section, we abstract from the role of taxation, but continue to assume that transfers to an employed worker must satisfy a minimum wage constraint.

We first consider the case of **employment**.[26] Following the analysis in Section 4, the principal's problem can be expressed as:

$$\max_{w \geq \underline{w},\, r \in [0,1],\, s \in [0,1]} \quad s\,(1-r)\,\pi - w \tag{38}$$

subject to:

$$w + sr\pi - C(s,\theta) \geq \underline{u} \tag{39}$$

$$s\,(1-r)\,\pi - w + \frac{\delta}{1-\delta}\left[s\,(1-r)\,\pi - w\right] \geq \pi\,(1-r) - w \tag{40}$$

We restrict attention to the case where $\underline{w} < \underline{u} + C(1,\theta)$, so that the parties' relationship can benefit from the use of relational agreements.[27] The following lemma then summarizes the solution to this problem.

Lemma 5. *Suppose that $\underline{w} < \underline{u} + C(1,\theta)$. The optimal contract sets $w = \underline{w}$ and r equal to:*

$$r^E(s) = \frac{\underline{u} - \underline{w} + C(s,\theta)}{s\pi} \tag{41}$$

[26]It is straightforward to verify that there continues to be no role for discretionary bonus payments under employment. Moreover, in the absence of a relational agreement the principal sets $s = 1$, which implies that the agent prefers to reject the contract offer after a deviation by either party, similar to the analysis in Section 4. Accordingly, each party's fallback utility remains zero.

[27]When $\underline{w} \geq \underline{u} + C(1,\theta)$, the principal sets $s = 1$, which is always implementable under employment.

For high values of δ, the principal can implement the first-best work design, s^*. Otherwise, if δ is sufficiently low, the optimal work design under employment, $s^E > s^*$, solves the equation:

$$\delta \left[s\pi - C(s,\theta) - \underline{u} \right] - (1-\delta) \left(s\pi - C(s,\theta) - \underline{u} + \underline{w} \right) \left(\frac{1-s}{s} \right) = 0 \tag{42}$$

The proof of Lemma 5 shows that as long as the minimum wage is not excessively high, $\underline{w} < \underline{u} + C(1,\theta)$, the agent's participation constraint continues to bind. The principal then faces a choice between compensating the agent's costs via the fixed payment w or through performance-contingent pay r. The lemma shows that it is optimal to set the former as low as possible (i.e. $w = \underline{w}$) with the remainder of the remuneration coming via a proportion of output following a successful assignment so that $r = r^E(s)$.

Intuitively, the principal's incentive to deviate from the relational agreement under employment comes from her ability to set $s = 1$, thereby increasing current-period expected output. When $r > 0$, this incentive to deviate is reduced—since she receives only $1 - r$ of the output after a successful assignment—relaxing the feasibility constraint. In contrast, variations in the fixed payment w do not affect her incentive to deviate. It follows that since she must remunerate the agent for his outside option and costs, this is best done by setting r as large as possible, subject to the minimum wage. Accordingly, the ability to formally condition compensation on the assignment outcome complements relational contracts, thereby increasing the surplus from the relationship, even though there is no need to create incentives for the agent under employment.[28] As in Section 4, for high values of δ the principal maximizes surplus by setting $s = s^*$. For sufficiently small δ it remains infeasible to implement s^*, in which case the feasibility constraint binds and the work design is given by (42).

Under **self-employment**, the formal elements of the contract similarly consist of a fixed wage w and a proportion of output $r \in [0, 1]$ to be paid to the agent when $x = 1$. In the

[28]We are not aware of any other paper that highlights this particular advantage of formal performance pay when parties relationally contract with one another.

absence of any relational agreement, the agent would choose the work design s that solves the following optimization problem:

$$\max_{s\in[0,1]} \quad w + sr\pi - C(s,\theta) \tag{43}$$

Let us denote the solution to this problem by $\hat{s} \equiv \hat{s}(r,\pi,\theta)$. Our conditions on $C(\cdot)$ then imply that $\hat{s}$ solves:

$$r\pi - C_s(s,\theta) = 0 \tag{44}$$

Intuitively, $r > 0$ implies that the agent implements a strictly positive s even in the absence of discretionary payments. This has two important effects. First, his optimal deviation from any relational agreement is no longer setting $s = 0$, but rather $s = \hat{s}$. Second, following a deviation by either party, both the principal and the agent find it profitable to trade on the terms of the existing contract, so that their interaction continues; specifically, for a given w and r their respective fallback utilities become:

$$f_P^{SE} = \hat{s}\,(1-r)\,\pi - w \geq 0 \tag{45}$$

$$f_A^{SE} = w + \hat{s}r\pi - C(\hat{s},\theta) \geq 0 \tag{46}$$

From the foregoing, the principal's problem under self-employment is:

$$\max_{w,\,b,\,r\in[0,1],\,s\in[0,1]} \quad s\,(1-r)\,\pi - b - w \tag{47}$$

subject to:

$$w, w + b \geq 0 \tag{48}$$

$$w + sr\pi + b - C(s,\theta) \geq \underline{u} \tag{49}$$

$$\begin{aligned} & w + sr\pi + b - C(s,\theta) + \frac{\delta}{1-\delta}(w + sr\pi + b - C(s,\theta)) \\ & \qquad \geq w + \hat{s}r\pi - C(\hat{s},\theta) + \frac{\delta}{1-\delta}\left(w + \hat{s}r\pi - C(\hat{s},\theta)\right) \end{aligned} \tag{50}$$

$$\begin{aligned} & s\,(1-r)\,\pi - b - w + \frac{\delta}{1-\delta}(s\,(1-r)\,\pi - b - w) \\ & \qquad \geq s\,(1-r)\,\pi - w + \frac{\delta}{1-\delta}\left(\hat{s}\,(1-r)\,\pi - w\right) \end{aligned} \tag{51}$$

The following lemma outlines the solution to this problem.

Lemma 6. *The optimal contract under self-employment with formal performance measurement sets $w = 0$ and $b = \delta\,(1-r)\,\pi\,(s - \hat{s})$. For high values of δ, the principal can implement the first-best work design, s^*. Otherwise, if δ is sufficiently low, we have two possible cases:*

(i) If the agent's participation constraint binds at the optimal solution, then s^{SE} and r^{SE} are implicitly defined by the following system:

$$C(s,\theta) + \underline{u} - sr\pi = \delta\,(1-r)\,\pi\,(s - \hat{s}) \tag{52}$$

$$\hat{s}r\pi - C(\hat{s},\theta) - [sr\pi - C(s,\theta)] = \delta\,(1-r)\,\pi\,(s - \hat{s}) \tag{53}$$

(ii) If the agent's participation constraint does not bind at the optimal solution, then s^{SE}

and r^{SE} are implicitly defined by the following system:

$$\pi - C_s(s,\theta) + \frac{\hat{s}\pi\left[\delta\left(1-r\right)\pi + r\pi - C_s(s,\theta)\right]}{\left(1-\delta\right)\pi\left(s-\hat{s}\right) - \delta\left(1-r\right)\pi\frac{\partial\hat{s}}{\partial r}} = 0 \tag{54}$$

$$\hat{s}r\pi - C(\hat{s},\theta) - \left[sr\pi - C(s,\theta)\right] = \delta\left(1-r\right)\pi\left(s-\hat{s}\right) \tag{55}$$

In either case, $s^{SE} < s^*$.

As in the case without formal performance measurement, for sufficiently large values of δ the principal can induce the first-best work design and capture the entire surplus from the relationship. However, for lower values of δ, the first-best outcome is unattainable. Under self-employment, formal performance measurement aids relational contracting by allowing for a reduction in the discretionary bonus payments. In particular, the agent's optimal deviation from the relational agreement is to implement $\hat{s} > 0$, which is increasing in r. It follows that, holding s constant, the stronger the agent's formal incentives are, the less he is tempted to deviate from the relational agreement to a less-demanding work design. This, in turn, implies that discretionary bonuses can be reduced, so that the principal has less of an incentive to deviate herself. Accordingly, the principal wishes to pay as much compensation as possible through formal performance pay r, implying a fixed transfer of zero.

To see why there are two possible cases, it is helpful to consider the principal's optimal contract in the absence of relational agreements (i.e. when $\delta = 0$). In this case, she is unable to create incentives without rewarding the agent with a strictly positive rent. As is standard in models with moral hazard and limited liability, she therefore induces a sub-optimal work design from the agent and is unable to achieve her first-best payoff using formal performance pay alone. While the ability to relationally contract can improve on this outcome, one can show that there exist cases in which it remains optimal for the principal to allow the agent to extract a rent, in order to improve the efficiency of the prescribed work design.

Proposition 7. *Suppose δ is sufficiently small such that the first-best work design cannot be implemented under either organizational form. Then:*

(i) s^E is weakly increasing in $\underline{u}$ and strictly increasing in $\underline{w}$.

(ii) s^{SE} is weakly increasing in $\underline{u}$.

Accordingly, increases in $\underline{u}$ or $\underline{w}$ make self-employment relatively more profitable.

An increase in $\underline{u}$ has two effects. First, the surplus from the relationship becomes lower, reducing the effectiveness of relational agreements. Second, the higher outside option implies that the agent's compensation must be increased. Under both organizational forms, the principal responds by increasing the proportion of output that is paid to the agent following a successful assignment. As discussed in the foregoing, more powerful outcome-contingent pay complements relational agreements, so that the two effects are countervailing.

Proposition 7 establishes that, under employment, the former effect is dominant. Intuitively, due to the over-demanding work design associated with employment, expected output is high, so that a relatively small increase in r^E is required to compensate the agent in response to a higher $\underline{u}$. Since formal performance pay is only marginally strengthened, the main impact of an increase in $\underline{u}$ is a reduction in the relationship's surplus that hinders relational agreements; in this case, the work design $s^E > s^*$ must be increased and thus becomes less efficient. In contrast, under self-employment, due to the under-demanding work design and low expected output, the increase in r^{SE} must be large relative to the case of employment; this large increase is sufficient to dominate the aforementioned 'surplus effect', so that $s^{SE} < s^*$ becomes higher and the efficiency of the work design is increased.

Overall, under either organizational form, the principal is worse off following an increase in $\underline{u}$. However, since the agent's utility is equal to his outside option in both cases, the fact that the work design becomes more efficient under self-employment but less efficient under employment immediately implies that a higher $\underline{u}$ favors the former organizational form. That is, in situations with formal performance pay and high outside options (e.g., due to competitive labor markets), one would expect self-employment to be prevalent.

Finally, under employment the principal prefers to compensate the agent through outcome-contingent pay as much as possible to strengthen the relational agreement. Following a raise in the minimum wage, the fixed payment must be increased so that r^E necessarily becomes smaller. Since this weakens the relational agreement, s must be increased to remain feasible, reducing efficiency. It follows that *any* strictly positive minimum wage has a negative impact on both profit and efficiency when formal performance pay is available, in contrast to the environment studied in Section 5.2 where small minimum wages did not affect contracting outcomes. Accordingly, minimum wages that are 'non-binding', in the sense that they are strictly less than total compensation, may nonetheless restrict the principal's ability to structure remuneration in a way that supports relational agreements, undermining efficiency. Our findings also suggest that the introduction of a minimum wage may be especially distortionary in industries where parties have access to objective measures of the worker's performance.

6 Conclusion

When procuring labor inputs, firms often face a choice between establishing employment relationships and outsourcing to self-employed workers. Which factors drive this decision is an important economic question, especially given recent developments in the labor market such as the rise of platform work. We provide a novel explanation, based on the relative efficiency of relational contracts when the legal structure imposes distinct restrictions on control over the production process under different organizational forms.

Our model yields several empirical predictions regarding differences in the characteristics of different labor relationships. We expect that, under employment, the manner in which work is performed is (i) more aligned with the firm's interests and (ii) more rigid, relative to self-employment. The model also predicts that employed workers should receive (*iii*) higher and (*iv*) less variable compensation. Finally, we expect that (*v*) employment relationships tend to arise in more favorable production environments, or when the efficient work design involves assignments being performed largely in accordance with the firm's preferences.

Throughout the paper, we discuss the extent to which these findings—and others—are consistent with existing empirical results regarding observed differences in the nature and utilization of different types of labor relationship. However, as noted by Mas and Pallais (2020), there is a "dearth of firm-level data, particularly recent ones, that can be used to better understand how firms make these decisions" (p. 650). Accordingly, we believe that further empirical studies in this direction represent an important agenda for future research, both for testing the predictions generated by our model and for developing further insights into what drives the decision between employment and self-employment.

7 Appendix

Proof of Lemma 1. We relax the problem (4)-(8). First, we replace the fallback options f_P^E and f_A^E with zero. Second, the aggregate feasibility constraint (10) is obtained by summing (7) and (8), and is hence necessarily satisfied by the solution to the problem (4)-(8). We replace (7) and (8) by (10) and, third, ignore the non-negativity constraints (5). Each of these changes relaxes the initial problem; we next characterize optimal transfers for the relaxed problem.

From the above, the relaxed problem can be expressed as maximizing (4) subject to (6) and (10). Since w does not appear in (10), it can be reduced until (6) binds. This yields:

$$w = \int_{\underline{\theta}}^{\overline{\theta}} [C(s(\theta), \theta) - b(\theta)] \cdot g(\theta) \cdot d\theta \tag{56}$$

Inserting this into (4) yields (9), which much be maximized subject to (10). Since $b(\theta)$ does not appear in the problem, it can be set arbitrarily; we set $b(\theta) = 0$ for all θ.

To conclude, we show that setting w according to (56) and $b(\theta) = 0$ for all θ satisfies the neglected constraints (5), (7) and (8). Constraint (5) is clearly satisfied. Substituting for w and $b(\theta)$, (8) becomes $b(\theta) \geq \frac{\delta}{1-\delta} f_E^A$ for all θ and (7) becomes:

$$\frac{\delta}{1-\delta}\left[\int_{\underline{\theta}}^{\overline{\theta}} [s(\theta)\pi - C(s(\theta),\theta)] \cdot g(\theta) \cdot d\theta\right] \geq \pi\left[1 - s(\theta')\right] + \frac{\delta}{1-\delta} f_P^E \quad \forall\, \theta' \in \left[\underline{\theta}, \overline{\theta}\right] \tag{57}$$

It remains to show that $f_E^A = f_E^P = 0$, which implies that $b(\theta) \geq \frac{\delta}{1-\delta} f_E^A$ becomes $b(\theta) \geq 0$ and (57) reduces to (10). Note that following a deviation by either party, contracting in the absence of a relational agreement must involve the principal setting $s = 1$ for every θ. This yields expected costs of $\int_{\underline{\theta}}^{\overline{\theta}} C(1,\theta) \cdot g(\theta) \cdot d\theta$ to the agent. If $\int_{\underline{\theta}}^{\overline{\theta}} C(s(\theta),\theta) \cdot g(\theta) \cdot d\theta < \int_{\underline{\theta}}^{\overline{\theta}} C(1,\theta) \cdot g(\theta) \cdot d\theta$, the contract's fixed wage is not sufficient to cover his expected costs, so that he prefers to reject the contract and hence $f_P^E = f_A^E = 0$. If $\int_{\underline{\theta}}^{\overline{\theta}} C(s(\theta),\theta) \cdot g(\theta) \cdot d\theta = \int_{\underline{\theta}}^{\overline{\theta}} C(1,\theta) \cdot g(\theta) \cdot d\theta$, the optimal contract must set $s = 1$ for all θ. However, this can be optimal only if the associated surplus is zero—otherwise (10) would be slack for all realizations of θ, contradicting the optimality of the contract—so that $f_P^E = f_A^E = 0$ in this case as well. $\square$

Proof of Proposition 1. Let the solution to the problem (9)-(10) be denoted by $s^{opt}(\theta)$ and define $\overline{s} = \inf_\theta\, s^{opt}(\theta) \geq 0$. From (10), it must then be the case that:

$$\frac{\delta}{1-\delta}\left[\int_{\underline{\theta}}^{\overline{\theta}} \left[s^{opt}(\theta)\pi - C(s^{opt}(\theta),\theta)\right] \cdot g(\theta) \cdot d\theta\right] \geq \pi\left[1 - \overline{s}\right] \tag{58}$$

is satisfied. Next, let us consider an alternative optimization problem, which is identical to the one under consideration, except that (10) has been replaced by the constraint:

$$s(\theta) \geq \overline{s} \quad \forall\, \theta \in \left[\underline{\theta}, \overline{\theta}\right] \tag{59}$$

Claim 1. $s^{opt}(\theta)$ also solves this alternative optimization problem.

Proof. Suppose to the contrary that $s^{opt}(\theta)$ does not solve the problem and that there exists a different solution that does; we denote this by $\xi(\theta)$. Since $s^{opt}(\theta)$ satisfies (59) (and hence all constraints), it must be that $\xi(\theta)$ yields a strictly greater value of the objective function (9) than $s^{opt}(\theta)$. However, using (58), this then implies that $\xi(\theta)$ satisfies the constraint (10), and hence all constraints of the original problem. In this case, it solves the original problem, which contradicts the fact that $s^{opt}(\theta)$ is the solution to the original problem. □

Accordingly, in order to find $s^{opt}(\theta)$, we solve the alternative problem using pointwise-optimization. The associated Lagrangian can be expressed as:[29]

$$\mathcal{L} = \int_{\underline{\theta}}^{\overline{\theta}} \left[s(\theta)\pi - C(s(\theta), \theta) \right] g(\theta) \cdot d\theta + \rho(\theta) \left(s(\theta) - \overline{s} \right) + \mu(\theta) \left(1 - s(\theta) \right) \tag{60}$$

with the associated first-order condition:

$$\left[\pi - C_s(s(\theta), \theta) \right] g(\theta) + \rho(\theta) - \mu(\theta) = 0 \tag{61}$$

We have three possible cases:[30]

1. Suppose $\rho(\theta) = 0$ and $\mu(\theta) = 0$ for some realization of θ. Then from (61):

$$\pi - C_s(s(\theta), \theta) = 0 \tag{62}$$

   so that we have $s(\theta) = s^*(\theta)$.

[29]Note that the requirement $s(\theta) \geq \overline{s}$ automatically implies that $s(\theta) \geq 0$, so that the latter requirement is dropped.

[30]The case where $\rho(\theta) > 0$ and $\mu(\theta) > 0$ implies that we have $s(\theta) = 1$ and $s(\theta) = \overline{s}$, so that $s(\theta) = 1$ over the entire work design schedule; however, as argued in Lemma 1, this can only be optimal if the associated surplus is zero.

2. Suppose $\rho(\theta) > 0$ and $\mu(\theta) = 0$ for some realization of θ. Then from (61):

$$[\pi - C_s(s(\theta), \theta)]\, g(\theta) = -\rho(\theta) < 0 \tag{63}$$

$$\Rightarrow \; \pi < C_s(s(\theta), \theta) \tag{64}$$

so that we have $s(\theta) = \overline{s}$ and $s(\theta) > s^*(\theta)$.

3. Suppose $\rho(\theta) = 0$ and $\mu(\theta) > 0$ for some realization of θ. Then from (61):

$$[\pi - C_s(s(\theta), \theta)]\, g(\theta) = \mu(\theta) > 0 \tag{65}$$

$$\Rightarrow \; \pi > C_s(s(\theta), \theta) \tag{66}$$

so that we have $s(\theta) = 1$ and $s(\theta) < s^*(\theta)$. But this is a contradiction to the assumption that $s^*(\bar{\theta}) < 1$ and we therefore rule this case out.

It follows that for all θ, we either have $s(\theta) = \overline{s}$ or $s(\theta) = s^*(\theta)$. Suppose that $s^*(\theta) \geq \overline{s}$, but we have $s(\theta) = \overline{s}$. Then from (64) and $C_{ss}(\cdot) > 0$ we have:

$$\pi < C_s(\bar{s}, \theta) \leq C_s(s^*(\theta), \theta) \tag{67}$$

which contradicts the definition of $s^*(\theta)$. Hence, if $s^*(\theta) \geq \overline{s}$, we must have $s(\theta) = s^*(\theta)$. Moreover, if $s^*(\theta) < \overline{s}$, we clearly cannot have $s(\theta) = s^*(\theta)$. We thus conclude that $s(\theta) = s^*(\theta)$ if and only if $s^*(\theta) \geq \overline{s}$. Altogether, this implies that the work design schedule can be expressed as (11).

□

Proof of Lemma 2. First, note that by the arguments in the text, $f_P^{SE} = f_A^{SE} = 0$. Second, following similar steps to the proof of Lemma 1, note that the aggregate feasibility constraint

(19) is obtained by summing (16) and (17), and is hence necessarily satisfied by the solution to the problem (13)-(17). We replace (16) and (17) by (19). Third, we ignore the non-negativity constraints (14). This relaxed problem can be expressed as maximizing (13) subject to (15) and (19). We show that there exists a contract that solves this relaxed problem and also satisfies all necessary constraints of the original problem (13)-(17).

First, note that w does not appear in (19) and hence will be reduced until (15) binds. We thus have:

$$w = \int_{\underline{\theta}}^{\overline{\theta}} [C(s(\theta), \theta) - b(\theta)] \cdot g(\theta) \cdot d\theta \tag{68}$$

Inserting this into (13) yields (18), which must be solved subject to (19), as stated in the Lemma. Since $b(\theta)$ appears in neither the objective function nor the constraint, its value is irrelevant to the problem; we thus set $b(\theta) = C(s(\theta), \theta)$ for all θ, which implies that $w = 0$. Substituting for w and $b(\theta)$ in (14), (16) and (17), it is straightforward to verify that all three constraints are satisfied. □

Proof of Proposition 2. We wish to maximize (18) subject to (19). Following the same procedure as in the case of employment, we instead solve the alternative optimization problem of maximizing (18) subject to the constraint:

$$\overline{b} \geq C(s(\theta), \theta) \tag{69}$$

for some constant $\overline{b}$ and for all θ. The Lagrangian is given by:[31]

$$\mathcal{L} = \int_{\underline{\theta}}^{\overline{\theta}} [s(\theta) \cdot \pi - C(s(\theta), \theta)] \, g(\theta) \cdot d\theta + \rho(\theta) \left(\overline{b} - C(s(\theta), \theta)\right) \tag{70}$$

[31]We ignore the constraint that $s(\theta) \in [0, 1]$, since this is trivially satisfied at the optimal solution for $\overline{b} \geq 0$.

with the associated first-order condition:

$$[\pi - C_s(s(\theta), \theta)]\, g(\theta) - \rho(\theta) C_s(s(\theta), \theta) = 0 \tag{71}$$

We have two cases. First, when $\rho(\theta) = 0$ for a particular realization of θ, from (71) we have:

$$\pi - C_s(s(\theta), \theta) = 0 \tag{72}$$

so that $s(\theta) = s^*(\theta)$. Second, when $\rho(\theta) > 0$ for a particular realization of θ, from (71) we have:

$$[\pi - C_s(s(\theta), \theta)]\, g(\theta) = \rho(\theta) C_s(s(\theta), \theta) > 0 \tag{73}$$

$$\Rightarrow \;\; \pi > C_s(s(\theta), \theta) \tag{74}$$

so that $s(\theta)$ satisfies $C(s(\theta), \theta) = \overline{b}$ and therefore $s(\theta) = \tilde{s}(\theta, \overline{b})$. Moreover, for all such $s(\theta)$, we have $s(\theta) < s^*(\theta)$ from (74).

Altogether, for each realization of θ we have either $s(\theta) = s^*(\theta)$ or $s(\theta) = \tilde{s}(\theta, \overline{b})$. Suppose that we have $s(\theta) = \tilde{s}(\theta, \overline{b})$, but $C(s^*(\theta), \theta) \leq \overline{b}$. But since $C_s(\cdot) > 0$, this implies that we have $\tilde{s}(\theta, \overline{b}) \geq s^*(\theta)$, contradicting (74). Hence, if $s(\theta) = \tilde{s}(\theta, \overline{b})$, it must be that $C(s^*(\theta), \theta) > \overline{b}$. Moreover, if $C(s^*(\theta), \theta) > \overline{b}$, then clearly $s(\theta) = s^*(\theta)$ is not feasible and hence we have $s(\theta) = \tilde{s}(\theta, \overline{b})$. Altogether, we conclude that $s(\theta) = \tilde{s}(\theta, \overline{b})$ if and only if $C(s^*(\theta), \theta) > \overline{b}$; hence, the optimal contract takes the form outlined in the main text. □

Proof of Proposition 3. As $\pi \to 0$, the first-best surplus, S^*, also tends to zero. In addition, the first-best work design $s^*(\theta)$ associated with each θ becomes small. The left-hand side of (10) is bounded by $\frac{\delta}{1-\delta} S^*$. Accordingly, as $\pi \to 0$, a work design schedule can only be implemented under employment if $s(\theta)$ is close to 1 for all θ. However, such schedules clearly cannot yield a non-negative profit for π sufficiently small, as they impose high costs for a small return.

Under self-employment, similar arguments imply that, as $\pi \to 0$, a schedule can only be implemented if $s(\theta)$ is close to 0 for all θ. However, $C(0,\theta) = C_s(0,\theta) = 0$ implies that for any θ, there always exist strictly positive values of s that are profitable. Accordingly, as π becomes sufficiently small, self-employment must become strictly optimal. $\square$

Proof of Proposition 4. Follows from the arguments in the text. $\square$

Proof of Proposition 5. We have $\int_{\underline{\theta}}^{\overline{\theta}} W^E(\theta) \cdot g(\theta) \cdot d\theta = \int_{\underline{\theta}}^{\overline{\theta}} C(s^E(\theta),\theta) \cdot g(\theta) \cdot d\theta$ and $\int_{\underline{\theta}}^{\overline{\theta}} W^{SE}(\theta) \cdot g(\theta) \cdot d\theta = \int_{\underline{\theta}}^{\overline{\theta}} C(s^{SE}(\theta),\theta) \cdot g(\theta) \cdot d\theta$; (ii) then follows immediately from $s^{SE}(\theta) \leq s^*(\theta) \leq s^E(\theta)$ and $C_s(s,\theta) > 0$; (i) follows from $\mathrm{Var}\left[W^E(\theta)\right] = 0$ with $\mathrm{Var}\left[W^{SE}(\theta)\right] > 0$ for sufficiently large δ. $\square$

Proof of Lemma 3. Application of the implicit function theorem to (29) yields:

$$\frac{ds^E}{d\tau} = \frac{[s+\delta-1]\,\pi}{(1-\tau)\,\pi - \delta C_s(s,\theta)} > 0 \tag{75}$$

The left-hand side of (29) is strictly concave and continuous in s, and attains a maximum when $(1-\tau)\,\pi = \delta C_s(s,\theta)$. Accordingly, (29) can be satisfied at (at most) two points; since we are restricting attention to cases where the feasibility constraint binds, the smaller of these is our solution, which implies an upward slope so that the denominator of the above term is strictly positive. In addition, note that $s+\delta-1 > 0$ at the solution of (29).

Application of the implicit function theorem to (37) yields:

$$\frac{ds^{SE}}{d\tau} = \frac{\delta s \pi}{\delta(1-\tau)\pi - C_s(s,\theta)} < 0 \tag{76}$$

The denominator of this term is strictly negative since the LHS of (37) must be downward sloping at the solution: it is (i) strictly concave in s, (ii) negative for the s that satisfies $\pi\,(1-\tau) = C_s(s,\theta)$, else the feasibility constraint would not bind and (iii) attains a maximum when $\delta(1-\tau)\pi = C_s(s,\theta)$. Accordingly, if the constraint is satisfied for some s, it must be that the principal's optimal choice is the highest s such that the constraint is satisfied, which implies a downward slope at the solution. $\square$

Proof of Proposition 6. From (26), the principal's objective function under employment is equal to zero when $s = 0$. For sufficiently large τ, there must exist another point at which profit is zero, for some $\tilde{s}_\tau$ implicitly defined by $(1-\tau)\tilde{s}_\tau\pi - C(\tilde{s}_\tau, \theta) = 0$. Clearly, $\tilde{s}_\tau$ is decreasing in τ and tends to 0 as $\tau \to 1$. Concavity of (26) then implies that the principal can only achieve a strictly positive profit by implementing $s < \tilde{s}_\tau$. Note that the surplus created by any such s also tends to zero as $\tau \to 1$. It follows from (27) that there exists a $\hat{\tau} < 1$ such that, for $\tau \geq \hat{\tau}$, the principal cannot earn a strictly positive profit under employment; i.e. the combination of a small surplus and a small s eventually becomes incompatible with the principal's feasibility constraint, for any $\delta \in (0, 1)$.

Next, let us rewrite the feasibility constraint under self-employment, (36), as:

$$\delta s\pi \geq \frac{C(s, \theta)}{(1-\tau)} \tag{77}$$

At the point $s = 0$, a marginal increase in s has a zero impact on the RHS (since $C_s(0, \theta) = 0$) and a positive impact on the LHS. Accordingly, sufficiently small values of s are feasible and lead to a strictly positive profit under self-employment for **any** $\tau < 1$ and $\delta \in (0, 1)$. It follows that for all $\tau \geq \hat{\tau}$, self-employment is optimal. $\square$

Proof of Lemma 4. Following the analysis in Section 4, the principal's problem under employment in the degenerate case becomes:

$$\max_{w \geq \underline{w}, s \in [0,1]} \quad s\pi - w \tag{78}$$

subject to:

$$w - C(s,\theta) \geq 0 \tag{79}$$

$$s\pi - w + \frac{\delta}{1-\delta}\left[s\pi - w\right] \geq \pi - w \tag{80}$$

As before, there can be no benefit to the use of relational bonus payments. First, suppose that $\underline{w} > C(1,\theta)$. In this case, (79) does not bind for any s; accordingly, in the case that $\pi - \underline{w} \geq 0$, so that the relationship is still profitable, the principal sets $s = 1$ and $w = \underline{w}$, which implies that the agent extracts a rent. This contract clearly satisfies (80).

In all other cases, (79) binds and the problem can be rewritten as:

$$\max_{s\in[0,1]} \;\; s\pi - C(s,\theta) \tag{81}$$

subject to:

$$C(s,\theta) \geq \underline{w} \tag{82}$$

$$s\pi + \frac{\delta}{1-\delta}\left[s\pi - C(s,\theta)\right] \geq \pi \tag{83}$$

The solution depends on which constraint binds. Clearly, for low values of $\underline{w}$, the minimum wage constraint does not bind and the analysis is identical to that of Section 4. Otherwise, for sufficiently large $\underline{w}$, (82) holds with equality so that s^E solves:

$$C(s,\theta) = \underline{w} \tag{84}$$

and is strictly increasing in $\underline{w}$. □

Proof of Lemma 5. In the absence of (39), the principal would maximize (38) and satisfy all other constraints by setting $w = \underline{w}$, $r = 0$ and $s = 1$. However, the restriction $\underline{w} < \underline{u} + C(1, \theta)$ then implies that (39) would not be satisfied. Hence, (39) binds at the optimum and we have $w = \underline{u} + C(s, \theta) - sr\pi$ so that the principal's problem can be rewritten as:

$$\max_{r \in [0,1], s \in [0,1]} \quad s\pi - C(s, \theta) - \underline{u} \tag{85}$$

subject to:

$$\underline{u} + C(s, \theta) - sr\pi \geq \underline{w} \tag{86}$$

$$\frac{\delta}{1-\delta}\left[s\pi - C(s, \theta) - \underline{u}\right] \geq (1-r)\,\pi\,[1-s] \tag{87}$$

Suppose that (86) does not bind at the optimal solution; the principal could then set $r = 1$ and implement s^* to maximize the objective function. However, this would imply that $-\left[s^*\pi - C(s^*, \theta) - \underline{u}\right] \geq \underline{w} \geq 0$, which is clearly a contradiction. Hence, (86) binds, yielding (41). For δ sufficiently large, (87) is slack and the first-best is implemented. Otherwise, this constraint binds; substituting for r then yields (42). Suppose that the resulting solution is $s' < s^*$, with associated payoff to the principal of:

$$\Phi := s'\pi - C(s', \theta) - \underline{u} > 0 \tag{88}$$

Since $s\pi - C(s, \theta) - \underline{u}$ is a strictly concave function which attains a maximum at $s = s^*$, there must also exist an $s'' > s^*$ such that we have:[32]

$$s''\pi - C(s'', \theta) - \underline{u} = \Phi \tag{89}$$

[32] If such an s'' does not exist, then it must be that $\pi - C(1, \theta) > s'\pi - C(s', \theta)$. Since setting $s = 1$ is always feasible for the principal (so long as the associated profit is non-negative), this contradicts the assertion that s' solves the principal's problem.

Moreover, since the optimal solution satisfies (42), we have:

$$\delta\left[s'\pi - C(s',\theta) - \underline{u}\right] - (1-\delta)\left(s'\pi - \underline{u} + \underline{w} - C(s',\theta)\right)\left(\frac{1-s'}{s'}\right) = 0 \tag{90}$$

or equivalently:

$$\delta\Phi - (1-\delta)\left(\Phi + \underline{w}\right)\left(\frac{1-s'}{s'}\right) = 0 \tag{91}$$

However, evaluating the feasibility constraint at $s = s''$ yields:

$$\delta\Phi - (1-\delta)\left(\Phi + \underline{w}\right)\left(\frac{1-s''}{s''}\right) > 0 \tag{92}$$

which is strictly positive since:

$$\frac{1-s'}{s'} > \frac{1-s''}{s''} \iff s'' > s' \tag{93}$$

Accordingly, the principal could set $s = s''$ and achieve the same payoff, but with a slack feasibility constraint; there hence exists an $\epsilon > 0$ such that implementing $s = s'' - \epsilon$ both satisfies the feasibility constraint and increases the principal's payoff, contradicting the assertion that s' is the solution to the principal's problem. Accordingly, we must have $s^E \geq s^*$. □

Proof of Lemma 6. Simplifying and cancelling terms, we rewrite the problem as:

$$\max_{w,b,r\in[0,1],s\in[0,1]} \quad s\,(1-r)\,\pi - b - w \tag{94}$$

subject to:

$$w, w + b \geq 0 \tag{95}$$

$$w + sr\pi + b - C(s, \theta) - \underline{u} \geq 0 \tag{96}$$

$$[sr\pi + b - C(s, \theta)] - [\hat{s}r\pi - C(\hat{s}, \theta)] \geq 0 \tag{97}$$

$$\delta\,(1 - r)\,\pi\,(s - \hat{s}) - b \geq 0 \tag{98}$$

The associated Lagrangian is:

$$\begin{aligned} L = s\,(1-r)\,\pi - b - w + \lambda_1 w + \lambda_2\,(w + sr\pi + b - C(s,\theta) - \underline{u}) \\ + \lambda_3\,([sr\pi + b - C(s,\theta)] - [\hat{s}r\pi - C(\hat{s},\theta)]) + \lambda_4\,(\delta\,(1-r)\,\pi\,(s-\hat{s}) - b) \end{aligned} \tag{99}$$

where we have ignored the constraint $w + b \geq 0$. The corresponding FOCs can be expressed as:

$$\lambda_1 + \lambda_2 = 1 \tag{100}$$

$$\lambda_2 + \lambda_3 = 1 + \lambda_4 \tag{101}$$

$$(1 + \lambda_4\delta)\,(1 - r)\,\pi + (\lambda_2 + \lambda_3)\,(r\pi - C_s(s, \theta)) = 0 \tag{102}$$

$$\lambda_3\left(s\pi - \hat{s}\pi - \frac{\partial \hat{s}}{\partial r}\,[r\pi - C_s(\hat{s}, \theta)]\right) - \lambda_4\left(\delta\pi\,(s - \hat{s}) + \delta\,(1 - r)\,\pi\frac{\partial \hat{s}}{\partial r}\right) = (1 - \lambda_2)\,s\pi \tag{103}$$

First, consider the case where (96) is slack at the optimal solution and $\lambda_2 = 0$. From (100) and (101), this immediately implies $\lambda_1 = 1 > 0$ and $\lambda_3 = 1 + \lambda_4 > 0$. Since $r\pi - C_s(\hat{s}, \theta) = 0$ by the definition of $\hat{s}$, (102) and (103) can then be rewritten as:

$$(1 + \lambda_4 \delta)(1 - r)\pi + (1 + \lambda_4)(r\pi - C_s(s, \theta)) = 0 \tag{104}$$

$$\lambda_4 \left[\pi (s - \hat{s})(1 - \delta) - \delta (1 - r) \pi \frac{\partial \hat{s}}{\partial r} \right] = \hat{s}\pi \tag{105}$$

where (105) implies that $\lambda_4 > 0$. Accordingly, when (96) is slack, all remaining constraints bind. This implies that $w = 0$ and, from (98), $b = \delta (1 - r) \pi (s - \hat{s})$. (97) then becomes (55). Combining (104) and (105) to cancel λ_4 yields (54).

Second, consider the case where (96) binds at the optimal solution, so that $\lambda_2 > 0$. Combining (101) and (102) yields (104). If $\lambda_4 = 0$, then we have $\pi - C_s(s, \theta) = 0$ and the first-best is implemented. Otherwise, if $\lambda_4 > 0$, combining (100) and (101) yields $\lambda_1 + \lambda_4 = \lambda_3$, so that $\lambda_3 > 0$. In this case, (96)-(98) bind, so that we have $w = C(s, \theta) + \underline{u} - sr\pi - b$, $b = \delta (1 - r) \pi (s - \hat{s})$ and the principal's problem can be rewritten as:

$$\max_{r \in [0,1], s \in [0,1]} \quad s\pi - C(s, \theta) - \underline{u} \tag{106}$$

subject to:

$$C(s, \theta) + \underline{u} - sr\pi - \delta (1 - r) \pi (s - \hat{s}) \geq 0 \tag{107}$$

$$\delta (1 - r) \pi (s - \hat{s}) + [sr\pi - C(s, \theta)] - [\hat{s}r\pi - C(\hat{s}, \theta)] = 0 \tag{108}$$

Suppose that (107) is slack at the optimal solution. In this case, the principal can set $r = 1$ and $s = s^*$, satisfying (108)—since $s^* = \hat{s}$ for $r = 1$—and maximizing the surplus (106). However, a combination of $w \geq 0$ and $r = 1$ clearly cannot yield a strictly positive profit to the principal; accordingly, (107) binds at the optimal solution, so that the pair (r, s) is defined by (52) and (53).

To conclude the proof, combining (101) and (102) yields:

$$\pi - C_s(s,\theta) + \lambda_4 \left[\delta\left(1-r\right)\pi + r\pi - C_s(s,\theta)\right] = 0 \tag{109}$$

If $\lambda_4 = 0$, the first-best is implemented. In all other cases, $\lambda_4 > 0$. Suppose that $\pi \leq C_s(s,\theta)$. For $\lambda_4 > 0$, (109) then implies that $\delta\left(1-r\right)\pi + r\pi \geq C_s(s,\theta)$ and hence $\delta\left(1-r\right)\pi + r\pi \geq \pi$. However, this is not possible for $\delta \in (0,1)$ unless $r = 1$, which clearly cannot be profitable. Accordingly, $\pi > C_s(s,\theta)$ and the work design is under-demanding. □

Proof of Proposition 7. First, consider the case of employment. Since $s\pi - C(s,\theta) - \underline{u}$ is strictly concave in s, when the first-best work design is unattainable, the principal chooses the lowest $s > s^*$ that satisfies the feasibility constraint. Since the LHS of (42) is continuous in s and (42) is not satisfied at $s = s^*$, this immediately implies that the LHS of (42) is upward sloping in s at the solution. Using this fact and applying the implicit function theorem to (42) then yields:

$$\frac{ds}{d\underline{u}} = \frac{\delta - (1-\delta)\left(\frac{1-s}{s}\right)}{(+ve)} \geq 0 \tag{110}$$

and:

$$\frac{ds}{d\underline{w}} = \frac{(1-\delta)\left(\frac{1-s}{s}\right)}{(+ve)} > 0 \tag{111}$$

where $\delta - (1-\delta)\left(\frac{1-s}{s}\right) \geq 0$ follows from (42). Next, consider self-employment. When the pair (r, s) is defined by (54)-(55), the system is independent of $\underline{u}$. When (r, s) is defined by (52)-(53), combining these two equations yields:

$$\hat{s}r\pi - C(\hat{s},\theta) - \underline{u} = 0 \tag{112}$$

Application of the implicit function theorem to the system formed by (52) and (112) yields:

$$\frac{\partial r}{\partial \underline{u}} = \frac{1}{\hat{s}\pi} > 0 \tag{113}$$

$$\frac{\partial s}{\partial \underline{u}} = \frac{\delta \hat{s}_r \left(\pi - C_s(\hat{s},\theta)\right) - (1-\delta)\pi(s-\hat{s})}{\hat{s}\pi\left((1-\delta)C_s(\hat{s},\theta) + \delta\pi - C_s(s,\theta)\right)} > 0 \tag{114}$$

where (114) is strictly positive by the following claim.

Claim 2. Let $z \in [0,1]$ be a constant. Then:

$$\frac{z\hat{s}_r \left(\pi - C_s(\hat{s},\theta)\right) - (1-z)\pi(s-\hat{s})}{\hat{s}\pi\left((1-z)C_s(\hat{s},\theta) + z\pi - C_s(s,\theta)\right)} \tag{115}$$

is strictly positive for any z.

Proof of Claim 2. For $z = 0$, (115) becomes:

$$\frac{\pi(s-\hat{s})}{\hat{s}\pi\left(C_s(s,\theta) - C_s(\hat{s},\theta)\right)} > 0 \tag{116}$$

which is strictly positive since $s > \hat{s}$. For $z = 1$, we have:

$$\frac{\hat{s}_r \left(\pi - C_s(\hat{s},\theta)\right)}{\hat{s}\pi\left(\pi - C_s(s,\theta)\right)} > 0 \tag{117}$$

since $s^* > s > \hat{s}$. Finally, differentiation of (115) with respect to z yields:

$$\hat{s}\pi\frac{\pi(s-\hat{s})\left[\pi - C_s(s,\theta)\right] + \hat{s}_r\left(\pi - C_s(\hat{s},\theta)\right)\left[C_s(\hat{s},\theta) - C_s(s,\theta)\right]}{\left[\hat{s}\pi\left((1-z)C_s(\hat{s},\theta) + z\pi - C_s(s,\theta)\right)\right]^2} \tag{118}$$

Clearly, the sign of (118) is determined by its numerator, which is independent of z. Accordingly, (118) has the same sign for all $z \in [0,1]$, so that (115) is bounded by (116) and (117), two strictly positive numbers; accordingly, (115) is strictly positive. □

□

References

Abraham, K. G. and Taylor, S. K. (1996), 'Firms' use of outside contractors: Theory and evidence', *Journal of Labor Economics* **14**(3), 394–424.

Alfaro, L., Bloom, N., Conconi, P., Fadinger, H., Legros, P., Newman, A. F., Sadun, R. and Van Reenen, J. (2023), 'Come together: Firm boundaries and delegation', *Journal of the European Economic Association, Forthcoming* .

Baker, G., Gibbons, R. and Murphy, K. J. (1994), 'Subjective performance measures in optimal incentive contracts', *The Quarterly Journal of Economics* **109**(4), 1125–1156.

Baker, G., Gibbons, R. and Murphy, K. J. (1999), 'Informal authority in organizations', *Journal of Law, Economics, and Organization* **15**(1), 56–73.

Baker, G., Gibbons, R. S. and Murphy, K. J. (2011), Relational adaptation. Working paper, Harvard Business School.

Baker, G. P., Gibbons, R. and Murphy, K. J. (2002), 'Relational contracts and the theory of the firm', *The Quarterly Journal of Economics* **117**(1), 39–84.

Boeri, T., Giupponi, G., Krueger, A. B. and Machin, S. (2020), 'Solo self-employment and alternative work arrangements: A cross-country perspective on the changing composition of jobs', *Journal of Economic Perspectives* **34**(1), 170–195.

Coase, R. H. (1937), 'The nature of the firm', *Economica* **4**(16), 386–405.

Demougin, D. and Fabel, O. (2019), 'The nexus between contract duration and the use of formal and informal incentive pay', *Labour* **33**(3), 351–370.

Eurofound (2018), Employment and working conditions of selected types of platform work. Publications Office of the European Union.

Faberman, R. J., Mueller, A. I., Şahin, A. and Topa, G. (2022), 'Job search behavior among the employed and non-employed', *Econometrica* **90**(4), 1743–1779.

Ganserer, A., Gregory, T. and Zierahn, U. (2022), 'Minimum wages and the rise in solo self-employment', *ZEW-Centre for European Economic Research Discussion Paper* (22-024).

Gentry, W. M. and Hubbard, R. G. (2000), 'Tax policy and entrepreneurial entry', *American Economic Review* **90**(2), 283–287.

Glasner, B. (2023), 'The minimum wage, self-employment, and the online gig economy', *Journal of Labor Economics* **41**(1), 103–127.

Grossman, S. J. and Hart, O. D. (1986), 'The costs and benefits of ownership: A theory of vertical and lateral integration', *Journal of Political Economy* **94**(4), 691–719.

Hart, O. and Moore, J. (1990), 'Property rights and the nature of the firm', *Journal of Political Economy* **98**(6), 1119–1158.

Hart, O. and Moore, J. (2008), 'Contracts as reference points', *The Quarterly Journal of Economics* **123**(1), 1–48.

Hurst, E. and Pugsley, B. W. (2011), What do small businesses do? Brookings Papers on Economic Activity.

Levin, J. (2003), 'Relational incentive contracts', *American Economic Review* **93**(3), 835–857.

Levin, J. and Tadelis, S. (2010), 'Contracting for government services: Theory and evidence from US cities', *The Journal of Industrial Economics* **58**(3), 507–541.

MacLeod, W. B. and Malcomson, J. M. (1989), 'Implicit contracts, incentive compatibility, and involuntary unemployment', *Econometrica* **57**(2), 447–480.

Marino, A. M., Matsusaka, J. G. and Zábojník, J. (2010), 'Disobedience and authority', *The Journal of Law, Economics, & Organization* **26**(3), 427–459.

Mas, A. and Pallais, A. (2020), 'Alternative work arrangements', *Annual Review of Economics* **12**, 631–658.

Muhl, C. J. (2002), 'What is an employee? The answer depends on the federal law', *Monthly Labor Review.* **125**, 3–11.

Oyer, P. (2020), 'The gig economy', *IZA World of Labor* **471**.

Pearce, D. G. and Stacchetti, E. (1998), 'The interaction of implicit and explicit contracts in repeated agency', *Games and Economic Behavior* **23**(1), 75–96.

Pohlan, L. (2019), 'Unemployment and social exclusion', *Journal of Economic Behavior & Organization* **164**, 273–299.

Raith, M. (2023), Employee or contractor? A theory of the labor boundaries of the firm. Working paper, University of Rochester.

Rantakari, H. (2023), 'Simon says? Equilibrium obedience and the limits of authority', *Journal of Law, Economics, & Organization, Forthcoming* .

Schmidt, K. M. and Schnitzer, M. (1995), 'The interaction of explicit and implicit contracts', *Economics Letters* **48**(2), 193–199.

Simon, H. A. (1951), 'A formal theory of the employment relationship', *Econometrica* **19**(3), 293–305.

Slaughter, S. and Ang, S. (1996), 'Employment outsourcing in information systems', *Communications of the ACM* **39**(7), 47–54.

Van den Steen, E. (2010), 'Interpersonal authority in a theory of the firm', *American Economic Review* **100**(1), 466–490.

Wernerfelt, B. (1997), 'On the nature and scope of the firm: An adjustment-cost theory', *The Journal of Business* **70**(4), 489–514.

Wernerfelt, B. (2004), 'Governance of adjustments', *The Journal of Business* **77**(S2), S3–S24.

Wernerfelt, B. (2015), 'The comparative advantages of firms, markets and contracts: A unified theory', *Economica* **82**(326), 350–367.