The Misaligned Incentives of Temporary Work Agencies and their Client Firms

Morgan Westéus  Tomas Raattamaa
Department of Economics  Department of Economics
USBE  USBE
Umeå university  Umeå university

Abstract
This paper adds to the theoretical literature on the incentives of Temporary Work Agencies (TWAs). Using a principal-agent model with hidden action to analyse two main types of contracts between a TWA and a Client Firm (CF), the TWA is shown to potentially act against the best interest of the CF when helping to fill a vacant position. The results also suggest that the adverse effect of the incentive misalignment is larger when workers are leased rather than hired by the CF. However, this effect could potentially be offset by introducing a sufficient level of competition among TWAs.

Acknowledgements: The authors wish to thank Thomas Aronsson, Magnus Wikström, Karl-Gustaf Löfgren, Christian Lundström (USBE) and Erik Geijer (SLU/CERE), discussants Tomas Sjögren, Stefanie Heidrich (both USBE), and Mattias Lundbäck (RATIO), as well as the seminar participants at the Department of Economics and at the RATIO institute for a gainful discussion and their insightful comments. The paper has also benefited from comments given by Ann-Sofie Kolm (SU) and the other participants at the Swedish National Conference in Economics 2014.
1 Introduction

The number of workers employed through a temporary work agency (TWA), either as a consultant employed by the TWA, or screened by the TWA before being hired by the client firm (CF), has risen substantially in the last decades (Andersson-Joona & Wadensjö, 2010; Forde & Slater, 2005). While it might seem as though the interests of the CF and TWA are aligned, i.e. one pays to fill a vacant position while the other gets paid to fill said position, we show that this is not necessarily the case (cf. Gibelman, 2005).

By outlining a principal-agent model with hidden action we are able to shed light on some important consequences of the differing incentive structures between the CF (principal) and the TWA (agent); among other things, the CF wants to lease the best possible worker for a given position, while the TWA may want to provide the least productive worker possible who is still sufficiently good at his/her job.

The TWA is usually assumed to have some advantage(s) in the recruitment process through which it is argued to be able to supply a worker more quickly than the CF would be able to, and/or hedge certain liabilities of the CF towards the contingent TWA worker (Autor, 2001, 2003; Baumann, Mechtel, & Stähler, 2011; ECORYS-NEI, 2002; Houseman, 2001; Mitlacher, 2007; Neugart, 2005; Westéus, 2014). The CF’s primary use of TWAs is thereby to either delegate search, i.e. identify prospective candidates for a vacant position and assess their productivity, and/or be able to (quickly) lease a worker to fill a vacant position (Beckmann & Kuhn, 2012; Forde, MacKenzie, & Robinson, 2008). Our principal-agent model therefore investigates two primary contract types: Recruitment Contracts and Recruit-and-Rent Contracts.

While not having been used in exactly this context, the principal-agent literature dates back to the 1970’s where Spence and Zeckhauser (1971) developed the first model, focusing on insurance. These types of models have since then been used to answer questions in various fields (see e.g. Miller, 2005, for a review). Bendor, Glazer, and Hammond (2001) offer a basic introduction to delegation, and Lewis (2012) contains a review of recent studies on delegated search. Milner and Pinker (2001) considered two types of problems; the first being when the productivity of a temporary worker is difficult to evaluate and subsequently the TWA is used for screening purposes, and the second one entailing the impossibility of creating a socially optimal labour supply contract between a CF and a TWA under hidden action, when the demanded quantity of sufficient, uniformly productive workers is uncertain when the unit price is negotiated.

Hidden action in this context implies that by only being able to observe the supplied worker(s) and not the process (cf. the findings in Beckmann & Kuhn, 2012; Connell & Burgess, 2002), the CF is only able to assess who they get, but not who the alternative workers were (cf. Gibelman, 2005), as any match is only privately observable by the TWA unless supplied to the CF (cf. Halac, Kartik, & Liu, 2016). This allows the TWA to match workers in a way that might not be optimal for the CF, and these differences have mostly been neglected in previous studies, even
though they are potentially highly influential in determining the type of worker that the TWA will supply. One exception is Postl (2004), who found that when an agent is given two alternatives to evaluate, he may only have the incentive to evaluate one and then lie about the quality of the other, resulting in an efficiency loss, as the principal would base its decision on (possibly) incorrect information. Additionally, this aspect of the matching problem has been overlooked in other studies where the agent is contracted to evaluate and provide information on the quality of an already available alternative, or a stream of alternatives, under hidden action (see e.g. Chade & Kovrijnykh, 2011).

This paper expands on the reasoning of Postl (2004) and related papers, but extends the model by letting the TWA itself search for any number of alternatives and also by removing the assumption of a fixed search cost per alternative. This allows us to carry out a more thorough investigation into the outcome of the incentive misalignment resulting from hidden action.

The present paper adds to the existing literature by providing a complementary (or perhaps even an alternative) and structural explanation for the ability to quickly match a worker to a vacant position, which in previous studies has been assumed to be the result of some superior matching technology (Baumann et al., 2011; Neugart, 2005). The results, similar to Beckmann and Kuhn (2012), suggest that CFs should use TWAs to only screen applicants (which are then employed directly at the CF), rather than continuously leasing the workers. We also discuss whether a logical extension of our theoretical results could provide an alternate (demand side) explanation to the increased level of education among the workers in the Swedish temporary work agency sector between 1999 (Joona & Wadensjö, 2008) and 2007 (Westéus, Raattamaa, & Lindgren, 2016), in contrast to the (supply side) rationale offered by Walter (2012).

Furthermore, these results are independent of the price level of the TWAs’ services (see Baumann et al., 2011; Neugart, 2005; Westéus, 2014), in that they only require the existence of a price that the CF is willing to pay and the TWA is willing to accept. This paper therefore does not need to consider optimal pricing for the TWAs’ services. The results instead rely on the assumptions of asymmetric information¹ and an imperfect labour market where the offered wage is related to the vacant position rather than the productivity of the matched worker.

The body of research on the productivity of temporary (agency) employees uses different productivity measures and yields somewhat inconclusive results (compare e.g. Beckmann & Kuhn, 2012; Kleinknecht, Oostendorp, Pradhan, & Naastepad, 2006) and has not considered the potential incentive misalignment suggested in this paper (see Hirsch & Mueller, 2012; Nielen & Schiersch, 2011). This is why it will be up to future research to measure the importance of our contribution. To our knowledge, there are no studies on the performance of workers on fixed-term contracts employed directly at the CF, relative to temporary agency workers at the same firm who perform the same type of jobs. Our model therefore does not fully

¹The CFs believe the TWAs to be superior in some aspect(s), making the CFs willing to use the TWAs’ services.
support or reject any of the results in previous studies from an individual worker productivity aspect, but rather it emphasises the difference in the type of worker the TWA might supply. It has similarly been argued by Walter (2012) that TWAs have an incentive to be able to continuously lease their workers, and therefore they might not only match for the specific traits requested by the CF.

The paper is outlined as follows: section two outlines the model for the two aforementioned main types of activities; a Recruitment Contract, or a Recruit-and-Rent Contract. The Recruitment Contract implies that the CF always employs the matched worker directly at the firm whereas the Recruit-and-Rent Contract allows the CF, at each point in time, to make a choice of either subsequently leasing or directly employing the matched worker\(^2\). Initially the analysis is concerned with the outcome of different types of contracts between the CF and a single TWA, whereas the last part of the analysis is concerned with the effects when there are several competing TWAs. The main results of the model are thereafter summarised, after which the final chapter contains a longer discussion of the models assumptions and implications.

### 2 Model

The model consists of two types of risk-neutral actors: a CF and one or more external recruitment agencies (TWAs), where the former is defined as any firm having established that there is a demand for an additional worker (i.e. a vacancy) that will be matched by a TWA.

For any vacancy we assume that there are \(J\) possible applicants and each individual \(j\) is identified by his/her unique productivity level, \(x_j\),\(^3\) where the set of available workers is assumed to follow a uniform distribution: \(X = \{x_1, ..., x_J\} \sim U[x, x]\).\(^4\)

For each vacancy we assume that there is an objective (i.e. true) exogenous\(^5\) minimum productivity required: \(x^*\), and that the position pays a fixed wage: \(w\), to the worker once filled.\(^6\) The analysis is therefore delineated to when there is a proper non-empty subset \(\chi = \{x \in X | x \geq x^*\}\) containing \(K < J\) elements where each individual has a unique \(x_k\). This allows us to ignore the special cases when there are either no applicants at all, or when only unqualified workers will apply (mar-

\(^2\)The choice to employ the worker is final, and the CF may thereafter not lease the worker.

\(^3\)For simplicity, we model the productivity as a scalar, but it could also be modeled as a multi-dimensional vector. The productivity vector would then consist of all possible traits that a worker may have (e.g. preferences on commute distance, age, education, previous job experience, family situation etc.) with a complete set of marginal rates of substitution between every pair of traits.

\(^4\)This simplification is done to keep the mathematics as simple as possible and not divert from the qualitative implications of the model.

\(^5\)We discuss the implications of endogenising this parameter at the end of the section.

\(^6\)We argue that this construct is empirically relevant. Optimal marginal wage-setting requires the employer to be able to estimate the individual’s marginal productivity which, outside of a perfect labour market with either piece-work pay or low-cost alternative employment opportunities that allow the workers to self-select, is often quite hard - or even impossible.
ket of lemons). It also implies that the TWA is expected to (asymptotically) be able to identify a worker with a sufficient productivity; \( x_m \in \mathcal{X} \), which is denoted as a *match*. We define \( x_L \) (\( x_H \)) as a subsequent match that has a lower (higher) productivity than the first match, while still being sufficient, i.e. \( x^* \leq x_L < x_m < x_H \).

The shape and size of the distribution of workers and its subset is assumed to be known by the TWA, due to its specialisation in creating matches between vacancies and job seekers, but for the same reasons it is assumed to be unknown by the CF. The CF is also either incapable, or unwilling, to monitor any search effort by the TWA other than the actual output: the productivity of the supplied worker.

The model is outlined in discrete time \( t \in (0, 1, ..., T) \) where only one worker may be evaluated at each sub-period. The model also includes a perfect credit market with interest rate \( r \). Following prior simplifications, the constant probability of finding a match at each point in time becomes \( p = p(x_m \geq x^*) = \frac{K}{T} = \frac{x^*-x}{x-x} \).

This allows the cumulative probability that a match will be found within a given amount of search periods to asymptotically approach one.\(^7\) For each time period the TWA searches, it incurs a constant cost \( c \).

Finally, a necessary (but not sufficient) condition for a TWA to accept the assignment is that the expected present value stream of payments from the CF, after netting off the TWA’s expected accumulated search cost, is non-negative. Because the model is not concerned with the explicit pricing of the TWA’s services, this condition will always be fulfilled by assumption, so that we are able to focus on the strategic search behaviour of the TWA.

### 2.1 Recruitment Contract

#### 2.1.1 Continuous Payment

A continuous payment contract stipulates that the CF will pay the TWA a fixed amount \( \phi \) at each time period until a sufficient worker is supplied, or the contract expires at time \( T \). The TWA’s present value decision rule for accepting the recruitment assignment can be written as:

\[
\pi = \sum_{t=0}^{\Gamma} \left\{ \frac{\phi}{(1+r)^t} \right\} - e \sum_{t=a}^{b} \left\{ \frac{c}{(1+r)^t} \right\} \geq 0, \tag{1}
\]

where \( e \in (0, 1) \) is the TWA’s discrete decision whether to exert effort or not over the time interval \([a, b] \subset [0, T] \). \( \Gamma \in [1, T] \) is when \( x_m \) is presented to the CF, or when the contract expires.

The assumption of hidden action makes the CF unable to monitor the actual (search) activity of the TWA. This removes the incentive for the TWA to carry out any search at all and thus the TWA always chooses \( e = 0 \) to maximise its

---

\(^7\)The search duration will follow a geometric distribution due to the fixed probability of finding a match at each turn. We assume that there are either enough applicants, or that individuals may enter and leave the set which makes the best approximation of the probability of finding a match to be constant over time.
profit. This payment scheme thereby does not create any incentive for the TWA to actually carry out any search.

The observant reader might notice that $G$ would be stochastic if the TWA would have incentives to reveal a match prior to $T$. This however will not be the case, as the profit maximising strategy of the TWA does not include any search activity at all. It is not the sequential nature of the contract that drives this result, but rather that payment is not conditioned on the CF actually being supplied a sufficient worker. Conditioning payment on delivery thereby becomes necessary for any effort to be exerted at all.

### 2.1.2 Payment on Delivery

In order to incentivise the TWA to search, the CF could offer a contract where a fixed payment $\phi$ is made when $x_m$ has been supplied. This payment-on-delivery contract ensures that upon accepting the contract, the TWA will start searching immediately, yet it also implies that the TWA will receive the same payment for supplying any $x_j \in \chi$. Thus there is no reason for the TWA to continue searching for $x_H$, after having found $x_m$, as any additional search effort will both increase the TWA’s costs and decrease its present value revenue.\(^8\) The optimal strategy by the TWA is therefore to immediately deliver its first match, whose expected productivity will be $E(x_m) = \frac{x^* + x^\ast}{2}$.

**Proposition 1.** The TWA will never attempt to find an alternative candidate after having found the first match.

This behaviour could create a different outcome compared to the option that the CF would have preferred, as the TWA will stop searching even if there were resources left to conduct additional search.\(^9\) If the CF had perfect monitoring, and thereby a better ability to enforce continuous effort, then the first match would still have a random (but still sufficient) productivity. However any additional resources could then be spent on finding an even better worker until the expected marginal cost of additional search would surpass its expected additional benefit. We define this as searching for the *marginally most productive worker*, whom would have the expected productivity $E(\tilde{x}_m) = \frac{x^* + x^\ast}{2} + \gamma$, where $\gamma \geq 0$ is the aforementioned productivity difference which is determined by how much additional search would have been profitable for the CF given the search cost and the residual probability of finding a more productive worker.

---

\(^8\)The model will later be extended to relate the productivity of the worker to the expected revenue of the TWA.

\(^9\)If the TWA (by chance) would find and supply the most productive worker in the population ($x_m = \bar{x}$), then matching is efficient even for the CF (as engaging in additional search would have a zero probability of finding a better worker). The same is true when the expected productivity gain of additional search is not expected to surpass its costs, had the CF carried out the search itself with the information and cost structure of the TWA. We will however focus our attention on when the best worker is not (necessarily) found right away and where additional search would have been an option.
The search decision for this type of contract is sequential in nature and denoting $\Xi_t$ as the set of draws prior to $t = \tau$ the two conditions for search in the subsequent time period $\tau$ become:

$$\Xi_t \cap \chi = \emptyset$$

$$p\phi - c (1 + r) \geq 0$$  \hspace{1cm} (2)

Equation (2) means that the TWA has not found a sufficient worker in the previous period, and Equation (3) is a marginal condition saying that any additional search effort must have a positive expected profit\(^{10}\).

It does not matter if the TWA expects the CF to terminate the contract at some point\(^{11}\), because even though this would set an upper bound on search activity, no incentives will be altered as payment is immediate upon delivery. Implications of contractually specifying a fixed point in time when payment will be made, conditioned on delivery, will be discussed in the next subsection.

\subsection*{2.1.3 Payment at a Pre-specified Point in Time, $T$, Conditioned on Delivery}

If the contract is designed in such a way that payment offered by the CF is made at time $T$, conditional on the TWA supplying $x_m$ at or prior to time $T$, then the TWA must consider its probability of finding $x_m$ over time. The present value at time $t = 0$ of the accumulated cost from searching an arbitrary number of periods $[a, b] \in [0, T]$ can be expressed as\(^{12}\):

$$C(s) = \sum_{t=a}^{b} \left\{ \frac{c}{(1 + r)^t} \right\}$$ \hspace{1cm} (4)

Denoting the fixed payment the TWA will obtain at $T$ if $x_m$ is supplied to the CF as $\phi$, the present value expected revenue at time $t = 0$ from searching up to $s$ time periods is:

$$E[R(s)] = \left[ 1 - (1 - p)^s \right] \left\{ \frac{\phi}{(1 + r)^s} \right\}$$ \hspace{1cm} (5)

where the first term signifies the cumulative probability that a match will have been found during the $s$ periods of searching.

If there is some exogenous accumulated-cost constraint making the TWA unable to fund search during all periods $t \in (0, ..., T)$, then searching will be deferred towards $T$, to maximise the constrained probability of delivery. Given a positive interest rate and denoting the maximum number of time periods the TWA will afford to search as $\bar{s}$ if starting to search immediately, and $\tilde{s}$ if searching towards the end, we have that $\bar{s} \leq \tilde{s}$ for any contract where $\bar{s} < T$\(^{13}\), as any given search effort will have a lower present value cost the later it is expended. When the

\(^{10}\)This condition will always be fulfilled, as the contract has been accepted.

\(^{11}\)E.g. that the TWA is not supplying a match within a “reasonable” amount of time.

\(^{12}\)The effort parameter $e$, previously used in Equation (1), is dropped in (4) as the present value of accepting a contract, and thereafter exerting no effort, will always be zero.

\(^{13}\)The first inequality is not strict due to the discrete nature of the model.
TWA has the option to choose when to search during the lifetime of the contract \( t \in (0, 1, ..., T) \), it is clear that the TWA will prefer to exert search efforts later rather than sooner in order to maximise its present value expected profit.

**Proposition 2.** The TWA will find it optimal to defer searching due to discounting if \( \bar{s} < T \).

Given that the TWA has found \( x_m \) prior to \( T \), the design of the contract makes the TWA indifferent between delivering \( x_m \) immediately, or waiting an arbitrary amount of time\(^{14} \), since the payment is independent of when \( x_m \) is presented. Moreover, this aspect is not optimal for the CF, as it can hardly be worse off by being presented with the (first and final) match as soon as it has been discovered – especially since the first actual day on the job can be subsequently negotiated between the CF and the matched worker.

The model will now be extended to also include the option for the CF to either employ the worker directly, or lease the worker through the TWA. This follows e.g. Houseman (2001) and ECORYS-NEI (2002) in that the main rationale for utilising TWAs is to screen the productivity of a worker before making the decision to either employ the worker, end the collaboration, or continue to lease the worker. It also follows the transitory pattern between temporary agency workers and workers on standard employment contracts in Westéus (2014), Baumann et al. (2011) and Neugart (2005).

### 2.2 Recruit-and-Rent Contracts

#### 2.2.1 A Single TWA

In this setup we assume that any supplied worker is initially always employed by the TWA and leased to the CF on an open-ended contract. At each subsequent point in time, the CF may either choose to continue to lease the worker, or employ the worker directly at the CF which will end the collaboration with the TWA. This corresponds to the TWA assuming all liabilities when the CF screens the productivity of the worker (see Neugart, 2005; Westéus, 2014).

The probability that a sufficient worker will become employed directly at the CF is defined as the worker’s *transition probability*, and is assumed to be an increasing function of the worker’s residual productivity: i.e., \( f(x_i) = \bar{f}(x_i - x^*) \in [0, 1] \), with the added simplification that it is assumed to be constant over time. This assumption implies that a more productive worker may leave as a result of obtaining a better offer elsewhere (as argued by Walter, 2005), or will have a greater chance of obtaining employment directly at the CF than a less productive worker. The duration of the open-ended contract is \( T \) periods, where \( T \sim Geo[f(x_i)] \) due to the assumed constant transition probability. This implies that the expected duration of the contract for a worker with productivity \( x_i \) can be expressed as

\(^{14}\text{But no longer than until } T.\)
\[ \mathcal{T}(x_i) = \frac{1}{f(x_i)} = N(x_i) + n(x_i); \text{ where } N(x_i) = \left\lfloor \frac{1}{f(x_i)} \right\rfloor \text{ is the integer part, and } n(x_i) = \frac{1}{f(x_i)} - N(x_i) \text{ is the fractional part.} \]

The accumulated search cost for a TWA searching from \( t = 0 \) to \( \hat{s} \) when the match \( x_m \) is found becomes:

\[ C(\hat{s}) = \sum_{t=0}^{\hat{s}} \left[ c(1+r)^t \right] \tag{6} \]

When the CF leases the worker, the TWA charges a wage-proportional fee \( \sigma \cdot w^* \) where \( \sigma > 1 \). The TWA in turn pays the worker \( \delta \cdot w^*; \delta < \sigma \), resulting in a revenue of \( (\sigma - \delta) \cdot w^* \) for the TWA.\(^{15}\)

Assuming that the profit from the fractional part is always incurred in the last time period, the expected present value profit for the TWA at time \( t = \hat{s} \) becomes:

\[ E(\pi \mid x_m) = -C(\hat{s}) + \left[ \frac{N(x_m)}{(1+r)^\hat{s}} + \frac{n(x_m)}{(1+r)^\mathcal{T}(x_m)} \right] (\sigma - \delta) \cdot w^* \tag{7} \]

The above definition of the expected duration, \( \mathcal{T}(x_i) \), of a contract states that a worker \( x_H(x_L) \) expected to transition to employment directly at the CF faster (slower) than the current match as \( f(x_H) > f(x_m) > f(x_L) \), and consequently \( \mathcal{T}(x_H) < \mathcal{T}(x_m) < \mathcal{T}(x_L) \). This implies that the TWA can only expect to compensate for any additional search costs by finding a worker with a lower productivity: \( x_L \), for which the TWA expects to be able to collect its fee for a longer period of time.\(^{16}\)

In order to provide comparable expression for the difference between the expected values of two non-linear stochastic processes, we utilise the short duration of temporary assignments (see e.g. Forde & Slater, 2005) and apply the limit argument to the non-linear term (i.e., allow the interest rate to approach zero). Furthermore, in order to keep the notation simple we will only explicitly model the case when \( \mathcal{T}(x_L) - \mathcal{T}(x_m) \geq 1 \) and \( n(x_\theta) \equiv 0 \) for \( \theta = (L, m) \) as this setup relates to the discrete nature of the model the most.

Since the productivity parameter is assumed to follow a uniform distribution, the conditional probability of finding a less productive match becomes \( q_L = q(x^* \leq x_L < x_m) = \frac{(x_m-x^*)-x+1}{x^*-x+1} \). The linearised expressions for the expected additional revenue and cost for the TWA associated with additional search can thereby be written as:

\[ \lim_{r \to 0^+} E(\Delta R) = \lim_{r \to 0^+} \left[ q_L \sum_{\tau=\mathcal{T}(x_m)}^{\mathcal{T}(x_L)} \frac{(\sigma - \delta) \cdot w^*}{(1+r)^\tau} \right] = q_L \left[ \mathcal{T}(x_L) - \mathcal{T}(x_m) \right] (\sigma - \delta) \cdot w^* \tag{8} \]

\(^{15}\)A number of studies find \( \delta < 1 \), however the EU Temporary and Agency Workers Directive (2008/104/EC) intends to ensure \( \delta = 1 \) (Westéus, 2014). Following Westéus (2014), Baumann et al. (2011) and Neugart (2005) we further assume that \( \sigma > 1 \) since paying the mark-up corresponds to a liability insurance for the CF.

\(^{16}\)If \( x_L \) is not found, then the TWA may still supply \( x_m \).
\[
\lim_{r \to 0^+} \Delta C = \lim_{r \to 0^+} \frac{c}{(1+r)} = c
\]  

**Proposition 3.** Using the above simplifications, additional search will be profitable for the TWA if
\[
q_L \left[ \hat{T} (x_i) - \hat{T} (x_m) \right] (\sigma - \delta) w^* \geq c.
\]
However, any additional search will always be for a worker with a lower, but still sufficient, productivity.

The above proposition shows that there are situations where the TWA will have incentives to actively act against the best interest of the CF. We denote the resulting misaligned-incentive induced expected productivity level difference as \( \kappa \).

Following Proposition 1, the expectation will go from \( E (\hat{x}_m) = \frac{x + x^*}{2} \) to \( E (\hat{x}_m) = \frac{x + x^*}{2} - \kappa \), where \( \kappa \geq 0 \) when using a Recruit-and-Rent Contract instead of a Recruitment Contract.

Assuming that TWAs have superior search capabilities, compared to the CF (similar to Baumann et al., 2011; Neugart, 2005), then this would further increase the size of the difference in expected match quality because better matching technology would allow the TWA to screen a larger number of workers for the same amount of resources.

The difference may be mitigated by introducing an opportunity cost for any TWA that finds a match but does not supply him/her to the CF. This can be done by increasing the number of TWAs competing for the assignment, and stating that only the TWA with the most productive match will get to supply the worker to the CF.

### 2.2.2 Incentives Caused by Competition Among TWAs

In this setup we assume that the CF has engaged \( z \) TWAs to find a match for the vacant position. We also assume that there are significantly fewer TWAs than applicants (\( z \ll J \)) in order to maintain the delineation to only analyse situations where there is actual strategic search behaviour on behalf of the TWAs when searching for a match to the vacant position (\( x^*, w^* \))\(^{17}\). Each search assignment resembles a Bertrand game to some extent, as we assume that only the first TWA to find a match (and in the case of several TWAs making a match in the same period, then the best match) will get to supply the entire worker demand (fixed at one).

To facilitate the analysis we simplify by assuming that the contract duration for a (sufficient) worker is defined by the linear function: \( \hat{T} (x_i \mid x_i \geq x^*) = T (x^*) - \beta (x_i - x^*) \), where \( T (x^*) \) is the maximum duration of the lease contract. We also define \( \hat{T} (\bar{x}) = \alpha \geq 1 \) and express the integer part and fractional part of the expected duration as before: \( \hat{T} (x_i) = N (x_i) + n (x_i) \).

We additionally make the assumption that each TWA makes the assessment that any other TWA will always present any match in the same period that the worker

\(^{17}\)As \( z \to \infty \) we would expect the best worker in the sample to always be found within one period of search and no strategic behaviour on behalf of the TWA could influence the outcome.
is found, and thus cannot expect any strategic search behaviour from its competitors. This highly restrictive assumption simplifies the model by allowing us to disregard any feedback effects among the competing TWAs. It also minimises the TWA’s incentive not to present a match when found.

Assuming that the accumulation of search costs, and the dynamics of how the TWA expects the CF to lease the supplied worker, follows the outline in the preceding subsection, the expected present value profit at time \( t = \hat{s} \) for any of the \( z \) TWAs having found a match, \( x_m \) at that point in time becomes:

\[
E (\pi \mid x_m, z) = - C (\hat{s}) + g (q_m) \cdot (\sigma - \delta) w^* \cdot \left[ \sum_{t=0}^{N(x_m)} \frac{1}{(1 + r)^t} + \frac{n (x_m)}{(1 + r)^T(x_m)} \right]
\]

(10)

where \( g (q_m) = \left( \frac{x_m - (x - 1)}{\hat{s} - x} \right)^{z-1} \) is the probability that none of the other TWAs have found an even more productive match than \( x_m \).

Finding \( x_H \) when conducting additional search will decrease the TWA’s expected lease time. However, \( x_H \) will also increase the overall chance of winning the contract – which is a necessary (but not sufficient) condition for obtaining any profit at all. As before, finding a less productive worker will increase the expected lease time for the TWA, conditional on winning the contract, but will now also reduce the probability that the given TWA is chosen to supply the worker. This is the main difference when adding competition as the TWA is no longer certain it will be awarded the contract when choosing to supply the matched worker to the CF.

To define when additional search is expected to be profitable for the TWA, we again assume that the expected lease duration difference for any two adjacent sufficient workers is at least one time period and \( n (x_k) \equiv 0 \), while at the same time we allow the interest rate to approach zero and apply the limit argument (cf. Subsection 2.2.1). We denote the limit expected revenue from any given match \( x_m \) as

\[
E \left[ R (x_m) \right] = g (q_m) \cdot \hat{T} (x_m) \cdot (\sigma - \delta) w^* \quad \text{and define} \quad \chi_\psi = \{ x \in \chi : E [ R (x) ] \geq E [ R (x_m) ] \}
\]

as the set of workers yielding a higher expected revenue than \( x_m \). We also define the number of elements in \( \chi_\psi \) as \( \mu_\psi \), and \( p_\psi \) as the probability of finding \( x_\psi \in \chi_\psi \). The linearised expression for the average expected revenue of the workers in \( \chi_\psi \), given \( x_m \) and \( \mu_\psi > 0 \), thereby becomes:

\[
E \left[ R (x_\psi \in \chi_\psi) \right] = \lim_{r \to 0^+} \frac{\sum_{r} E [ R (x_\psi \in \chi_\psi) ]}{\mu_\psi} = \frac{\sum_{x_\psi \in \chi_\psi} g (q_\psi) \hat{T} (x_\psi)}{\mu_\psi} (\sigma - \delta) w^*
\]

(11)

\[\text{If } \mu (\chi_x) = 0, \text{then } x_m \text{ is the most profitable match and the TWA has no incentives to search for another worker.}\]
where \( g(q_y) \) is defined analogously to \( g(q_m) \) above. The expected revenue of searching for a worker with a higher expected revenue (while still retaining the possibility to provide \( x_m \)) becomes:

\[
E \left[ R \left( x_y \mid x_m \right) \right] = p_y E \left[ R \left( x_y \in \chi_y \right) \right] + (1 - p_y) E \left[ R \left( x_m \right) \right]
\]

and taking into account that no other TWA supplies a worker in the current period, i.e. \( g(q^*) = \left( \frac{x^* - x - 1}{x - x^*} \right)^{z-1} \), the expected change in revenue from additional search can be expressed as:

\[
E (\Delta R) = g(q^*) \left\{ E \left[ R \left( x_y \mid x_m \right) \right] - E \left[ R \left( x_m \right) \right] \right\} =
\]

\[
= g(q^*) p_y \left\{ \frac{\sum x \in \chi_y g(q_x) \hat{T}(x)}{\mu_y} - g(q_m) \hat{T}(x) \right\} \cdot (\sigma - \delta) w^* \quad (13)
\]

The linearised additional search cost follows Equation (9), which enables us to define a condition for additional search.

**Proposition 4.** After introducing competition among \( z \) TWAs within the given framework, additional search will still take place if

\[
g(q^*) p_y \left\{ \frac{\sum x \in \chi_y g(q_x) \hat{T}(x)}{\mu_y} - g(q_m) \hat{T}(x) \right\} \cdot (\sigma - \delta) w^* \geq c.
\]

The resulting expression is quite intuitive; given an initial match, the TWA will conduct additional search if it expects additional revenue to surpass its costs. As previously mentioned, competition introduces an additional trade-off regarding the type of worker that will be matched, compared to the preceding subsection with only one TWA where the only trade-off was between expected contract duration and additional search cost. As increasing the number of competing TWAs decreases all individual TWA’s probability of winning, the contract becomes relatively more important to the TWA than the expected duration of the lease. Therefore the worker with the highest expected revenue for the TWA moves to the right in the distribution – i.e. towards a more productive worker. However, increasing the number of competing TWAs also lowers the expected revenue as the probability of being the TWA with the most productive match decreases accordingly.

To facilitate the analysis we plot the expected revenue curves for \( z = \{2, 3, 5, 10\} \) in Figure 1, while assuming that the number of sufficient workers is no larger than the number of non-sufficient workers. Given these parameters, the worker yielding the highest expected revenue for the TWA in the \( z = 2 \) case is still the \( x^* \) worker (similar to the result in Subsection 2.2.1).

Figure 1 shows that even though the CF would benefit from having several TWAs competing for the contract\(^{19}\), the decreasing expected revenue for each TWA will

\(^{19}\)Indeed, if \( z \) would be very large, then \( \bar{x} \) would most likely be supplied within one search period.
most likely restrict the number of TWAs that are willing to compete. Determining the optimal number of competing TWAs from the CF’s point of view will be left for future research, as the model should then include a hiring cost per TWA and also an opportunity cost for the CF for recruiting in-house and assuming all liabilities, which is outside the scope of this paper.

Figure 1: Expected revenue for varying levels of competition

3 Summary

The model in this paper provides a number of important insights regarding the search behaviour of an external TWA hired by a CF to match a worker with a vacant position when the CF is unable to monitor anything else other than the productivity of the supplied worker. Subsection 2.1 utilises the TWA as a filter to find an appropriate candidate, whereas in Subsection 2.2 the TWA (initially) employs the worker while the CF leases the worker. The established search behaviour of the TWA for each contract type is shown to generally differ from trying to find the marginally most productive worker given the available resources. The implications of the differing incentives establish that a payment-on-delivery contract is a necessary (but not sufficient) prerequisite for the TWA to carry out any actual search, as the TWA could otherwise merely claim to be searching. Subsection 2.1.3 then shows that any long-term worker supply planning on behalf of the CF (modelled as a fixed delivery date prior to which the CF does not need the matched worker; i.e. when planning vacations, etc.) could very well be negated by the TWA, as any search will occur as close to the last time period as possible. The main result from Subsection 2.1 is nevertheless that the TWA will never have any incentives to provide another worker other than the first sufficient candidate. As this corresponds to a random match from the subset of sufficient workers, it may result in a match with a lower average productivity compared to an in-
stance when any remaining resources would have been spent searching for the marginally most productive worker.

Arguably, this also implies that a vacancy (on average) is likely to be filled more rapidly by a TWA even if it would screen prospective candidates somewhat slower than the CF. The relatively quick vacancy/worker matching by TWAs (Autor, 2001, 2003; Houseman, 2001; Mitlacher, 2007), suggested to be the consequence of better matching technologies (Baumann et al., 2011; Neugart, 2005), could thereby be explained by the differing incentives shown in this paper – either in conjunction with actual differences in the available search technologies, or by the incentive structure itself.

Subsection 2.2 outlines a Recruit-and-Rent contract setup where the TWA can be shown to under certain circumstances (Proposition 3), gain from spending additional resources to find the marginally least productive (but still sufficient) worker after a first (random) match has been found. This will result in an even lower average productivity among the supplied matches than in the recruitment case. The theoretical predictions of our model thereby match the empirical results from a panel data study by Beckmann and Kuhn (2012) which found that firms which only use the TWA for screening purposes are found to be more productive than firms that continuously lease their temporary workforce. However, as the TWA now has incentives to conduct additional search, it is less certain that they, on average, will be able to recruit faster than the CF, as argued in the preceding subsection.

The introduced competition in Subsection 2.2.2 reduces the TWA’s incentives to perform additional search by adding an opportunity cost in that it also provides the other TWAs with another possibility to find an even better match. Here, the results are less clear and ultimately depend on the parameters of the model.

At a low level of competition the TWA would like to approach $x^*$, but as the competition increases there is a possibility that the TWA may decide to supply a worker with relatively high productivity in order to win the contract. However, there is also a trade-off in that the number of TWAs that the CF will be able to engage depends on the revenue that each TWA expects to make from being awarded the contract – a variable that decreases with the number of TWAs.

4 Discussion

The principal-agent model with hidden action presented here is a relevant framework for studying the recruitment of labour, since one of the main reasons to use a TWA is disengaged from the recruitment process (and thereby reduce the foregone productivity associated with having to filter all prospective candidates; cf. Beckmann & Kuhn, 2012; Connell & Burgess, 2002). Leasing workers from a TWA is also similar to when the employer signs redundancy insurance, which directly relates back to the first principal-agent model by Spence and Zeckhauser (1971). We further claim that the offered wage is more often associated with the specific position rather than the (maximum) productivity of the worker.
The assumed constant probability of finding a match is arguably less intuitive than allowing the set of remaining applicants to shrink after each screening. However, a constant probability of finding a match is a more restrictive assumption that will not only keep the mathematics more comprehensible, but will also provide more conservative results. We also argue that it mimics, to some extent, a dynamic distribution of applicants where individuals could both enter and leave during the search duration.

Any limited liability for the TWA towards the CF (as the model assumes that there is no penalty for not supplying a worker) could potentially facilitate moral hazard problems by inducing TWAs to accept assignments that they do not expect to complete. The TWA will also always have plausible deniability since the time until a match is found (or not) is stochastic. Determining any suboptimal behaviour would thereby require the CFs to either share information among each other, or to utilise the same TWA for a long succession of similar assignments – which are both highly implausible. Furthermore, the CF will have a hard time proving any suboptimal behaviour of the TWA whenever a worker is supplied since a match is always sufficient by definition.

Every TWA will have incentives to claim that they are able to provide the best match, which will make the screening of available TWAs a delicate task for the CF. Ironically, to some extent, this also corresponds to the underlying problem of choosing the right applicant to fill a vacancy – particularly in the presence of complex pricing, the minimum involvement by the CF, and that a number of verdicts from the Swedish Market Court (Marknadsdomstolen) suggest that no TWA has been able to objectively verify that they have any comparative advantage over their competitors that allows them to supply a better match.

Any advantages for the TWA in screening prospective candidates in the recruitment case would also increase the misaligned-incentive induced expected productivity difference even further. This is because any sufficient relative difference (that could be quite small in absolute terms) would allow the TWA to supply a match faster than the CF, while still rejecting any candidates with a higher productivity than the current least productive match.

With an exogenously determined minimum productivity level (and in the absence of several competing TWAs) the model suggests that while the most productive individuals have the same (random) chance as any other sufficient worker of being matched in the recruitment case, they would suffer the greatest penalty in their probability of being chosen for leased contingent work.

However, the CF should arguably be able to mitigate the negative effect on the expected productivity of the supplied worker by endogenising the requested minimum productivity level; \( x^{**} : x^{**} > x^{*} \). Using formal education as a proxy for the productivity parameter and comparing the situation in Sweden in 1999 (when private employment mediation agencies had only been available for 6-7 years; Joona & Wadensjö, 2008) to 2007 (Westéus et al., 2016), there is evidence

---

that the education level in the temporary work sector has increased considerably, and even surpassed that in the regular sector (i.e. any employment that is not through a TWA. See Andersson-Joona & Wadensjö, 2012; Petersson, 2013; Walter, 2012). This could suggest that the CFs have realised the need to overstate the minimum requirements for a vacant position to be filled with a TWA worker. In this case, the results from our model offer a demand side explanation for the increased education level in the TWA sector that is based on the TWA’s profit maximisation, without the need to introduce supply-side effects such as reputation (see e.g. Walter, 2012).

A worker that is matched to a position claimed to require $x^{**}$, but that objectively only requires $x^*$, would arguably also feel overqualified and/or mismatched (cf. Loughlin & Barling, 2001; Petersson, 2013) to a larger extent – especially in combination with the sectors lower wages (Andersson-Joona & Wadensjö, 2012), and adverse working conditions (Håkansson, Isidorsson, & Strauss-Raats, 2013). Evidence of this is found by de Graaf-Zijl (2012) in that agency workers with the highest educational attainment show the largest negative difference in job satisfaction (which relates mostly to the content of their job and to job insecurity only to a lesser extent) compared to what the author denotes regular workers.

This paper contributes to the theoretical literature and leaves it open to future research to measure its importance. The next step could be to empirically analyse the assumptions of limited liabilities when a worker is not supplied, and thereafter aim to de-construct the pricing mechanism of the services of TWAs under the misaligned incentives framework that has been outlined above. Future research could also focus on evaluating the relative productivity of agency workers compared to temporary workers employed directly at the CF to test this model’s theoretical predictions. Another possible direction would be to investigate potential gains from eliminating any incentive misalignment, as they could arguably be significant.

References


Department of Economics, University of Birmingham.