Contract Manufacturing Relationships

Can Urgun[*,†]

August 2017

Abstract

Contract manufacturing enables intellectual property holders to enjoy scale economies, reduce labor costs and free up capital. However, in many scenarios contract manufacturing is a double-edged sword, rife with entrenchments, threats of predation or hold up. I explore these contract manufacturing problems in a non-recursive relational contract setting. These non-recursivities appear in at least two scenarios: First, a setting where there is learning by doing, but the accumulated expertise can also be used by the agent to compete against the principal. Second, a setting where there are multiple potential producers, but these contract manufacturers have prior entrenchments effecting their costs and can hold up the client. The analysis of these relations requires a novel methodological approach. A key contribution is that despite the non-recursive nature of these relationships, in both settings the principal optimal contract is characterized by a simple index rule, which does not depend on history or other agents.

JEL-Classification: C73, D86, L14, L21

[*] I am grateful to Alvaro Sandroni, Niko Matouschek, Bruno Strulovici, Dan Barron, Mike Powell, Ehud Kalai and Jin Li who provided support, guidance and helpful suggestions. I am also thankful to seminar participants at Northwestern, Chicago, Princeton, UCSD, London Business School and Indiana

[†] Princeton University, e-mail: curgun@princeton.edu

1 Introduction

Firms often maintain relationships with trading partners to outsource production. The main form of outsourcing in many industries is a contract manufacturing agreement. However, these agreements are seldom complete and are often backed with informal promises. Outsourcing allows a firm to concentrate on enhancing products by focusing on R&D, marketing and design, while enjoying the cost advantages brought in by the expertise of contract manufacturers (CM).

Global contract manufacturing had an expected volume of $515 billion in electronics industry and $40.7 billion in pharmaceutial industry in 2015. It is expected to grow even more with projected annual growth rates of 8.6% and 6.4% for the respective industries (Rajaram 2015, Pandya and Shah 2013). Moreover, contract manufacturing agreements are not limited to electronics and pharmaceuticals; they are used in a broad range of other industries including automotive, food and beverages (Tully 1994). In a contract manufacturing agreement, a client engages a contractor to manufacture a product in exchange for a negotiated fee.[1] If a product is novel and complex, a client will gravitate towards a single source contract manufacturing agreement as switching between CMs becomes costly. On the other hand, as products commodify, clients gain a wide choice of interchangeable CMs (Arrunada and Vázquez 2006).

Despite the cost advantages, contract manufacturing entails some inescapable hazards. In many contract manufacturing agreements parties soon find themselves immersed in a "melodrama replete with promiscuity, infidelity, and betrayal" (Arrunada and Vázquez 2006). On one hand, if the product is novel and sole sourced, then the CM is in a prime position to compete or even overtake the client. "Adding insult to injury, if the client had not given its business to the traitorous contract manufacturer, the CM's knowledge might have remained sufficiently meager to prevent it from entering its patron's market".(Arrunada and Vázquez 2006). Indeed, Intel, Cisco Systems and Alcatel retain some plants despite outsourcing most of their production. These firms juggle their production between the CM and their own inefficient, in-house production capabilities in order to curb the learning and efficiency of the CM. On the other hand, if the product is commodified

[1]For ease of referral, I address the principal/client as female and the agents/contractors as male.

and in a mature market then another problem arises. In case of commodified products as McCoy (2003) notes, when a client approaches a contractor she may discover that he is entrenched with other clients. The relationships become necessarily intertwined despite their bilateral nature. Contractors manage these diverse relationships by trying to keep their facilities running at $70-80\%$ capacity and they meet extra demands by working overtime (Tully 1994). Thus a contractor who is already employed may have to over-utilize his assets if he works for another client, which increases his contracting costs.

The main difference of contract manufacturing problems from other relational contracting problems is that employment has an impact on future costs. I explore two main settings that have this feature: sole sourcing and multiple sourcing. The sole sourcing model is motivated by the electronics industry, where firms retain some in-house building capacity to limit the efficiency gains of their contract manufacturer in order to prevent the contract manufacturer from competing against them. In the sole sourcing model, employment leads to learning by doing. If the agent works for the principal, he gets more efficient and hence becomes a potential threat to the principal, if the principal opts for in-house production the agent slowly becomes less efficient. Multiple sourcing model is motivated by the pharmaceutical industry. The product is more commodified and an intellectual property holder wishes to produce a drug via a facility that satisfies some requirements (e.g. FDA regulations). I model this interaction as a relational contracting setting with a single principal and multiple agents. The principal wishes to produce a good but lacks manufacturing capacity. Thus, she outsources production to at most one agent every period. The agents are sometimes entrenched with other clients and repeated employment forces the agents to over-utilize their assets, increasing their costs. If an agent who has other entrenchments is not employed by the principal, than he has an opportunity to catch-up to his other obligations, rest, and decrease future costs.

Despite the complexities in these economies, the principal optimal employment schedule is achieved by a simple index rule and an accompanying payment rule. The index of an agent depends only on the current cost of that agent and the payment is tied to the indices.

The simplicity of this policy reveals striking characteristics of the optimal contract. When making an employment decision, the principal could potentially rely on many factors. These include the entire history of relationships, all the informal promises she made, or even calendar time. However, the index does not depend on these factors, it simply depends on the current cost of an agent and the mechanics i.e. the underlying law of motion, governing the cost.

The indices do not depend on calendar time which is particularly significant in the sole sourcing model, because the length of the relationship does not effect the decisions of the principal. The level of efficiency, no matter how long it took to reach, is the deciding factor.

The indices do not depend on history which has important ramifications in the multiple sourcing model. In particular, the indices do not depend on whether an agent has ever worked for the principal or not. This rules out the insider-outsider phenomena, where preferential treatment is given to agents with whom the principal has worked before. Loyalty in the form of keeping promises occurs in the optimal contract, but loyalty in the form of preferential treatment does not.

The analysis of employment impacting future costs requires novel techniques. In the existing literature, dependence of future costs on employment is usually abstracted away because it breaks any inherent recursion. The absence of recursion turns the game into a *reducible* stochastic game. Unlike *irreducible* stochastic games, there are no Folk theorems or equivalent applicable techniques for characterizing the payoff space and the set of equilibria.[2] Hence, a new methodology is required for this paper. I show that the lack of recursion in this game can be tackled by index policies and the principal's problem is a relaxed version of a non-standard bandit problem where I build upon the Whittle (1988) index. Despite the various incentive frictions and complex relationships, the indices in this paper share some of the characteristics of the Gittins (1979) index, which was celebrated for its surprising simplicity.

The index policy approach has several advantages. First, although I focus on economies where costs are the only phenomena that is dependent on employment, the methodology is broadly applicable to other scenarios where recursion might be broken by actions available to the parties such as persistent capital investments, liquidity constraints that are tied to performance, or reputation build up in different markets. In fact, the methodology can be further generalized by using multi-mode bandit indices to capture different effort levels. Second, the optimal policy will always be time consistent.

[2]A stochastic game is irreducible if no player has a strategy that can reduce the payoff relevant stochastic process.

This is in contrast with the pioneering work of Board (2011) where the optimal contract had a discrepancy between insiders and outsiders, which led to time inconsistencies even with recursion. Finally, the index policy is fully described and the indices are identified in closed form. For example in the learning by doing model the closed form enables comparative statics about the speed of learning and profits.

This paper is organized as follows. After a brief literature review section 2 investigates the sole sourcing model. Section 3 investigates the commodified product case with entrenchments. Section 4 gives a brief overview of the methodology. Finally, section 5 concludes. All the proofs are relegated to the appendix.

1.1 Related Literature

This paper builds on a large number of relational contracting papers. This is a vast literature that I do not survey here. Malcomson et al. (2010) provides an excellent survey.

The sole sourced contract manufacturing model explores learning by doing in a dynamic setting. A comprehensive survey for learning by doing is Thompson (2010), as noted, dynamic models quickly become intractable, hence much of the existing analysis has been made in restricted settings. The closest models are Plambeck and Taylor (2005) and Gray, Tomlin, and Roth (2009) as they explore learning by doing in a contract manufacturing setting, albeit limited to two periods.

The classic references to the entrenchment model are Levin (2002), Board (2011), and Andrews and Barron (2013) since they feature multiple agents in a relational contracting setting.

The critical problem for the principal in both settings is to find an optimal employment schedule even though there is no inherent recursion in the game. Bandit problems are also scheduling problems which need not recur thus I build upon techniques in the bandit literature. From a methodological perspective, approaching the principal's problem as a bandit problem is different from canonical papers in relational incentive contracting, such as Levin (2003), Baker, Gibbons, and Murphy (2002), and Malcomson et al. (2010). Most of the literature utilizes inherent recursion to take advantage of various Folk theorems, which enables a strong characterization of the payoff space and from this space pick the principal optimal one. The main advantage of a bandit approach is that it allows for an easily implementable policy when there is no recursion while still delivering the principal optimal behavior.

Within the bandit literature this paper builds upon restless bandit problems. Gittins, Glazebrook, and Weber (2011) provides an excellent treatment of this literature, and Nino-Mora et al. (2001), Glazebrook, Nino-Mora, and Ansell (2002), Nino-Mora (2002), Glazebrook, Ruiz-Hernandez, and Kirkbride (2006), Glazebrook, Hodge, Kirkbride, et al. (2013) are notable contributions. Restless bandits are bandit problems where even the arms that are not operated continue to give rewards and to change states, albeit at different rates. The pioneering work in that literature is Whittle (1988), where he derives a heuristic index based on a Lagrangian relaxation of the undiscounted problem. Papadimitriou and Tsitsiklis (1999) showed that general restless bandits are intractable and even indexability of the problem is hard to ascertain. However, here I show that in this special case of bi-directional restless bandits the intractability can be bypassed via using an equivalence of policies to calculate indices in closed form. In order to do this, I build upon the work of Glazebrook, Ruiz-Hernandez, and Kirkbride (2006). Finally I build upon Jacko (2011) to show that the index policy is optimal.

Finally, as a generalization of bandit problems this paper also utilizes some general existence results on Markov decision problems. This is also a vast literature that I do not survey here, Blackwell (1965) is an important pioneering work, and Puterman (2014) provides a remarkable treatment of the literature.

2 Single Sourcing and Learning by Doing

In this section, I investigate the optimal relational contract between a single powerful contract manufacturer and a single client. In situations where the client's product is novel and complex, it becomes nearly impossible for the client to replace a CM. In such cases "an ambitious, upstart CM may decide to build its own brand and forge its own relationships". (Arrunada and Vázquez 2006)

Formally, there are two players, 1 principal (she) and 1 agent (he). Time is discrete and the horizon is infinite, and both players discount future payoffs with a discount factor $\delta \in (0, 1)$. At each period t the following events unfold:

- The agent's production cost c^t is realized and becomes publicly known.
- The principal makes an offer to the agent that specifies a production

source that is either in-house or the agent, and a payment to the agent.

- The agent decides whether to accept the offer not. If the agent is accepts, production and payments happen as contractually specified, in addition if the source of production is the agent, the principal also covers the cost c^t. Alternatively the agent can reject the contract and enter into competition with the principal by paying a fixed fee F. If the agent enters into competition the principal is forced to produce in-house from that period onward.

A history at period t consists of all the employment decisions and all the payments made as well as all the costs, $h^t = ((I^0, p^0, c^0), (I^1, p^1, c^1), \ldots, (I^t, p^t, c^t))$ up to period t. Given that the strategic interaction ends once the agent enters I implicitly assume that the agent has not entered until period t while considering a period t history. The set of all histories at period $t, t+1, \ldots$ generate a growing sequence of σ-algebras, i.e. a filtration, $\{\mathcal{F}_t\}$. The probability triple (i.e. the filtered probability space) is given by $(\Omega, \{\mathcal{F}_t\}, P)$ where, Ω is the set of all histories P is the probability measure over Ω and $\{\mathcal{F}_t\}$ is the natural filtration. Let E denote the expectation operator associated with P. Throughout the analysis I focus on pure strategies. The principal's strategy denoted as π is a $\mathcal{F}_t$ measurable plan of employment decisions and payments $\{I^t, p^t\}_{t\in\mathbb{N}}$. The agent's entry decision is an extended stopping time τ_e on $(\Omega, \{\mathcal{F}_t\}, P)$, $\tau_e : \Omega \to \mathbb{N}\cup\{\infty\}$ such that the event $\tau_e \leq t$ is $\mathcal{F}_t$ measurable. A tuple (π, τ_e) is called a relational contract.

Throughout I assume there is open book accounting and the costs of production is borne by the agent but paid in full by the principal. In this section only I abstract away from explicitly modeling potential frictions such as hold up problems which are prevalent in contract manufacturing. However hold up will be considered more explicitly in the following section. In this section such cases can readily be encompassed into the model. The fixed cost F in principle could take into account future benefits and liabilities of holding up the principal.

2.1 Costs and Per Period Payoffs

2.1.1 Costs

If the agent works for the principal he may get more efficient. However, when the principal opts for in-house production the agent is away from the technological frontier and may become less efficient in producing cutting edge products. Formally, the costs of production for the agent is a controlled Markov chain with a finite state space. For simplicity I assume the following cost structure and law of motion: The cost of the agent in period t is $c^t \in C \equiv \{c_1, c_2, \ldots, c_{n-1}, c_n\} \subset \mathbb{R}_+$ with $n \in \mathbb{N}$. The costs are weakly increasing i.e., $c_1 \leq c_2 \leq \ldots c_n$ and satisfy the following convexity assumption: $c_k - c_{k-1} \leq c_{k+1} - c_k$ for $1 < k < n$. That is the efficiency gains get smaller as the agent moves along the learning curve.

If the production is outsourced to the agent in period t, then the costs decrease according to the following distribution:

$$(c^{t+1}|I^t = 1, c^t = c_k) = \begin{cases} c_k & \text{with } 1-r \text{ probability if } c_k > c_1 \\ c_{k-1} & \text{with } r \text{ probability if } c_k > c_1 \\ c_k & \text{if } c_k = c_1 \end{cases} \tag{CD}$$

Here a single parameter (r) captures the speed of learning. When r is higher the agent is likely to experience cost reduction when working. When r is low the agent is likely to stay with the same cost even if he works. On the other hand when the principal produces in-house, the agents cost increase according to the following distribution:

$$(c^{t+1}|I^t = 0, c^t = c_k) = \begin{cases} c_{k+1} & \text{with } q \text{ probability if } c_k < c_n \\ c_k & \text{with } 1-q \text{ probability if } c_k < c_n \\ c_k & \text{if } c_k = c_n \end{cases} \tag{CU}$$

Similarly, the parameter (q) captures the speed of forgetting in this formulation. When q is higher the agent is likely to lose efficiency when not working. When q is low the agent is likely to stay with the same cost even if he doesn't work.

This cost structure captures learning by doing in a simple manner. As the agent does more work for the principal he gets more efficient, when he doesn't work he slowly loses this efficiency.

An alternative interpretation of the cost structure that fits the electronics industry is as follows. Over time the production of cutting edge technology incrementally becomes more demanding and technically involved, working enables the contract manufacturer to at least keep up with the requirements of the current frontier, while not working the agent may fall behind if the frontier moves forward and become less proficient.

In order to capture learning by doing, I assume that at the beginning of the game the agent is at the beginning of the learning curve, i.e. $c^0 = c_n$.

2.1.2 Actions and Per Period Payoffs

The profits from production depend on the market structure (i.e. the number of producing firms) which, in this model is determined by whether the agent has entered the market or not. I assume that there are no learning opportunities in the in-house production as the principal has already done the R & D. The inefficiency in the in-house production arises because the principals production capabilities are small compared to the agent and hence, does not have similar scale economies.

Given a relational contract (π, τ_e) the period t payoff of the principal $u_p(c^t, \pi, \tau_e)$ is given by:

$$u_p(c^t, \pi, \tau_e) = \begin{cases} v_o - c^t - p^t & \text{if } I^t = 1 \text{ and } P(\tau_e > t|\mathcal{F}_t) = 1 \\ w_o - p^t & \text{if } I^t = 0 \text{ and } P(\tau_e > t|\mathcal{F}_t) = 1 \\ w_e & \text{o.w.} \end{cases}$$

The period t payoff of the agent denoted by u_o^a is as follows:

$$u_a(c^t, \pi, \tau_e) = \begin{cases} p^t & \text{if } P(\tau_e > t|\mathcal{F}_t) = 1 \\ v_e - c^t - F & \text{if } P(\tau_e = t|\mathcal{F}_t) = 1 \\ v_e - c^t & \text{if } P(\tau_e < t|\mathcal{F}_t) = 1 \end{cases}$$

I assume $w_e < w_o$ and $v_e < v_o$ so competition reduces the profits for both parties.

2.2 Payoffs and Constraints

Given a relational contract (π, τ_e) the total discounted payoff to agent $U_a(\pi, \tau_e)$ is as follows:

$$U_a(\pi, \tau_e) = E(\sum_{t=0}^{\tau_e - 1} \delta^t p^t + \sum_{t=\tau_e}^{\infty} \delta^t (v_e - c^t) - \delta^{\tau_e} F|\pi, \tau_e) \tag{2.1}$$

Given a relational contract (π, τ_e) the total discounted payoff of the principal $U_p(\pi, \tau_e)$ is as follows:

$$U_p(\pi, \tau_e) = E(\sum_{t=0}^{\tau_e - 1} \delta^t (I^t(v_o - c^t) + (1 - I^t)w_o - p^t) + \sum_{t=\tau_e}^{\infty} \delta^t w_e | \pi, \tau_e) \quad (2.2)$$

The payoff of the agent when he enters the market is dependent only on his cost level when he enters. Thus the payoff of the agent after he enters the market with a c_k is:

$$\begin{aligned} U_a^e(k) &= E(\sum_{t=0}^{\infty} \delta^t (v_e - c^t) | c^0 = c_k) - F \\ &= \sum_{n=0}^{k-2} \left(\frac{r\delta}{1 - \delta(1 - r)} \right)^n \frac{v_e - c_{k-n}}{1 - \delta(1 - r)} + \left(\frac{r\delta}{1 - \delta(1 - r)} \right)^{k-1} \frac{v_e - c_1}{1 - \delta} - F \end{aligned}$$

Finally the agents profits net of the fixed cost after entry with cost level c_k is denoted $A(k)$ and is given by:

$$A(k) = U_a^e(k) + F$$

I assume the following:

Assumption 1.

$$\frac{(1 - \delta + \delta r)w_o + \delta q(v_o - c_n)}{(1 - \delta)(1 - \delta + \delta r + \delta q)} - U_a^e(n - 1) \geq \frac{w_e}{1 - \delta}$$

Assumption 1 guarantees that that the principal doesn't want to compete in the long run. That is the payoff from maintaining the relationships is larger than eventually allowing the agent to enter. When assumption 1 is not satisfied, the length of a principal optimal relationship is almost surely finite, which yields an optimal control problem, there is large body of literature on such problems, thus I restrict attention to relationships of infinite length.

Assumption 2.

$$U_a^e(n) \geq 0$$

Assumption 2 guarantees that entry yields non-negative profits for the agent.[3]

In order to make sure that the agent doesn't enter, it must be the case that for any period the continuation utility for the agent from that period onward is greater than entering the market. For any period $\hat{t}$

$$E(\sum_{t=\hat{t}}^{\infty} \delta^{t-\hat{t}} p^t | \mathcal{F}_{\hat{t}}) \geq E(\sum_{t=\hat{t}}^{\infty} \delta^{t-\hat{t}} v_e - c^t | \mathcal{F}_{\hat{t}}) - F. \tag{IC_A}$$

Analogously the incentive constraint of the principal takes the following form for any period $\hat{t}$:

$$E(\sum_{t=\hat{t}}^{\infty} \delta^{t-\hat{t}} (I^t(v_o - c^t) + (1 - I^t) w_o) - \sum_{t=\hat{t}}^{\infty} \delta^{t-\tau} p^t | \mathcal{F}_{\hat{t}}) \geq \frac{w_e}{1 - \delta} \tag{IC_P}$$

With all the constraints identified, the principal's problem can be summed up as follows:

Problem 1 (Principal's Problem)**.**

$$\max_{\{\{p^t\},\{I^t\}\}_{t\in\mathbb{N}}} E(\sum_{t=0}^{\infty} \delta^t I^t (v - c^t + (1 - I^t) w_o) - \sum_{t=0}^{\infty} \delta^t p^t)$$
$$\textit{subject to } IC_A$$
$$IC_P$$

2.3 The Principal Optimal Contract

Under assumption 1, in the principal optimal relational contract the agent is sometimes employed and sometimes isn't but is paid enough payments such that agent never enters. In this section I first identify the payment rules that satisfy the agents incentive constraints, then I identify the optimal schedule of outsourcing.

[3] Assumption 2 is for simplifying the strategy space, it is possible to consider an extended model where the agent can reject the offer of the principal but not enter. Since rejection only happens off-path, the assumption guarantees that the optimal punishment is entry in this broader model. This broader model without the assumption also results in a qualitatively similar contract but with a slightly different payment scheme.

2.3.1 The Payment Rule and Thresholds

The following definitions will be important to find the optimal contract.

Definition 1 (Monotone Payments)**.** *A monotone payment scheme is identified by Markovian payments that depends on the current efficiency level of the agent and whether or not the agent is employed,* $(p^t|c^t = c_k, I^t = 1) = p_a(c_k)$ *and* $(p^t|c^t = c_k, I^t = 0) = p_p(c_k)$ *where,*

$$\begin{aligned} p_a(c_k) =& v_e - c_k - (1-\delta)F \\ p_p(c_k) =& \frac{A(k-1)\left(1-2\delta+\delta^2+\delta q+\delta r-\delta^2 q-\delta^2 r\right)}{1-\delta+\delta r} \\ & - \frac{F\left(1-2\delta+\delta^2-\delta r-\delta^2 r\right)-(v_e-c_k)\delta q}{1-\delta+\delta r} \end{aligned}$$

The agent can threaten to enter into the market regardless of whether the principal wants to outsource in that period or not. However, the threat depends on the current efficiency level of the agent thus the monotone payments reflect only this dependence.

Before fully describing the optimal contract, it is useful to introduce the following definition and highlight its importance.

Definition 2 (Monotone Utilization Policies)**.** *A policy is called monotone utilization policy if* $\exists k \in \{1, 2, \ldots, n\}$ *such that for all* $j < k \quad c^t = c_j \Rightarrow I(t) = 0$ *and for all* $l \geq k \quad c^t = c_l \Rightarrow I(t) = 1$

A policy is called monotone utilization policy with threshold k if the agent is utilized whenever his costs are greater or equal to c_k and he is never employed if his costs are below c_k.

2.3.2 An Optimal Monotone Policy Pair

In the appendix propositions 4 and 5 show that a monotone utilization policy with monotone payments is an optimal solution to the principal's problem. The threshold for the monotone utilization policy is characterized by indices.

Definition 3. *The index of state* c_x *denoted by* $\lambda(c_x)$ *is given by*

$$\lambda(c_x) = \frac{f_x^x - f_{x+1}^x}{g_x^x - g_{x+1}^x}$$

Where

$$
\begin{aligned}
f_x^x &= \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)}(v_o - c_x - p_a(c_x)) \\
&+ \frac{(\delta r)}{(1-\delta)(1-\delta+\delta r+\delta q)}(w_o - p_p(c_{x-1})) \\
f_{x+1}^x &= \frac{\delta q}{(1-\delta)(1-\delta+\delta r+\delta q)}(v_o - c_{x+1} - p_a(c_{x+1})) \\
&+ \frac{(1-\delta+\delta r)}{(1-\delta)(1-\delta+\delta r+\delta q)}(w_o - p_p(c_x)) \\
g_x^x &= \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)} \\
g_{x+1}^x &= \frac{\delta q}{(1-\delta)(1-\delta+\delta r+\delta q)}
\end{aligned}
$$

The index of state c_x captures the time normalized marginal change in profits by adding the state c_x to the monotone utilization policy $x+1$.

Theorem 1. *In the optimal policy, there $\exists k^*$ such that, for all $k \geq k^*$, $c^t = c_k \Rightarrow I^t = 1$ and for all $l < k^*$, $c^t = c_l \Rightarrow I^t = 0$. And k^* is the smallest integer that satisfies:*

$$
\lambda(c_{k^*}) \geq \frac{(v_o - v_e + (1-\delta)F - \delta w_e)(1-\delta+\delta r+\delta q)}{1-\delta+\delta r}
$$

The dynamics of the relationship is driven by the learning opportunities available to the agent hence comparative statics with respect to learning are of particular importance. Due to the relatively simple policy and the closed form of the index it is possible to investigate these relationships.

Proposition 1. *As the speed of learning, r, increases, the profits of the principal and the expected discounted time the agent spends working, decrease.*

Proposition 1 highlights the tension in the relationship as the learning opportunities increase. The decreasing profits for the principal that arises from higher speed of learning may at first seem surprising but has a very clear intuition. A higher speed of learning increases the predatory threat of the agent, that is an agent who learns faster may be able enter the market and compete with the principal sooner. Thus, the principal has to pay higher wages to prevent entry at the threshold level. Moreover due to faster learning the time that the agent is utilized for production at the threshold level also decreases leading to loss of profits for the principal since the agent is more efficient. Finally, due to monotone payments the gains arising from faster learning before the threshold is reached goes to the agent. Thus there is no change to the principal's profits before the threshold is reached. The three effects combined leads to a loss for the principal arising from a faster learning agent.

When choosing the optimal threshold the principal must consider two factors, first is the effect of utilizing the agent on profits, second the effect of this utilization on the threat of entry. The index $\lambda(c_x)$ captures the marginal gains in profits from a utilization decision. An important feature of the optimal policy is that the payments become larger as the agent gets more efficient. That is, as the agent gets more efficient the profit gains from utilization decrease, whereas the payments necessary to maintain the relationship increase. The driving factor for this result is that a gain in efficiency increases the outside option of the agent. Thus the efficiency gained precisely by working for the principal is used to threaten the principal. At the most efficient level that is reached the agent does not work but is paid just to make sure he does not enter.

3 Commodified Product and Entrenched CMs

In this section I investigate the optimal relational contract between a single principal and multiple, imperfectly substitutable contract manufacturers. In situations where the product of the principal is commodified a principal usually maintains relationships with multiple contract manufacturers. This is due to the fact that sometimes the contract manufacturers are entrenched with other principals (unmodeled). In this case a contract manufacturer can still work albeit at higher costs due to the overwork required. Moreover since the product is commodified hold-up becomes a more relevant threat.

Formally, there are $N+1$ players; 1 principal (she) and N agents (he). Time is discrete and the horizon is infinite i.e., $t \in \mathbb{N}$. All parties share a common discount rate $\delta \in (0,1)$. At each period t the following events unfold:

- Agents' production costs $c^t = (c_1^t, \ldots, c_N^t)$ are realized and become publicly known.

- The principal chooses on average one agent $i \in N$ to utilize and pays the cost c_i^t, and promises p_i^t out of production. Agents can refuse the investment before it is made. Formally, let $I_i^t \in \{0, 1\}$ denote the principal utilizing agent i in period t.

- The utilized agent manufactures the product of value v and p_i^t is given to agent $v - p_i^t$ to the principal. Alternatively the agent can hold up the principal during production and demand up to $v - l_i$, where $l_i < v$.

In case of contract manufacturing, firms often outsource the entire production, and many aspects of the process are not contractible. The maximum on the hold up, $v - l_i$, captures the liabilities associated with intellectual property rights of the principal and potential costs of litigation associated if the agent tries to hold up the principal. The bargaining power of an agent depends on the size of his liabilities l_i. The case of 0 liabilities becomes a pure hold-up problem where agents enjoy full bargaining power.

A history at period t consists of all the employment decisions and all the payments made as well as all the costs, $h^t = ((\{I_i^0\}_{i \in N}, \{p_i^0\}_{i \in N}, \{c_i^0\}_{i \in N}), (\{I_i^1\}_{i \in N}, \{p_i^1\}_{i \in N}, \{c_i^1\}_{i \in N}), \ldots, (\{I_i^t\}_{i \in N}, \{p_i^t\}_{i \in N}, \{c_i^t\}_{i \in N}))$ up to period t. The set of all histories at period $t, t + 1, \ldots$ generate a growing sequence of σ-algebras, i.e. a filtration, $\{\mathcal{F}_t\}$. The probability triple (i.e. the filtered probability space) is given by $(\Omega, \{\mathcal{F}_t\}, P)$ where, Ω is the set of all histories P is the probability measure over Ω and $\{\mathcal{F}_t\}$ is the natural filtration. Let E denote the expectation operator associated with P. Principals strategy denoted as π is a $\mathcal{F}_t$ measurable plan of employment decisions. Formally, $\{\{I_i^t\}_{i \in N}\}_{t \in \mathbb{N}}$, where $I_i^t = 1$ if agent i is selected at period t and 0 otherwise. Similarly p_i^t denotes the non-contractible fee that i receives from the principal. I assume that payment can only be made when an agent is producing for the principal, thus $I_i^t = 0 \Rightarrow p_i^t = 0$. An agents strategy denoted by σ_i is a $\mathcal{F}$ measurable plan of fees demanded $\{p_i^t\}_{t \in \mathbb{N}}$ A tuple (π, σ), where $\sigma = (\sigma_1, \ldots, \sigma_N)$ is called a relational contract.

3.1 Costs and Per Period Payoffs

3.1.1 Cost Structure, Entrenchment and Average Demand

In the previous section working could lead to a decrease in costs due to learning. However, it is often the case that learning is not the main aspect of the dynamics in a commodified product. The main aspect is rather based upon the effect of over-utilizing assets. A frequent use of employment may lead to higher costs and not lower as in the case of learning by doing. In addition in this section I allow for multiple agents. When an agent is not working for the principal, he does not encounter any capacity constraints. Thus his costs decrease to a lower bound. On the other hand if an agent is employed by the principal, he may encounter capacity problems. He may have to over-utilize his assets (e.g., pay overtime) and his contracting costs will increase due to over-utilization. Formally, principal's costs of investing in an agent is a controlled Markov chain with a finite state space. For simplicity I assume the following cost structure and law of motion: The cost of an agent i at period t is $c_i^t \in C_i \equiv \{c_{i,1}, c_{i,2}, \ldots, c_{i,n_i-1}, c_{i,n_i}\} \subset \mathbb{R}_+$ with $n_i \in \mathbb{N}$. The costs are weakly increasing i.e., $c_{i,1} \leq c_{i,2} \leq \ldots c_{i,n_i}$ and satisfy the following convexity assumption: $c_{i,k} - c_{i,k-1} \leq c_{i,k+1} - c_{i,k}$ for $1 < k < n_i$. The whole vector of costs in period t is denoted $c^t = (c_1^t, \ldots, c_N^t) \in C_1 \times \ldots C_N$.

I assume $v > c_{i,n_i}$ for all i, so if there were no payments to the agents, the principal would always like to produce. Similarly the liabilities are smaller than the cost of production, thus for all i, $c_{i,1} > l_i$. If agent i is chosen by the principal in period t, his costs increase according to the following distribution:

$$(c_i^{t+1} | I_i^t = 1, c_i^t = c_{i,k}) = \begin{cases} c_{i,k+1} & \text{with } q_i \text{ probability if } c_{i,k} < c_{i,n_i} \\ c_{i,k} & \text{with } 1 - q_i \text{ probability if } c_{i,k} < c_{i,n_i} \\ c_{i,k} & \text{if } c_{i,k} = c_{i,n_i} \end{cases} \tag{CU}$$

If agent i is not chosen by the principal in period t, he catches up with his entrenchments and his costs decrease to their initial levels:

$$(c_i^{t+1} | I_i^t = 1, c_i^t = c_{i,k}) = \begin{cases} c_{i,1} & \text{with } r_i \text{ probability} \\ c_{i,k} & \text{with } 1 - r_i \text{ probability} \end{cases} \tag{CD}$$

This cost structure displays entrenchments and spillovers in the economy in a simple manner. CU captures the upward movements in costs due to reaching capacity constraints after being employed and CD captures the reinitialization. So, demand for an agent increases his costs (price of employment), and the spillover from demand for an agent is the opportunity generated for other agents to catch up on their entrenched work/rest so that their future costs are not as high.

Finally I assume that the principal faces a long term average demand of 1, with potential fluctuations hence the total amount produced satisfies the following.

$$E(\sum_{t=0}^{\infty}\sum_{i\in N} I_i^t) \leq 1/(1-\delta)$$

3.1.2 Actions and Per Period Payoffs

The period t payoff of the principal denoted $u_p(c^t, \pi)$ given a relational contract (π, σ) is given by

$$u_p(c^t, \pi, \sigma) = \sum_{i=1}^{N} I_i^t(v - c_i^t - p_i^t)$$

The period t payoff of an agent i denoted $u_i(c^t, \pi)$ given a relational contract (π, σ) is given by

$$u_i(c^t, \pi, \sigma) = I_i^t p_i^t$$

3.2 Payoffs and Constraints

The total discounted payoff of agent i denoted $U_i(\pi, \sigma)$ given (π, σ) is as follows:

$$U_i(\pi, \sigma) = E(\sum_{t=0}^{\infty} \delta^t I_i^t p_i^t | \pi, \sigma) \tag{3.1}$$

Similarly the profits of the principal $U_p(\pi, \sigma)$ given (π, σ) is as follows:

$$U_p(\pi, \sigma) = E(\sum_{t=0}^{\infty}\sum_{i=1}^{N} \delta^t I_i^t(v - c_i^t - p_i^t) | \pi, \sigma) \tag{3.2}$$

From the profits it is easy to isolate the profits raised from agent i as follows:

$$U_p^i(\pi, \sigma) = E(\sum_{t=0}^{\infty} \delta^t I_i^t(v - c_i^t - p_i^t) | \pi, \sigma)$$

Since the principal only incentivizes agents when they are employed, the incentive constraint depends on the times she employs agent i. Let $\tau_{i,1} = \inf\{t \geq 0 : I_i^t = 1\}$ denote the first time agent i is employed by the principal. Inductively, let $\tau_{i,n} = \inf\{t > \tau_{i,n-1} : I_i^t = 1\}$ denote the nth time agent i is employed by the principal. Thus the sequence of random variables $\{\tau_{i,n}\}_{n\in\mathbb{N}}$ denotes all the periods that agent i is utilized by the principal. Hence, the incentive constraint takes the following form:

$$E(\sum_{k=n}^{\infty} \delta^{\tau_{i,k}-\tau_{i,n}} p_i^{\tau_{i,k}} | \pi, \sigma) \geq E(I_i^{\tau_{i,n}}(v - l_i)) \text{ for all } n. \tag{IC_i}$$

Analogously the incentive constraint of the principal takes the following form:

$$E(\sum_{k=n}^{\infty} \delta^{\tau_{i,k}-\tau_{i,n}} (v - c_i^{\tau_{i,k}} - p_i^{\tau_{i,k}}) | \pi, \sigma) \geq 0 \text{ for all } n \text{ for all } i. \tag{IC_P}$$

Notice that this incentive condition implies that the punishments are bilateral. The main reason for this is to broaden the scope of industries captured by the model. Depending on the industry, agents may or may not be able to jointly punish a principal, but if a relational contract is sustainable under bilateral punishments, then it is necessarily sustainable when multiple agents cooperate to punish the principal.

With all the constraints identified, the principal's problem can be summed up as follows:

Problem 2 (Principal's Problem)**.**

$$\begin{aligned} \max_{\{\{p_i^t\}_{i\in N}, \{I_i^t\}_{i\in N}\}_{t\in\mathbb{N}}} \quad & E(\sum_{t=0}^{\infty} \sum_{i=1}^{N} \delta^t I_i^t (v - c_i^t - p_i^t)) \\ \textit{subject to} \quad & E(\sum_{t=0}^{\infty} \sum_{i\in N} I_i^t) \leq 1/(1-\delta) \\ & IC_i \quad \forall i \\ & IC_P \end{aligned}$$

3.3 The Principal Optimal Contract

The principal optimal contract consists of employment decisions and payments. In this section, first I identify a simple payment rule that gives away the minimal economic rents for any employment rule. Then, I identify the optimal employment rule.

3.3.1 Fastest Prices

Once an employment rule is chosen payments can be done in a myriad of ways while satisfying the incentive constraints of the agents. The fastest prices were introduced in Board (2011) as a payment scheme that satisfies the incentive conditions for all agents at every period with equality, i.e., while giving minimum economic rents. The remarkable feature of fastest prices is that they tie payments directly to employment, without relying on any other variables.

Definition 4 (Fastest Prices)**.** *For any employment rule $\{I_i^t\}_{t\in\mathbb{N}}$, the fastest prices are given by*

$$p_i^{\tau_{i,n}} = (v - l_i)E(1 - \delta^{\tau_{i,n+1}}|\tau_{i,n}) \quad \textit{for all } n \in \mathbb{N}$$

Proposition 2 (Board)**.** *For any employment rule $\{I_i^t\}_{t\in\mathbb{N}}$, no pricing rule can yield higher profits than fastest prices.*

The proof of proposition 2 is identical to Board (2011), hence omitted.

3.4 Optimal Employment Rule

The optimal employment rule is crucial since the payment rule can be readily characterized for a given employment rule using fastest prices. Hence, the behavior of the entire principal optimal contract depends on the properties of the employment rule.

Despite the complex nature of the problem, the optimal employment rule is surprisingly simple. The optimal employment rule is an index rule that is augmented by a threshold. In particular, each agent is assigned an index that is only dependent on his current cost, and the principal employs all agents that have an index higher than the threshold.

Definition 5. *The index of agent i at state $c_{i,x}$ is given by :*

$$\lambda_i(c_{i,x}) = \frac{f_{i,x}^x - f_{i,x-1}^x}{g_{i,x}^x - g_{i,x-1}^x}$$

Where

$$g_{i,x}^x = \frac{\frac{1}{1-\delta(1-q_i)} + \delta \frac{q_i\delta}{1-\delta(1-q_i)} \left[\sum_{n=0}^{x-2} \frac{1}{1-\delta(1-q_i)} \left(\frac{\delta q_i}{1-\delta(1-q_i)}\right)^n\right]}{1-\delta\left(\frac{\delta q_i}{1-\delta(1-q_i)}\right)^x}$$

$$g_{i,x-1}^x = \frac{\delta \left[\sum_{n=0}^{x-2} \frac{1}{1-\delta(1-q_i)} \left(\frac{\delta q_i}{1-\delta(1-q_i)}\right)^n\right]}{1-\delta\left(\frac{\delta q_i}{1-\delta(1-q_i)}\right)^{x-1}}$$

$$f_{i,x}^x = \frac{\frac{v-c_{i,x}}{1-\delta(1-q_i)} + \delta \frac{q_i\delta}{1-\delta(1-q_i)} \left[\sum_{n=0}^{x-2} \frac{v-c_{i,n+1}}{1-\delta(1-q_i)} \left(\frac{\delta q_i}{1-\delta(1-q_i)}\right)^n\right] - (v-l)}{1-\delta\left(\frac{\delta q_i}{1-\delta(1-q_i)}\right)^x}$$

$$f_{i,x-1}^x = \frac{\delta \left[\sum_{n=0}^{x-2} \frac{v-c_{i,n+1}}{1-\delta(1-q_i)} \left(\frac{\delta q_i}{1-\delta(1-q_i)}\right)^n\right] - \delta(v-l)}{1-\delta\left(\frac{\delta q_i}{1-\delta(1-q_i)}\right)^{x-1}}$$

Theorem 2. *The optimal employment rule is characterized by a set of indices* $\{\{\lambda_i(c_{i,x})\}_{c_{i,x}\in C_i}\}_{i\in N}$ *and a threshold* $\lambda^* \geq 0$. *For all* t *and all* i, $I_i^t = 1 \Leftrightarrow \lambda_i(c_{i,x}) > \lambda^*$ *and* $I_i^t \in [0,1] \Leftrightarrow \lambda_i(c_{i,x}) = \lambda^*$.

Where λ^* *is the smallest* $\lambda \in \mathbb{R}$ *that satisfies*

$$\sum_{i\in N}\left[\sum_{n=0}^{x-2} \frac{1}{1-\delta(1-q)} \left(\frac{\delta q}{1-\delta(1-q)}\right)^n + \left(\frac{\delta q}{1-\delta(1-q)}\right)^{x-1} g_{i,c_i(\lambda)}^{c_i(\lambda)}\right] = 1/(1-\delta)$$

The randomization at the exact threshold level is necessary to ensure that the average demand is met. In general cases where no two agents are identical, only one agent will ever need to randomize.

When making employment decisions the principal could potentially be relying on the entire history of all the relationships she maintains, as well as all the promises she made. However, the optimal employment scheme takes a rather simple form. The threshold is fixed at the very beginning of the game and the indices *only* depend on the current cost level of an agent. In an economy with heterogenous agents and laws of motion, the indices provide a simple employment scheme to maximize profits, where just employing the cheapest agent is simple, but not necessarily profit maximizing.

An important feature of the index is that it does not depend on employment history. Thus, in an optimal employment rule, whether an agent has worked or not for the principal does not factor into the employment decision.

In particular the index of an agent does not depend on the employment history of any agents, including himself. So, in the optimal contract an agent who has worked for the principal does not receive preferential treatment over an agent who has never worked for the principal before.

Moreover initial costs do not factor into the indices either. In particular, at the start of a relationship an agent might have significant cost advantages or disadvantages, resulting in very frequent or very rare employment in the early periods. However, these frequencies do not necessarily last throughout the game. Where an agent starts from has no bearing on the relationship in the long run.

Finally, the indices do not depend on other agents at all. Any employment decision necessarily implies that some agents are preferred over others. However, in the optimal contract this preference is not a cross comparison. The principal does not compare agents and pick the *best*, she employs anyone that is *good enough*. How "good" an agent is dependent on both the current cost of an agent and his entrenchments (the law of motion for costs) which is captured by his index. The threshold represents the criterion for what is "good enough". This criterion for being good enough is fixed at the beginning of the game and doesn't change according to the realized costs or actions of players throughout, although the initial determination does depend on the total number of agents, their potential cost structures as well as their starting costs. In essence, who is good enough depends on what every agent *should* do but which history is actually realized does not change this criterion.

There are two kinds of loyalty that could be considered in this setting. The first one is being loyal to promises, that is, if the principal promises future work she will indeed employ the agents in the future. The principal is loyal to her promises, in the sense that any employment promise she makes will be fulfilled in finite time. This is especially important since the entire economy need not follow a recurring pattern, even in the very long run, unless the strategies chosen by the players are precisely intended to cause the economy to recur. When recursion is easily avoidable, breaking promises is a very plausible strategy, unlike a repeated environment. However, even though the optimal contract does not start with a cyclical pattern, a patient principal will maintain her relationships by being loyal to her promises and will eventually converge to a cyclical employment pattern. The second kind of loyalty that could be considered is preferential treatment of agents based on a longer employment history. As the indices do not depend on history at all, this kind of loyalty is not present in the optimal contract.

3.5 A Simplification with Full Resetting

Although the setting explored here with different randomized recovery times is broader, it comes with the cost of relaxing the employment decisions to an average. However if desired, by setting $r_i = 1$ for all i and using the same indices a sharper result can be obtained.

Proposition 3 (Jacko)**.** *If $r_i = 1$ for all i, an index policy with fastest prices and with indices identified as*

$$\lambda_i(c_{i,x}) = \frac{f_{i,x}^x - f_{i,x-1}^x}{g_{i,x}^x - g_{i,x-1}^x}$$

will yield at most one agent being employed every period, that is the employment constraint can be reduced to

$$\sum_{i \in N} I_i^t \leq 1$$

.

The result of Jacko (2011) was to show that the marginal productivity indices identified in a deterministically fully resetting restless bandit setting is optimal, which interpreted in this setup yields the above proposition. However, this result also highlights two aspects of the analysis. First, by putting more structure onto the restless relationships might deliver sharper results. Second, as the sufficient nature of the result might imply it is a daunting task to identify necessary conditions for particularly desirable results. In the settings that I explored bi-directionality, that is employment decisions moving the state in the opposite direction of unemployment was extremely helpful, but again only sufficient and certainly not necessary for indexability of the problems.

4 A Short Description of the Methodology

The optimal contracts identified in this paper are mostly tied to the optimal utilization schedules which are characterized by indices. Index policies are unorthodox in relational contracting settings, but they are prevalent in bandit problems. The pioneering work of Gittins (1979) showed that some bandit scheduling problems can be solved by a rather simple index policy, which is characterized arm by arm via identifying indifference points. The methodology for solving the principal's problem is slightly more involved than a bandit problem but capitalizes on the same idea.

The main reason for utilizing a bandit approach is that existing Folk theorem based approaches are inapplicable. In the setting that I explore the inapplicability arises from the fact that costs are tied to employment. In particular players have the power to reduce the payoff relevant state space by refusing to work/employ, turning the game into a *reducible* stochastic game. Unlike irreducible games, reducible games do not have Folk Theorems. Reducibility necessarily breaks down recursion since some continuation payoffs that are currently available need not be available ever again, which prohibits Folk like results relying on the payoff space to be independent from the strategies of players at least in the very long run. This lack of ability to use recursion is not limited to the current setting either, such a phenomena is bound to occur in other settings where a player's decisions have long lasting effects; such as liquidity constraints tied to performance, long lasting investments, or settings where the recurring payoff space is very hard to deal with such as cases where players have different discount factors. Hence, a new methodology is required. Since relational contracts are essentially scheduling problems, (whether it is a problem of when to employ or when to induce effort), bandit techniques are applicable.

In order to utilize bandit techniques, the first step is to transform the principal's problem. The critical feature that is utilized here is the existence of payments, which enables handling forward looking constraints in a Markovian fashion by tying them to utilization. This is a relaxation of the original problem, but the transferable utility achieved by the presence of money ensures that the solution to the relaxed problem is feasible in the constrained problem as well. Once the forward looking constraints of the agents are settled the optimization readily resembles a standard bandit problem, with one caveat: the arms that are not pulled also change their state. Such a problem is readily found in the the bandits literature as *restless* bandits.

The work on restless bandits was pioneered by Whittle (1988), based on a Lagrangian relaxation of the utilization constraint, as the non-relaxed problems are generally intractable as shown by Papadimitriou and Tsitsiklis (1999). However, even the relaxed problems are not guaranteed to be solved by index policies, and ascertaining that a problem is *indexable* is a daunting task. Furthermore like the Gittins (1979) index, calculating indices in closed form is also usually not feasible. In order to tackle these problems I utilize the bi-directional nature of the cost structure, building upon the work of Glazebrook, Kirkbride, and Ruiz-Hernandez (2006) and Niño-Mora (2007) and acquire indices in closed form, which enables ascertaining the indexability of the problem via monotonicity of the indices. Finally, for the optimality of the index policy I build upon the work of Jacko (2011).

The index has a natural interpretation in an economic setting. Consider a single agent problem where employment yields the same returns, and unemployment yields a subsidy equal to λ and the problem is to choose when to employ the agent. Utilizing the fact that employment and unemployment force the costs to go into opposite directions, the optimal policies in the single agent problem can be shown to be threshold policies. Utilizing threshold policies, the subsidy level for indifference of employment at particular cost values can be identified in closed form and yields the indices much like Gittins (1979) index. The main characteristics of the indices are quite similar, they provide a way to capture the marginal value of employing an agent/activating an arm in a state, taking into account the entire law of motion. They do not depend on other agents, they do not depend on history, they only depend on the current state. The original Gittins index policy is for a problem where arms that are not pulled remain the same, thus the Gittins index has the natural interpretation as the time normalized average returns of utilization and looks like a stopping problem. The index identified here captures the time normalized marginal returns to utilization and is a problem of selecting an active and passive set.

The two different settings are essentially solved by utilizing similar techniques highlighting the applicability of the methodology. Despite differences in incentive conditions, differences in the number of agents and the non-recursive nature, this new approach promises wide applications to economic problems where lack of recursion causes technical challenges.

5 Conclusion

Outsourcing entire manufacturing of a product allows original equipment manufacturers to reduce labor costs, free up capital, and improve worker productivity (Arrunada and Vázquez 2006). Global contract manufacturing had an expected volume of \$515 billion in electronics industry and \$40.7 billion in pharmaceutial industry (Rajaram 2015, Pandya and Shah 2013). Under such potential gains contract manufacturing is inevitable, though it entails inescapable hazards. (Arrunada and Vázquez 2006). This paper explores two frequent problems, under different incentive conditions. First, I explore potential predation by a single contract manufacturer when there are opportunities for learning by doing. Second, I explore a setting with multiple contract manufacturers where entrenchments and hold up opportunities are need to be navigated.

In the first setting the opportunity to learn by working for the client has two effects, while it becomes cheaper to employ the contract manufacturer as he gets more efficient, under the threat of predation the efficiency gains are only enjoyed by the contract manufacturer. The principal tries to limit the rents that need to be paid by stopping utilization once the agent becomes too efficient. As the contract manufacturer loses its edge by being kept out of production, he is employed again. The rents that need to be paid slowly increases as the agent becomes more efficient and reaches a maximum when utilization is stopped.

In the second setting a commodified product with multiple potential contractors poses new incentive frictions. As McCoy (2003) notes, entrenchment is a frequent problem, and changes the dynamics of not one, but multiple bilateral relationships. This paper characterizes the optimal policy completely. The optimal policy is time consistent, and is in a simple closed form characterized by indices. The principal keeps her promises, but she does not prefer one agent over another just because she had employed one of them in the past. So, past employment does not lead to preferential treatment.

6 Appendix

In most calculations it is necessary to utilize a common version (see Serfozo (2009)) of Wald's identity for discounted partial sums with stopping times. For convenience I will include the identity here as well.

Identity 1 (Wald's Identity for Discounted Sums)**.** *Suppose that* $X_1, X_2, \ldots$ *are i.i.d. with mean* $\bar{x}$*. Let* $\delta \in (0,1)$ *and* τ *be a stopping time for* $X_1, X_2, \ldots$ *with* $E(\tau) < \infty$ *and* $E(\delta^\tau)$ *exists. Then*

$$E(\sum_{t=0}^{\tau} \delta^t X_t) = \frac{\bar{x}(1 - \delta E(\delta^\tau))}{1-\delta}$$

6.1 Sole Sourcing Model

This is the principals problem in the sole sourcing model(PPS).

$$\max_{\{\{p^t\},\{I^t\}\}_{t\in\mathbb{N}}} E(\sum_{t=0}^{\infty} \delta^t I^t(v - c^t + (1-I^t)w_o) - \sum_{t=0}^{\infty} \delta^t p^t) \tag{PPS}$$

subject to

$$E(\sum_{t=\hat{t}}^{\infty} \delta^{t-\hat{t}} p^t | \mathcal{F}_{\hat{t}}) \geq E(\sum_{t=\hat{t}}^{\infty} \delta^{t-\hat{t}} v_e - c^t | \mathcal{F}_{\hat{t}}) - F. \text{ for all } \hat{t}$$

$$E(\sum_{t=\hat{t}}^{\infty} \delta^{t-\hat{t}} (I^t(v_o - c^t) + (1-I^t)w_o) - \sum_{t=\hat{t}}^{\infty} \delta^{t-\tau} p^t | \mathcal{F}_{\hat{t}}) \geq \frac{w_e}{1-\delta} \text{ for all } \hat{t}$$

Consider the following relaxation of PPS:

$$\max_{\{\{p^t\},\{I^t\}\}_{t\in\mathbb{N}}} E(\sum_{t=0}^{\infty} \delta^t I^t(v - c^t + (1-I^t)w_o) - \sum_{t=0}^{\infty} \delta^t p^t) \tag{RPPS}$$

subject to

$$E(\sum_{t=0}^{\infty} \delta^t p^t | \mathcal{F}_0) \geq E(\sum_{t=0}^{\infty} \delta^t v_e - c^t | \mathcal{F}_0) - F.$$

Letting V_{PPS} denote the optimal value of PPS and V_{RPPS} denote the optimal value of RPPS, it must be the case that

$$V_{RPPS} \geq V_{PPS}$$

Focusing on RPPS, since the action space is compact and the returns function is upper semi-continous, Puterman (2014) shows that there is an optimal Markovian solution.

Proposition 4. *Any Markovian policy has an equivalent monotone utilization policy.*

Proof.

Observation 1. *Any Markov policy, will map states into payments and utilization decisions. Let π be any Markov policy and let S^{π} denote its active set, such that $I^t = 1 \Leftrightarrow c^t \in S^{\pi}$.*

Notice that the initial cost is c_n, and consider any Markov policy π, identified with its active set S^{π}. Let $c_{\underline{x}} = \max\{c \in C : c \notin S^{\pi}\}$. Then by definition under policy π, for all t $c^t \in \{c_{\underline{x}+1}, \dots c_n\}$. Moreover, for all t, $I^t = 1 \Leftrightarrow c^t > c_{\underline{x}}$. Now, consider the monotone utilization policy that has the same payment rule as π, identified with $c_{\underline{x}+1}$ denoted by $\pi_{\underline{x}+1}$. Then, by definition under policy $\pi_{\underline{x}+1}$ for all t, $c^t \in \{c_{\underline{x}+1}, \dots c_n\}$. Moreover, for all t, $I^t = 1 \Leftrightarrow c^t > c_{\underline{x}}$. Thus the two policies are equivalent.

Thus for any Markov policy, starting from any initial state, there is an equivalent monotone utilization policy. □

Proposition 5. *For a monotone utilization policy with threshold k, no payment scheme can yield higher profits than a monotone payments.*

Proof. Consider any monotone utilization policy π_k with threshold k. Under a monotone utilization policy π_k the CM will be utilized repeatedly until his costs reach the threshold c_k and after reaching the threshold he will be rested whenever his costs hit c_{k-1} and will be utilized again when the costs again rise to c_k. Thus, under a monotone utilization policy, all states c_l with $l > k$ are transient and the states c_k and c_{k-1} are recurrent. Given this a monotone payment scheme satisfies the incentive constraints with equality at all states.

Let $T(z)^y_x$ denote the expected discounted time spent in c_z starting from state c_y under π_x.

Starting from the recurrent states, c_k and c_{k-1}, the incentive conditions have to satisfy the following:

$$p_a(c_k)T(k)_k^k + p_p(c_{k-1})T(k-1)_k^k = E(\sum_{t=0}^{\infty} \delta^t v_e - c^t | c^0 = c_k) - F$$
$$p_p(c_{k-1})T(k-1)_k^{k-1} + p_a(c_k)T(k)_k^{k-1} = E(\sum_{t=0}^{\infty} \delta^t v_e - c^t | c^0 = c_{k-1}) - F$$

Lemma 1. *For any* $0 < r, q \leq 1$,

$$T(x)_x^x = \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)} \tag{6.1}$$
$$T(x-1)_x^x = \frac{\delta r}{(1-\delta)(1-\delta+\delta r+\delta q)} \tag{6.2}$$
$$T(x)_x^{x-1} = \frac{\delta q}{(1-\delta)(1-\delta+\delta r+\delta q)} \tag{6.3}$$
$$T(x-1)_x^{x-1} = \frac{1-\delta+\delta r}{(1-\delta)(1-\delta+\delta r+\delta q)} \tag{6.4}$$

Proof of lemma 1. $T(x)_x^x$ is the unique solution to the following system

$$T(x)_x^x = 1 + \delta(1-r)T(x)_x^x + \delta r T(x)_x^{x-1}$$
$$T(x)_x^{x-1} = \delta q T(x)_x^x + \delta(1-q)T(x)_x^{x-1}$$

Since expected discounted time is equal to $1/(1-\delta)$, $T(x-1)_x^x$ satisfies

$$T(x-1)_x^x = 1/(1-\delta) - T(x)_x^x$$

Thus,

$$T(x)_x^x = \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)}$$
$$T(x-1)_x^x = \frac{\delta r}{(1-\delta)(1-\delta+\delta r+\delta q)}$$

Similarly $T(x-1)_x^{x-1}$ is the unique solution to the following system

$$T(x-1)_x^x = \delta r T(x-1)_x^{x-1} + \delta(1-r)T(x-1)_x^x$$
$$T(x-1)_x^{x-1} = 1 + \delta q T(x-1)_x^x + \delta(1-q)T(x-1)_x^{x-1}$$

Identically, $T(x-1)_x^{x-1}$ satisfies

$$T(x-1)_x^{x-1} = 1/(1-\delta) - T(x)_x^{x-1}$$

Thus,

$$\begin{aligned} T(x)_x^{x-1} &= \frac{\delta q}{(1-\delta)(1-\delta+\delta r+\delta q)} \\ T(x-1)_x^{x-1} &= \frac{1-\delta+\delta r}{(1-\delta)(1-\delta+\delta r+\delta q)} \end{aligned}$$

□

Letting $A(k)$ denote $E(\sum_{t=0}^{\infty} \delta^t v_e - c^t | c^0 = c_k)$, by utilizing Strong Markov property and Wald's identity we have

$$A(k) = \sum_{n=0}^{k-2} \left(\frac{r\delta}{1-\delta(1-r)}\right)^n \frac{v_e - c_{k-n}}{1-\delta(1-r)} + \left(\frac{r\delta}{1-\delta(1-r)}\right)^{k-1} \frac{v_e - c_1}{1-\delta}$$

Finally, noticing $A(k)$ and $A(k-1)$ satisfies the following identity

$$A(k) = \frac{\delta r}{1-\delta(1-r)} A(k-1) + \frac{v_e - c_k}{1-\delta(1-r)}$$

A straightforward application of Farka's Lemma implies that there are two positive numbers $p_a(c_k)$ and $p_p(c_{k-1})$ that satisfies the incentive conditions with equality. Solving the system yields

$$\begin{aligned} p_a(c_k) &= v_e - c_k - (1-\delta)F \\ p_p(c_{k-1}) &= \frac{A(k-1)\,(1-2\delta+\delta^2+\delta q+\delta r-\delta^2 q-\delta^2 r)}{1-\delta+\delta r} \\ &\quad - \frac{F\,(1-2\delta+\delta^2-\delta r-\delta^2 r)-(v_e-c_k)\delta q}{1-\delta+\delta r} \end{aligned}$$

The incentive condition is satisfied with equality starting from the state c_k, thus for $p_a(c_{k+1})$, letting τ_k denote $\inf_{t\geq 0}\{t : c^t = c_k\}$ for the incentive condition to hold with equality we must have:

$$E(\sum_{t=0}^{\tau_k - 1} \delta^t p_a(c_{k+1}) + \delta^\tau (A(k) - F)) = \frac{r\delta}{1-\delta(1-r)} A(k) + \frac{v_e - c_{k+1}}{1-\delta(1-r)} - F$$

Plugging in the expectation yields:

$$\frac{(p_a(c_{k+1})))}{1-\delta(1-r)} + \frac{\delta r}{1-\delta(1-r)}(A(k)-F) = \frac{\delta r}{1-\delta(1-r)}A(k) + \frac{v_e - c_{k+1}}{1-\delta(1-r)} - F$$

Simplifying yields:

$$p_a(c_{k+1}) = v_e - c_{k+1} - (1-\delta)F$$

Inductively for all $x \geq k+1$ we must have

$$p_a(c_x) = v_e - c_x - (1-\delta)F$$

Where the incentive conditions hold with equality at every state. Since the incentive conditions hold with equality, no payment scheme can yield higher profits under a monotone utilization policy. □

Now, with the monotone payments baked in I introduce the following augmented return function $R(c^t|I^t)$, that is conditional on employment.

$$R(c^t|I^t = 1) = v_o - c^t - p_a(c^t)$$
$$R(c^t|I^t = 0) = w_o - p_p(c^t)$$

Now, consider the following relaxed hypothetical problem with no constraints, and the law of motion for c^t is identical to the principals problem.

$$\max_{\{I^t\}} E(\sum_{t=0}^{\infty} \delta^t R(c^t|I^t)) \tag{6.5}$$

Due to proposition 4 any Markovian policy is equivalent to a monotone policy. Let π_x denote a monotone policy, such that $I^t = 1 \Leftrightarrow c^t \geq c_x$. Let f_x^k denote expected discounted returns under policy π_x with initial state c_k. Similarly let g_x^k denote expected discounted utilization under policy π_x with initial state c_k. Formally:

$$f_x^k = E(\sum_{t=0}^{\infty} \delta^t R(c^t|I^t)|\pi_x, c^0 = c_k)$$

$$g_x^k = E(\sum_{t=0}^{\infty} \delta^t I^t|\pi_x, c^0 = c_k)$$

Since monotone policies automatically induce a family of nested sets, utilizing Niño-Mora (2007), the marginal productivity index for any state c_x denoted $\lambda(c_x)$ for the relaxed problem can be readily computed as

$$\lambda(c_k) = \frac{f_x^x - f_{x+1}^x}{g_x^x - g_{x+1}^x}$$

Utilizing the strong Markov property along with Wald's Identity, the components of the index can be calculated in closed form.

$$\begin{aligned} f_x^x &= \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)}(v_o - c_x - p_a(c_x)) \\ &+ \frac{(\delta r)}{(1-\delta)(1-\delta+\delta r+\delta q)}(w_o - p_p(c_{x-1})) \\ f_{x+1}^x &= \frac{\delta q}{(1-\delta)(1-\delta+\delta r+\delta q)}(v_o - c_{x+1} - p_a(c_{x+1})) \\ &+ \frac{(1-\delta+\delta r)}{(1-\delta)(1-\delta+\delta r+\delta q)}(w_o - p_p(c_x)) \end{aligned}$$

Similarly,

$$\begin{aligned} g_x^x &= \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)} \\ g_{x+1}^x &= \frac{\delta q}{(1-\delta)(1-\delta+\delta r+\delta q)} \end{aligned}$$

Due to Gittins, Glazebrook, and Weber (2011) we know that that the index being monotone in the state is a sufficient condition for the problem to be indexable, hence the index identified indeed captures the marginal productivity. Furthermore as there is just a single restless arm as shown by Jacko (2009) the relaxed problem is optimally solved by indices.

The next step is to observe that the optimal solution of the relaxed problem is feasible in PPS. Due to proposition 5 the incentive constraint of the agent binds with equality on every history that is reachable under the index policy. The only thing that needs to be checked is the principals constraint is satisfied.

Proposition 6. *Under a monotone utilization policy with threshold x the principal's incentive condition is satisfied at both state c_x and c_{x-1}.*

Proof. The proof is by induction, for the basis step observe that under assumption 1, $\lambda(c_n) > 0$ and the principal's incentive condition is satisfied under a monotone policy with threshold n, at both the state c_n and state c_{n-1}.

For the inductive step the following two lemmas are necessary

Lemma 2. *If $f_x^x \geq w_e/(1-\delta)$ then $f_{x+1}^{x+1} \geq w_e/(1-\delta)$.*

Proof of lemma 2. Observe that

$$\begin{aligned} f_x^x =& \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)}(v_o - v_e + (1-\delta)F) \\ &+ \frac{(\delta r)}{(1-\delta)(1-\delta+\delta r+\delta q)}(w_o - p_p(c_{x-1})) \end{aligned}$$

and

$$\begin{aligned} f_{x+1}^{x+1} =& \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)}(v_o - v_e + (1-\delta)F) \\ &+ \frac{(\delta r)}{(1-\delta)(1-\delta+\delta r+\delta q)}(w_o - p_p(c_x)) \end{aligned}$$

But since $p_p(c_x) \leq p_p(c_{x-1})$ we must have, $f_{x+1}^{x+1} \geq f_x^x \geq w_e/(1-\delta)$. □

Lemma 3. *If $f_{x+1}^x \geq w_e/(1-\delta)$, and $\lambda(c_x) > \frac{(v_o-v_e+(1-\delta)F-\delta r w_e)(1-\delta+\delta r+\delta q)}{1-\delta+\delta r}$, then $f_x^x, f_x^{x-1} > w_e/(1-\delta)$.*

Proof of lemma 3. $\lambda(c_x) > \frac{(v_o-v_e+(1-\delta)F-\delta r w_e)(1-\delta+\delta r+\delta q)}{1-\delta+\delta r} > 0$ implies that $f_x^x - f_{x+1}^x > 0$ since $g_k^k - g_{k+1}^k > 0$ for all k. Hence it must be the case that $f_x^x > f_{x+1}^x > w_e/(1-\delta)$. Moreover we must have

$$\begin{aligned} f_x^x &= \frac{v_o - v_e + (1-\delta)F}{1-\delta+\delta r} + \frac{\delta r}{1-\delta+\delta r} f_x^{x-1} \\ &\geq f_{x+1}^x + \frac{(v_o - v_e + (1-\delta)F - \delta r w_e)}{1-\delta+\delta r} \end{aligned}$$

Since $f_{x+1}^x \geq w_e/(1-\delta)$, it must be the case,

$$\frac{v_o - v_e + (1-\delta)F}{1-\delta+\delta r} + \frac{\delta r}{1-\delta+\delta r} f_x^{x-1} \geq \frac{w_e}{1-\delta} + \frac{(v_o - v_e + (1-\delta)F - \delta r w_e)}{1-\delta+\delta r}$$

Rearranging the terms yield

$$f_x^{x-1} \geq w_e/(1-\delta)$$

$\square$

For the inductive step, assume that for thresholds $k+1$ the statement is true. For state k to be the threshold, it must be the case that $\lambda(c_k) > \frac{(v_o - v_e + (1-\delta)F - \delta r w_e)(1-\delta+\delta r+\delta q)}{1-\delta+\delta r}$. By the inductive hypothesis, it must be the case that $f_{k+1}^k \geq w_e/(1-\delta)$, but then by lemma 3 it must also be the case that both $f_k^k > f_{k+1}^x > w_e/(1-\delta)$, concluding incentive compatibility on behalf of the principal.

$\square$

6.1.1 Proof of Proposition 1

Proof of Proposition 1. The proposition follows directly from the following lemmas and corollaries.

Lemma 4. *For any x and $-1 \leq k \leq n-x$, as r increases g_x^{x+k} decreases.*

Proof of lemma 4. Starting from the recurrent states:

$$g_x^x = \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)}$$
$$g_x^{x-1} = \frac{\delta q}{(1-\delta)(1-\delta+\delta r+\delta q)}$$

are both decreasing in r. For any higher state $x+k$ for $k \leq n-x$.

$$g_x^{x+k} = \frac{1}{1-\delta(1-r)} \frac{1-\left(\frac{r\delta}{1-\delta(1-r)}\right)^{k+1}}{1-\left(\frac{r\delta}{1-\delta(1-r)}\right)} + \left(\frac{r\delta}{1-\delta(1-r)}\right)^{k+1} \frac{(\delta q)}{(1-\delta)(1-\delta+\delta r+\delta q)}$$

Differentiating yields

$$\begin{aligned}\frac{\partial g_x^{x+k}}{\partial r} =& -k\frac{\delta(1-\frac{r\delta}{1-\delta(1-r)})}{(1-\delta(1-r))^2}(\frac{r\delta}{1-\delta(1-r)})^k - \frac{q\delta^2}{1-\delta}\frac{(\frac{r\delta}{1-\delta(1-r)})^{k+1}}{(1-\delta+\delta q+\delta r)^2}\\ &+\frac{kq\delta^2}{(1-\delta)(1-\delta+\delta r+\delta q)}\frac{\delta(1-\frac{r\delta}{1-\delta(1-r)})}{1-\delta(1-r)}(\frac{r\delta}{1-\delta(1-r)})^k \leq 0\end{aligned}$$

$\square$

Lemma 5. *For any $k > 1$ as r increases $U_a^e(k)$ increases.*

Proof of lemma 5. The proof is done by straight forward calculations, for any state c_k,

$$U_a^e(k) = \sum_{l=0}^{k-2} \left(\frac{r\delta}{1-\delta(1-r)} \right)^l \frac{v_e - c_{k-l}}{1-\delta(1-r)} + \left(\frac{r\delta}{1-\delta(1-r)} \right)^{k-1} \frac{v_e - c_1}{1-\delta} - F$$

Differentiating yields

$$\begin{aligned} \frac{\partial U_a^e(k)}{\partial r} =& \frac{1}{r^2\delta} \left[(k-1) \left((v - c_1) \frac{r\delta}{1-\delta+\delta r} \right)^k + r^2 \right] \\ &+ \sum_{l=0}^{k-2} r^{l-1}\delta^l \frac{v - c_{k-l}}{(1-\delta+\delta r)^{l+2}} (l(1-\delta) + r\delta) \geq 0 \end{aligned}$$

□

From lemma 5 it is easy to ascertain by a straightforward calculation, the following corollary.

Corollary 1. *For any x as r increases $p_p(c_x)$ increases.*

Lemma 6. *Holding the threshold constant, the expected profits of the principal decrease as r increases.*

Proof of lemma 6. Let the threshold level be c_x. First, observe that before due to monotone payments before at any cost level $c_k \geq c_x$ higher than the threshold the monotone payments to the agent for utilization is $v_e - c_l - (1-\delta)F$. But this also implies that above the threshold level the profits of the principal is independent of the cost level of the agent and equal to $v_o - v_e + (1-\delta)F$. Thus the principal's expected profits with monotone payments starting from any state $c_k \geq c_x$ is equal to:

$$f_x^k = g_x^k(v_o - v_e + (1-\delta)F) + \left(\frac{1}{1-\delta} - g_x^k \right) (w_o - p_p(c_x))$$

But from lemma 4 and corollary 1 we know that that g_x^k is decreasing in r, and $p_p(c_x)$ is increasing in r. Since $(v_o - v_e + (1-\delta)F) \geq (w_o - p_p(c_x))$ for any x, it must be the case that f_x^k is decreasing in r. □

Lemma 7. *Expected profits of the principal is weakly decreasing in r.*

Proof of lemma 7. Suppose not, then there exists $r' \geq r$ such that the profits of the principal at the beginning of the game with r' denoted by $f^n_{x^*(r')}(r')$ is greater than the expected profits of the principal at the beginning of the game with r, denoted $f^n_{x^*(r)}(r)$. Let $c_{x^*(r')}$ denote the optimal threshold with learning speed r' and similarly let $c_{x^*(r)}$ denote the optimal threshold with learning speed r. By assumption we must have

$$f^n_{x^*(r')}(r') > f^n_{x^*(r)}(r)$$

Now consider the expected payoffs of the principal with the suboptimal policy, that adopts the threshold $c_{x^*(r')}$ with learning speed r, denoted by $f^n_{x^*(r')}(r)$. Since $r' \geq r$ by lemma 6 it must be the case that:

$$f^n_{x^*(r')}(r) \geq f^n_{x^*(r')}(r') > f^n_{x^*(r)}(r)$$

But $c_{x^*(r)}$ is the optimal threshold for learning speed r, thus it must also be the case that:

$$f^n_{x^*(r)}(r) \geq f^n_{x^*(r')}(r) \geq f^n_{x^*(r')}(r') > f^n_{x^*(r)}(r)$$

leading to the desired contradiction. □

□

6.2 Multiple Sourcing Model

Fixing the payment rule to be the fastest prices, let $R_i(c^t_i) = v - c^t_i - p^t_i$. Similar to the sole sourcing model, consider the following relaxed version of the principal's problem.

$$\begin{aligned}
&\left[\max_{\{\{I^t_i\}_{i\in N}\}_{t\in\mathbb{N}}} E(\sum_{t=0}^{\infty}\sum_{i=1}^{N}\delta^t I^t_i R_i(c^t_i))\right] \\
\text{subject to } & E(\sum_{t=0}^{\infty}\sum_{i=1}^{N}\delta^t(1-I^t_i)) \geq 0 \\
& IC_P
\end{aligned} \tag{6.6}$$

Ignoring the incentive constraint of the principal, the Lagrangian for the relaxed problem is as follows

$$\max_{\{\{I_i^t\}_{i\in N}\}_{t\in\mathbb{N}}} \left[E(\sum_{i=1}^{N}\sum_{t=0}^{\infty}\delta^t(I_i^t R_i(c_i^t))|c^0) + \lambda\left(E(\sum_{i=1}^{N}\sum_{t=0}^{\infty}\delta^t(1-I_i(t))|c^0)\right)\right]$$

Rearranging the terms yield,

$$\max_{\{\{I_i^t\}_{i\in N}\}_{t\in\mathbb{N}}} E(\sum_{i=1}^{N}\sum_{t=0}^{\infty}\delta^t(I_i^t R_i(c_i^t) + (1-I_i(t))\lambda)|c^0) \tag{6.7}$$

Due to the linearity of expectations equation 6.7 can be further rearranged as follows:

$$\max_{\{\{I_i^t\}_{i\in N}\}_{t\in\mathbb{N}}} \left[\sum_{i=1}^{N} E(\sum_{t=0}^{\infty}\delta^t(I_i^t R_i(c_i^t) + (1-I_i(t))\lambda|c^0))\right] \tag{6.8}$$

The final iteration yields a relaxed restless bandit problem. Whittle (1988) has shown that this relaxation is solved optimally arm by arm by index policies if the problem is indexable. In particular, it is easy to see that the entire sum is going to be maximized if each individual summand is maximized. A single summand is a single arm restless bandit problem where passive rewards are equal to λ.

6.2.1 Single Arm Problems

Each of the summands in problem 6.8 is as follows:

$$\max_{\{I_i^t\}_{t\in\mathbb{N}}} E(\sum_{t=0}^{\infty}\delta^t(I_i^t R_i(c_i^t) + (1-I_i^t)\lambda)|c_i^0) \qquad (\lambda\text{-passive problem for agent i})$$

For this problem consider the following policies

Definition 6 (Monotone Policies)**.** *A policy is called monotone if* $\exists \hat{c}_i \in \{c_{i,1}, c_{i,2}, \ldots, c_{i,n_i-1}, c_{i,n_i}\}$ *such that for all* $x_i > \hat{c}_i \quad I_i(t) = 0$ *and for all* $z_i \leq \hat{c}_i \quad I_i(t) = 1$

A policy is called monotone if the agent is employed whenever his costs are below a level $\hat{c}_i$ and he is never employed if his costs are higher than $\hat{c}_i$.

Proposition 7. *Any Markovian policy has an equivalent monotone policy.*

Proof of proposition 7. Since definition is based on a single agent the notation regarding the agent is suppressed.

Observation 2. *Any Markov employment policy π can be identified by its active set S^π, such that $I^t = 1 \Leftrightarrow c^t \in S^\pi$.*

Let $c^0 = \hat{c}$ denote the initial state and consider any Markov employment policy π, identified with its active set S^π. Let $\underline{c_x} = \max\{c \in S^\pi : c \leq \hat{c}\}$ and let $\bar{c_x} = \min\{c \in C \setminus S^\pi : c \geq \hat{c}\}$. There are two possible cases

Case 1. $\hat{c} \in S^\pi$. *Under policy π for all t, $c^t \in \{\hat{c}, \ldots \bar{c_x}\}$. Moreover, for all t, $I^t = 1 \Leftrightarrow c^t < \bar{c_x}$. Now, consider the monotone policy, identified with $\bar{c_x}$ denoted by $P_{\bar{x}}$. Then, by definition under policy $P_{\bar{x}}$ for all t, $c^t \in \{\hat{c}, \ldots, \bar{c_x}\}$. Moreover, for all t, $I^t = 1 \Leftrightarrow c^t < \bar{c_x}$. Thus the two policies are equivalent.*

Case 2. $\hat{c} \notin S^\pi$. *Under policy π for all t, $c^t \in \{\underline{c_x}, \ldots, \hat{c}\}$. Moreover, for all t, $I^t = 0 \Leftrightarrow c^t > \underline{c_x}$. Now, consider the monotone policy, identified with $\underline{c_x}$ denoted by $P_{\underline{x}}$. Then, by definition under policy $P_{\underline{x}}$ for all t, $c^t \in \{\underline{c_x}, \ldots, \hat{c}\}$. Moreover, for all t, $I^t = 0 \Leftrightarrow c^t > \underline{c_x}$. Thus the two policies are equivalent.*

Thus for any Markov policy, starting from any initial state, there is an equivalent monotone policy. □

Similar to the previous section the single agent problems can be solved optimally via an index policy.

For $x < y$, let σ_x^y denote the time when an agent i who starts in state x at time 0, and who works every period reaches the state y. Formally:

$$\sigma_x^y = \inf\{t > 0 : (c_i^t) = y \text{ and } c_i^0 = x \text{ and } I^s = 1 \quad \forall s < t\}.$$

The expected waiting just before changing state $E(\delta^{\sigma_x^{x+1}-1})$ is given by

$$\begin{aligned} E(\delta^{\sigma_x^{x+1}-1}) &= \sum_{n=0}^{\infty} \delta^n (1-q)^n q \\ &= \frac{q}{1-\delta(1-q)} \end{aligned}$$

Similarly for state any state let γ denote the time it takes an agent to reinitialize from any state $c_k > c_1$ at time 0, who rests every period. Formally:

$$\gamma = \inf\{t > 0 : (c_i^t) = c_1 \text{ and } c_i^0 \neq c_1 \text{ and } I^s = 0 \quad \forall s < t\}.$$

The expected discounted waiting time before reinitializing $E(\delta^{\gamma-1})$ is given by

$$\begin{aligned} E(\delta^{\gamma-1}) &= \sum_{n=0}^{\infty} \delta^n (1-r)^n r \\ &= \frac{r}{1-\delta(1-r)} \end{aligned}$$

Let π_x denote a monotone policy, such that $I^t = 1 \Leftrightarrow c^t \geq c_x$. Let f_x^k denote expected discounted returns under policy π_x with initial state c_k. Similarly let g_x^k denote expected discounted utilization under policy π_x with initial state c_k. Formally:

$$\begin{aligned} f_x^k &= E(\sum_{t=0}^{\infty} \delta^t R(c^t|I^t)|\pi_x, c^0 = c_k) \\ g_x^k &= E(\sum_{t=0}^{\infty} \delta^t I^t|\pi_x, c^0 = c_k) \end{aligned}$$

Since monotone policies automatically induce a family of nested sets, utilizing Niño-Mora (2007), the marginal productivity index for any state c_x denoted $\lambda(c_x)$ for the relaxed problem can be readily computed as

$$\lambda(c_k) = \frac{f_x^x - f_{x-1}^x}{g_x^x - g_{x-1}^x}$$

The nested sets are decreasing now since the law of motion under the active action is reversed.

Finally, once again utilizing Wald's identity along with strong markov property the components of the index can be calculated as follows:

$$
\begin{aligned}
f_x^x &= \sum_{t=0}^{\sigma_x^{x+1}-1} \delta^t (v - c_x) + \delta^{\sigma_x^{x+1}+\gamma} \left[\sum_{n=0}^{\sigma_1^x - 1} \delta^n v - c_{n+1} + \delta^{\sigma_1^x} f_x^x \right] - (v - l) \\
&= \frac{v - c_x}{1 - \delta(1-q)} - (v - l) \\
&+ \frac{\delta r}{1 - \delta(1-r)} \frac{q\delta}{1 - \delta(1-q)} \left[\sum_{n=0}^{x-2} \frac{v - c_{n+1}}{1 - \delta(1-q)} \left(\frac{\delta q}{1 - \delta(1-q)} \right)^n + \left(\frac{\delta q}{1 - \delta(1-q)} \right)^{x-1} f_x^x \right] \\
&= \frac{\frac{v - c_x}{1-\delta(1-q)} + \frac{\delta r}{1-\delta(1-r)} \frac{q\delta}{1-\delta(1-q)} \left[\sum_{n=0}^{x-2} \frac{v - c_{n+1}}{1-\delta(1-q)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^n \right] - (v - l)}{1 - \frac{\delta r}{1-\delta(1-r)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^x}
\end{aligned}
$$

$$
\begin{aligned}
f_{x-1}^x &= \frac{\delta r}{1 - \delta(1-r)} \left[\sum_{n=0}^{\sigma_1^x - 1} \delta^n (v - c_{n+1}) + \delta^{\sigma_1^x} f_{x-1}^x - (v - l) \right] \\
&= \frac{\frac{\delta r}{1-\delta(1-r)} \left[\sum_{n=0}^{x-2} \frac{v - c_{n+1}}{1-\delta(1-q)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^n \right] - \delta(v - l)}{1 - \frac{\delta r}{1-\delta(1-r)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^{x-1}}
\end{aligned}
$$

$$
\begin{aligned}
g_x^x &= \sum_{t=0}^{\sigma_x^{x+1}-1} \delta^t 1 + \delta^{\sigma_x^{x+1}+\gamma} \left[\sum_{n=0}^{\sigma_1^x - 1} \delta^n 1 + \delta^{\sigma_1^x} g_x^x \right] \\
&= \frac{1}{1 - \delta(1-q)} \\
&+ \frac{\delta r}{1 - \delta(1-r)} \frac{q\delta}{1 - \delta(1-q)} \left[\sum_{n=0}^{x-2} \frac{1}{1 - \delta(1-q)} \left(\frac{\delta q}{1 - \delta(1-q)} \right)^n + \left(\frac{\delta q}{1 - \delta(1-q)} \right)^{x-1} g_x^x \right] \\
&= \frac{\frac{1}{1-\delta(1-q)} + \frac{\delta r}{1-\delta(1-r)} \frac{q\delta}{1-\delta(1-q)} \left[\sum_{n=0}^{x-2} \frac{1}{1-\delta(1-q)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^n \right]}{1 - \frac{\delta r}{1-\delta(1-r)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^x}
\end{aligned}
$$

$$g_{x-1}^x = \frac{\delta r}{1-\delta(1-r)} \left[\sum_{n=0}^{\sigma_1^x - 1} \delta^n 1 + \delta^{\sigma_1^x} g_{x-1}^x \right]$$

$$= \frac{\frac{\delta r}{1-\delta(1-r)} \left[\sum_{n=0}^{x-2} \frac{1}{1-\delta(1-q)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^n \right]}{1 - \frac{\delta r}{1-\delta(1-r)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^{x-1}}$$

6.2.2 Optimal Threshold

Given that each λ-passive problem has a solution in monotone policies, the final step is to identify which monotone policy is optimal for each arm. In that direction we first introduce the following notions

Definition 7 (Index policy augmented with threshold λ)**.** *An employment policy identified by indicies $\{\{\lambda_i(c_{i,x})\}_{c_{i,x} \in C_i}\}_{i \in N}$ and threshold λ is an index policy with threshold λ if*

$$I_i^t = 1 \Leftrightarrow \lambda_i(c_i^t) \geq \lambda \text{ and } \lambda_i(c_i^t) \geq 0$$

An index policy augmented with threshold λ chooses a monotone policy for each arm in 6.7 since the identified indices arise from the decoupling of problem 6.7 the selection of a single threshold is without loss of generality.

Definition 8 (Workload limit with threshold λ)**.** *For any index policy augmented with threshold λ, the workload limit for an agent (when it exists) is given by*

$$c_i(\lambda) = max\{c_i \in \{c_{i,1}, c_{i,2}, \ldots, c_{i,n_i-1}, c_{i,n_i}\} : \lambda_i(c_i) \geq \lambda\}.^4$$

Here it is helpful to introduce the following notation as well, $k(\lambda)$ denotes the ordinal level of $c_i(\lambda)$, i.e. $c_i(\lambda) = c_{i,k(\lambda)}$ where $k(\lambda) \in \{1, 2, \ldots, n_i\}$. The workload limit for an agent indicates the highest level of costs that the agent is employed at a given threshold λ. If the workload limit does not exist for an agent at some threshold, then that agent is never employed with a threshold λ. Furthermore it is easy to see that for each arm $c_i(\lambda)$ is decreasing in λ.

[4] Notice I used max here since indices are decreasing in state.

For identifying the expected discounted total utilization of a single arm, we return back to the components of the index.

$$\begin{aligned} g^1_{i,c_i(\lambda)} &= E(\sum_{t=0}^{\infty} \delta^t I^t | I^t = 1 \Leftrightarrow c_i^t \leq c_i(\lambda)) \\ &= \sum_{n=0}^{x-2} \frac{1}{1-\delta(1-q)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^n + \left(\frac{\delta q}{1-\delta(1-q)} \right)^{x-1} g^{c_i(\lambda)}_{i,c_i(\lambda)} \end{aligned}$$

Observe that for each arm since the workload limits are decreasing as the threshold decreases, the sum of expected utilizations denoted as

$$\sum_{i \in N} g^1_{i,c_i(\lambda)} = E(\sum_{t=0}^{\infty} \sum_{i \in N} I_i^t)$$

is also decreasing in λ in a monotone fashion. Thus by starting from a very low λ and slowly increasing λ it is possible to make sure that the constraint

$$E(\sum_{t=0}^{\infty} \sum_{i \in N} I_i^t) \leq 1/(1-\delta)$$

is satisfied, with one caveat, there might be a need to randomize for one agent at the workload limit for a given λ if the constraint is to be satisfied with equality since for a fixed discount factor the utilizations make discrete jumps as λ changes. There are two potential cases that needs to be considered for finding the optimal λ and whether the constraint will be satisfied with equality, which ties the problem back to the principal's incentive constraint. Again to identify the connection between the indices and the principal's incentive constraint, we return back to another component of the index, where

$$\begin{aligned} f^{c_i(\lambda)}_{i,c_i(\lambda)} &= E(\sum_{t=0}^{\infty} \delta^t R_i(c_i^t) I^t | I^t = 1 \Leftrightarrow c_i^t \leq c_i(\lambda)) \\ &= \frac{\frac{v-c_i(\lambda)}{1-\delta(1-q)} + \frac{\delta r}{1-\delta(1-r)} \frac{q\delta}{1-\delta(1-q)} \left[\sum_{n=0}^{k(\lambda)-2} \frac{v-c_{n+1}}{1-\delta(1-q)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^n \right] - (v-l)}{1 - \frac{\delta r}{1-\delta(1-r)} \left(\frac{\delta q}{1-\delta(1-q)} \right)^{k(\lambda)}} \end{aligned}$$

Notice that for any $m \in \{1, 2, \ldots, k(\lambda)\}$ we also have $f_{i,c_i(\lambda)}^{c_i(\lambda)-m} \geq f_{i,c_i(\lambda)}^{c_i(\lambda)}$. Thus the only point where the IC of the principal binds is at the workload limit. However observe that by definition since indices are restricted to be positive, $f_{i,c_i(\lambda)}^{c_i(\lambda)} \geq 0$ whenever $I_i^t = 1$, thus the principal's IC will always be satisfied. With this case ruled out the optimal threshold is purely driven by the total utilization.

Observation 3. *Optimal threshold λ^* is the smallest $\lambda \in \mathbb{R}$ such that*

$$\sum_{i \in N} g_{i,c_i(\lambda)}^1 \leq E(\sum_{t=0}^{\infty} \sum_{i \in N} I_i^t)$$

Due to the discrete nature of the $g_{i,c_i(\lambda)}^1$'s some randomization at exactly the threshold level will be necessary, however it is easy to see if the agents are not identical, only one agent will ever be randomizing.

References

Andrews, I., and D. Barron (2013): "The allocation of future business: Dynamic relational contracts with multiple agents," *American Economic Review*.

Arrunada, B., and X. H. Vázquez (2006): "When your contract manufacturer becomes your competitor," *Harvard business review*, 84(9), 135.

Baker, G., R. Gibbons, and K. J. Murphy (2002): "Relational Contracts and the Theory of the Firm," *Quarterly Journal of economics*, pp. 39–84.

Blackwell, D. (1965): "Discounted Dynamic Programming," *The Annals of Mathematical Statistics*, 36(1), pp. 226–235.

Board, S. (2011): "Relational contracts and the value of loyalty," *The American Economic Review*, pp. 3349–3367.

Gittins, J., K. Glazebrook, and R. Weber (2011): *Multi-armed bandit allocation indices*. John Wiley & Sons.

Gittins, J. C. (1979): "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 148–177.

Glazebrook, K., D. Hodge, C. Kirkbride, et al. (2013): "Monotone policies and indexability for bidirectional restless bandits," *Advances in Applied Probability*, 45(1), 51–85.

Glazebrook, K., C. Kirkbride, and D. Ruiz-Hernandez (2006): "Spinning plates and squad systems: Policies for bi-directional restless bandits," *Advances in applied probability*, pp. 95–115.

Glazebrook, K., J. Nino-Mora, and P. Ansell (2002): "Index policies for a class of discounted restless bandits," *Advances in Applied Probability*, pp. 754–774.

Glazebrook, K., D. Ruiz-Hernandez, and C. Kirkbride (2006): "Some indexable families of restless bandit problems," *Advances in Applied Probability*, pp. 643–672.

Gray, J. V., B. Tomlin, and A. V. Roth (2009): "Outsourcing to a Powerful Contract Manufacturer: The Effect of Learning-by-Doing," *Production and Operations Management*, 18(5), 487–505.

Jacko, P. (2009): "Marginal productivity index policies for dynamic priority allocation in restless bandit models," .

——— (2011): "Optimal index rules for single resource allocation to stochastic dynamic competitors," in *Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools*, pp. 425–433. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Levin, J. (2002): "Multilateral contracting and the employment relationship," *Quarterly Journal of economics*, pp. 1075–1103.

——— (2003): "Relational incentive contracts," *The American Economic Review*, 93(3), 835–857.

Malcomson, J. M., et al. (2010): *Relational incentive contracts*. Department of Economics, University of Oxford.

McCoy, M. (2003): "Serving emerging pharma," *Chemical & engineering news*, 81(16), 21–33.

Nino-Mora, J. (2002): "Dynamic allocation indices for restless projects and queueing admission control: a polyhedral approach," *Mathematical programming*, 93(3), 361–413.

Niño-Mora, J. (2007): "Dynamic priority allocation via restless bandit marginal productivity indices," *Top*, 15(2), 161–198.

Nino-Mora, J., et al. (2001): "Restless bandits, partial conservation laws and indexability," *Advances in Applied Probability*, 33(1), 76–98.

Pandya, E., and K. Shah (2013): "CONTRACT MANUFACTURING IN PHARMA INDUSTRY.," *Pharma Science Monitor*, 4(3).

Papadimitriou, C. H., and J. N. Tsitsiklis (1999): "The complexity of optimal queuing network control," *Mathematics of Operations Research*, 24(2), 293–305.

Plambeck, E. L., and T. A. Taylor (2005): "Sell the plant? The impact of contract manufacturing on innovation, capacity, and profitability," *Management Science*, 51(1), 133–150.

Puterman, M. L. (2014): *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Rajaram, S. (2015): "Electronics Contract Manufacturing and Design Services: The Global Market," Discussion paper, BCC Research.

Serfozo, R. (2009): *Basics of applied stochastic processes*. Springer Science & Business Media.

Thompson, P. (2010): "Learning by doing," *Handbook of the Economics of Innovation*, 1, 429–476.

Tully, S. (1994): "Youll never guess who really makes," *Fortune*, 130(7), 124–128.

Whittle, P. (1988): "Restless bandits: Activity allocation in a changing world," *Journal of applied probability*, pp. 287–298.