DISCUSSION PAPER SERIES

IZA DP No. 464

The Hold-Down Problem and the Boundaries of the Firm: Lessons from a Hidden Action Model with Endogenous Outside Option

Wendelin Schnedler
Uwe Sunde

March 2002

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Hold-Down Problem and the Boundaries of the Firm: Lessons from a Hidden Action Model with Endogenous Outside Option

Wendelin Schnedler
IZA, Bonn and Bonn Graduate School of Economics

Uwe Sunde
IZA, Bonn and Bonn Graduate School of Economics

Discussion Paper No. 464
March 2002

IZA

P.O. Box 7240
D-53072 Bonn
Germany

Tel.: +49-228-3894-0
Fax: +49-228-3894-210
Email: iza@iza.org

IZA Discussion Paper No. 464
March 2002

ABSTRACT

The Hold-Down Problem and the Boundaries of the Firm: Lessons from a Hidden Action Model with Endogenous Outside Option*

This paper offers a rationale for limiting the delegation of (real) authority, which neither relies on insurance arguments nor depends on ownership structure. We analyse a repeated hidden action model in which the actions of a risk neutral agent determine his future outside option. Consequently, the agent can improve his future bargaining position, which gives the principal an incentive to retain sufficient control over the agent's actions. Using respective one-period contracts, the principal can implement the efficient outcome while "selling the shop" to the agent is sub-optimal. This provides an argument for integration if the boundary of the firm is defined by control rights rather than the entitlement to revenues.

JEL Classification: D23, D82, L23, L33

Keywords: hidden action, moral hazard, endogenous outside option, authority, outsourcing

Wendelin Schnedler
IZA
P.O. Box 7240
D-53072 Bonn
Germany
Tel.: +49 228 38 94 121
Fax: +49 228 38 94 510
Email: schnedler@iza.org

* The authors wish to thank Georg Nöldeke, Jörg Oechssler, Nicola Pavoni, Urs Schweizer and seminar participants at the University of Bonn for helpful comments. All remaining errors are our own. Financial support by DFG is gratefully acknowledged. First version: April 10, 2001; this version: March 15, 2002.

1 Introduction

In recent years there has been an increased tendency of firms to flatten hierarchies and to delegate responsibilities. In particular, many employer-employee relationships were re-designed to transfer some of the entrepreneurial responsibilities to the employees or agents and to set better incentives by making them freelancers or franchisees. Besides the potential advantage of reducing costs through improved incentives, the additional freedom may be used by the agent to improve his bargaining position –possibly by choosing counterproductive actions. For example, an employee (agent) who is engaged in customer contact might acquire a close relationship with customers by being extraordinarily obliging. Not only will this increase his performance as measured by his employer (principal), but the good relations with the customers might allow the agent to continue cooperating with the customers without the principal. The agent could use this as a threat when negotiating next time with the principal about his salary. This means that the principal faces costs when leaving the decision to the agent. The key insight of this paper is that, in order to avoid these costs, the principal has to exercise some degree of authority, which she cannot if the respective activities are outsourced.

The model presented in this paper complements and extends the existing literature in several ways. As in most papers on incomplete contracts, the possibility of renegotiation as well as the ownership structure with respect to who has residual control over payoffs play a role. Similar to a strand of this literature, ownership is an element of the contract (see e.g. Nöldeke and Schmidt, 1998). Eventually, the boundaries of the firm are determined endogenously in the model. Here, it is not the ownership structure, but the power to control the agent's actions which defines the boundary. In the usual hold-up context, the investment is determined by the fact that renegotiation results do not allow to internalise all the benefits of the investment, leading to underinvestment. In the model presented below, the agent's investment improves his position in the renegotiation process, leading to overinvestment; the gains from the agent's investment are higher outside than inside the contractual relationship. In the literature on incomplete contracts, states of nature play an important role as they determine the ex-post efficient outcome and cannot be contracted upon. Like in the standard textbook hidden action problem (see e.g. Salanié, 1998), states of nature here are signals about agent's actions and are verifiable. However, we assume risk neutrality of principal and agent, so that risk sharing issues do not prevent delegation to the agent ("selling the shop").

The concept of "integration" used in this paper is linked to the idea of authority rather than ownership. Aghion and Tirole (1997) analyse the role of formal and real authority and its dependence on informational asymmetry. In contrast, our focus is not on informational asymmetry. Rather, asymmetric intertemporal interests and the possibility to renegotiate render it optimal to control actions, i.e. exercise authority. Like in the present article, Baker, Gibbons, and Murphy (2002) also consider the question of integration in a repeated interaction but they understand integration in terms of ownership and focus on the effect of market conditions on the integration decision. Market conditions also play a role in the paper by Grossman and Helpman (2002), who develop an equilibrium model of integration versus outsourcing. In this model outsourcing reduces production costs but increases the search costs for finding appropriate partners. Complementarily, our (partial) analysis looks at the adverse incentive effects of leaving discretion to the agent.

Our model also touches upon the literature on short-term vs. long-term contracts. Fudenberg, Holmström, and Milgrom (1990) devise four sufficient conditions under which, in a repeated game setting with recontracting, common knowledge and identical discount rates for both players, short term contracts can replace long term contracts. In our context, one of their conditions is violated: not all public information is contractible – only the signal about agent's actions. Nevertheless, it will be shown that complex multi-period contracts are not necessary and that the principal can restrict herself to short term contracts and still obtain the maximal payoff.

Repeated hidden action games with intertemporal effects have also been studied by Abreu, Pearce, and Stacchetti (1990), Ma (1991), Holmström (1999) and Fernandes and Phelan (2000). In the model presented below, actions have intertemporal effects regardless of whether there is a relationship or not. In contrast, Ma (1991) only considers effects inside the relationship. Unlike Abreu, Pearce, and Stacchetti (1990) and Holmström (1999), we consider direct intertemporal effects of current actions on future payoffs, not the informational consequences. While informational asymmetries are essential for Fernandes and Phelan (2000), the results in our model are driven by asymetric intertemporal incentives. While Posner and Triantis (2001) ask the question when to prevent the agents from using an investment outside the relationship by a "covenant not to compete" clause, we assume that the agent cannot commit himself to forfeit the alternative usage.

Marcet and Marimon (1998) devise general solution methods for dynamic incentive (or implementability) problems in infinite horizon settings. As our analysis deals with a finite horizon, it is simpler to solve the problem directly rather than adapting their method.

The following section introduces the model. Section 3 explains the mechanisms at work in a two-period model. In Section 4 , the results are extended to finite games and multi-period contracting. Finally, Section 5 concludes.

2 The model

There are two players: a risk-neutral principal (P) and a risk-neutral agent (A) who play a repeated hidden action game. Later, we will analyse a two-period as well as a multi-period version of the model. In each period t, the interaction between principal and agent creates a revenue $\pi(\cdot,\cdot,\cdot)$. Future payoffs are discounted with a discount factor $\delta\in]0;1]$ which is the same for principal and agent. By default the agent is the receiver of the revenue ($r_t = A$); but contractual arrangements can assign this role to the principal ($r_t = P$).

Both, principal and agent, have control over some determinants of the surplus $\Phi(s_t, a_t, a_{t-1})$ which is defined as the revenue minus the costs:

$$\Phi(s_t, a_t, a_{t-1}) = \pi(s_t, a_t, a_{t-1}) - c(s_t, a_t, a_{t-1}).$$

The principal can costlessly set a support level $s_t \in \{0, 1\}$.[1] The agent can influence the surplus by his present and past activity level a_t and a_{t-1}, where a_t and a_{t-1} can take a low (L) or a high (H) value. The costs $c(\cdot,\cdot,\cdot)$ for the activity a_t, which may depend on support s_t and past activity a_{t-1}, are always born by the agent. The revenue $\pi(\cdot,\cdot,\cdot)$ is non-decreasing in the

[1]Acquiring the ability to set a support in itself may be costly but we assume this costs to be sunk at the moment of decision. If support were costly, we would have to deal with a two-sided moral hazard problem in which it is not surprising that selling the shop is not optimal.

present activity level a_t. For most of the analysis it will be sufficient to make assumptions on the surplus function.

To illustrate the effect of the principal's choice variable, suppose that she owns an asset which helps the agent to produce. We assume that she can either support the agent by allowing him to use the asset ($s_t = 1$) or not support him by preventing the use of the asset ($s_t = 0$) in t. The following assumption determines the effect of support on the surplus:

Assumption 1. *Independently from the agent's activity, support by the principal increases the surplus:*

$$\forall\, a,\, a' :\ \Phi(0, a, a') < \Phi(1, a, a'). \tag{1}$$

For the current period, the costs of engaging in the activity $c(\cdot,\cdot,\cdot)$ outweigh the benefits, so the surplus decreases in current period's activity:

Assumption 2. *Low activity maximises the current surplus:*

$$\forall\, s,\, a :\ \Phi(s, H, a) < \Phi(s, L, a). \tag{2}$$

Note that this and the non-decreasingness of the revenue $\pi(\cdot,\cdot,\cdot)$ in the current activity level imply increasing costs $c(\cdot,\cdot,\cdot)$ in the current activity level.

Next period's surplus is increased by the activity a so that a can be interpreted as an investment. If the principal sticks to her support choice s in both periods, the total intertemporal surplus generated by the low activity exceeds the total intertemporal surplus generated by the high activity. In other words:

Assumption 3. *Given identical support in two consecutive periods, the discounted future gains from the investment cannot recoup the present loss implied by the investment:*

$$\forall\, a\,, a'\,, s\ :\ \delta(\Phi(s, a', H) - \Phi(s, a', L)) < \Phi(s, L, a) - \Phi(s, H, a). \tag{3}$$

However, the discounted future gains of the investment outweigh the costs if the agent is currently supported and will not receive support in the future. This is expressed in the following assumption:

Assumption 4. *Given support in the current period and no support in the following period, the higher costs accrued through exerting the high activity rather than the low activity are more than outweighed by a higher future payoff:*

$$\forall a\,, a' :\; c(1, H, a) - c(1, L, a) < \delta(\Phi(0, a', H) - \Phi(0, a', L)). \tag{4}$$

This completes the set of assumptions on revenue and cost functions.[2]

To illustrate the economic situation reflected by these assumptions, consider a two-period example of an experienced advocate and her junior partner. The advocate has established a large client base. She may ($s = 1$) or may not ($s = 0$) support the junior partner (who only has a few faithful clients) by directing her client's attention to the junior partner. Suppose that the number of clients determines revenues. The junior partner can handle clients with an extra portion of care ($a = H$) at high costs or just act normally ($a = L$).

Whatever the junior partner does, the surplus will be higher if the senior advocate supports him, e.g. by calling up her clients. This is reflected by Assumption 1. But regardless of support and previous behaviour towards the clients, current surplus is highest when the junior partner is not too obliging as indicated by Assumption 2. This could be because the extra portion of care, like inviting clients for dinner is very costly.

If the junior advocate can count on the senior advocate's help in both periods, there is no point for the junior advocate in being particularly nice to the clients which is expressed by Assumption 3 for $s = 1$. Given that the junior advocate has no help by the senior advocate and will get no help in the future, there is no opportunity for the junior advocate to alienate the former's customers and hence it does not help to put in extra effort to improve the familiarity with the clients ($s = 0$ in Assumption 3).

If, however, the senior advocate only provides support in the first period, being particularly nice to customers in this period may turn some of them

[2]For an example which satisfies Assumptions 1 to 4 consider the revenue function $\pi(s, a, a') \equiv 2s + (1 - s)a'$ and the cost function $C(s, a, a') \equiv (1 - \frac{s}{2})a$, where $a \in \{0, 1\}$ and $s \in \{0, 1\}$.

into faithful customers who stick with the junior partner even when he leaves the senior advocate. This effect is described by (4).

Given the environment described so far, one can now show that the efficient outcome is obtained when the principal provides support in all periods, and the agent exerts low activity:

Lemma 1. *The efficient allocation is achieved by choosing support ($s = 1$) and low activity ($a = L$) in all periods.*

Proof. To see this result, consider the overall surplus over a horizon of $\mathcal{T}$ periods, $\sum_{t=1}^{\mathcal{T}} \delta^{t-1}\Phi(s_t, a_t, a_{t-1})$. By inequality (1), this surplus is higher when support is provided ($s = 1$), independently of the activity chosen. Then, the surplus is maximised by the low activity ($a = L$) in the last period ($t = \mathcal{T}$) due to (2) and in earlier periods ($t < \mathcal{T}$) due to (3). □

The framework presented allows to deduce the following regularities:

Lemma 2.

$$\begin{aligned}
\forall\, a\,, a' : &\quad \Phi(1, L, a) - \Phi(1, H, a) &< \delta(\Phi(0, a', H) - \Phi(0, a', L)), \quad (5)\\
\forall\, a\,, a'\,, s : &\quad \delta\,(\Phi(1, a', H) - \Phi(1, a', L)) &< \Phi(s, L, a) - \Phi(s, H, a), \quad (6)\\
\forall\, a' : &\quad \Phi(1, a', H) - \Phi(1, a', L) &< \Phi(0, a', H) - \Phi(0, a', L). \quad (7)
\end{aligned}$$

Proof. Inequality (5) follows from Assumption 4 together with the non-decreasingness of the revenue in the second argument. By using (5) together with (3), one obtains:

$$\begin{aligned}
\forall\, a\,, a' \,:\; \delta(\Phi(1, a', H) - \Phi(1, a', L)) &\overset{\text{by (3) for } s=1}{<} \Phi(1, L, a) - \Phi(1, H, a) \overset{\text{by (5)}}{<} \\
\delta(\Phi(0, a', H) - \Phi(0, a', L)) &\overset{\text{by (3) for } s=0}{<} \Phi(0, L, a) - \Phi(0, H, a) \quad (8)
\end{aligned}$$

The second result, (6), is certainly true for $s = 1$ by (3). By comparing the smallest with the largest value in (8) it is also true for $s = 0$. Finally, (7) results from comparing the term on the left in the first line with the term of the left in the last line of (8). □

What is the intuition behind Lemma 2? Inequality (5) is a reformulation of inequality (4) in surplus terms. It supplies a motive to exert high activity as a precautionary measure against the loss of support. Inequality

(6) asserts that such a precautionary measure is redundant when support can be expected to prevail in the following period. Thus, the effect of the precautionary investment must be lower when support is given as stated by inequality (7).

The support of the principal s, the activity of the agent a, and the surplus $\Phi(\cdot,\cdot,\cdot)$ are not verifiable. The reason might be that support is only observed by the principal, the activity by the agent, and the surplus by whoever receives it. However, there is dichotomous signal Z_t which can be contracted upon and which takes on the value one with probability $P(a_t)$ and the value zero with probability $1 - P(a_t)$ where $P(H) > P(L)$.

Each period t of the model consists of the following sequence of events:

1. The principal proposes a contract.
2. The agent accepts or rejects.
3. The principal chooses support s_t and the agent decides on activity a_t.[3]
4. Nature draws the signal Z_t given the activity choice a_t.
5. Transfers are carried out according to the contract.

This time line is depicted in Figure 1.

The possibility to reject the contract, i.e. produce independently, creates an outside option for the agent which leaves the principal with nothing.

For simplicity, it should be assumed that the principal knows the past activity of the agent at the beginning of a new period. This is for example the case, when the agent has accepted the contract and it induces a unique activity. Then, there is no asymmetric information at the time when the principal proposes the contract and, unlike Fernandes and Phelan (2000), it is only necessary to consider participation and incentive constraints. If the last activity were not observable, the essential link between agent's activity and future payoffs is retained given sufficient discriminatory power of the

[3]For our results it is irrelevant whether principal and agent move simultaneously or in a particular sequence.

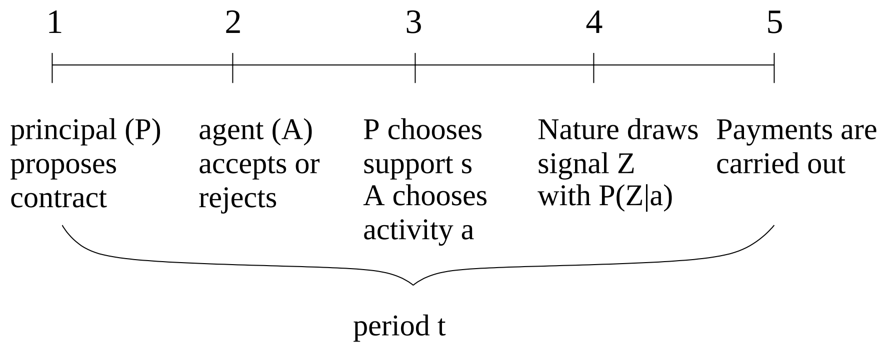

Figure 1: Seqence of events within a period

signal (see Appendix A).

The purpose of formal contracts in this model is to allow for externally enforcable commitments on the basis of signals occuring while the contract is valid ("conditioning on signals"). A contract lasting from t_0 to T specifies for each period t the payments and the identity of the revenue recipient depending on the respective history of signal realisations $\xi_t = \{z_i\}_{0 \leq i \leq t}$. Formally, each contract $\mathcal{C}$ is a set of subcontracts $\mathcal{C}_t$: $\mathcal{C} = \{\mathcal{C}_t\}_{t_0 \leq t \leq T}$. Each subcontract is a function:

$$\begin{aligned} \mathcal{C}_t : \{0,1\}^t &\rightarrow I\!R \times \{A, P\} \\ \xi_t &\mapsto (\tau_t(\xi_t), r_t(\xi_t)), \end{aligned} \tag{9}$$

where $\tau_t(\xi_t)$ is the dependent payment in period t from the principal to the agent and $r_t(\xi_t)$ constitutes the identity of the revenue recipient. In what follows, the identity of the recipient does not depend on signals. A practical argument for this would be that in reality the employment status is specified before work begins. Theoretically, any incentive effect of an alternating recipient can be reflected by an appropriate change in transfers (see Appendix B). Moreover, as we will see, the first-best can be achieved without such conditioning.

For later reference, we call a contract which lasts longer than one period ($T - t_0 + 1 > 1$) *multi-period* contract and *one-period* contract, otherwise. Until Section 4, attention will be limited to one-period contracts only.

If a contract is proposed and accepted while a multi-period contract still lasts, the former will be assumed to be the only valid contract: it can mimic any element of the multi-period contract. If both parties agree to a contract which "alters" the multi-period contract, they will also agree on replacing the contract by a new contract which combines the old contract and the alterations.

A contract which uses the current signal enables the principal to control stochastically the agent's current behaviour and is therefore called a *control* contract. We refrain from analysing contracts which employ past signals to determine current transfers. It can be shown that such constracts can be substituted by simple control contracts which condition only on the current signal (see Appendix B). Furthermore, it will turn out that simple control contracts already achieve the first best, so that there is no need for intertemporal conditioning.

For the sake of brevity, we introduce a terminology describing the production arrangements between principal and agent. If no contractual agreement is in place, the agent is the receiver of the revenue and we call him *self-employed*. Later, this situation will define the agent's outside option. If a contractual arrangement is in place, we distinguish between four different types of contracts depending on whether the principal or the agent is recipient of the revenue and whether the contract is conditioning on the current signal (see Table 1).

		Recipient of revenue	
		principal (r=P)	agent (r=A)
transfers dependent on current signal (*simple contracts*)	yes	performance pay	franchise
	no	fixed salary	freelance

Table 1: Typology of Contracts

The surplus function depends on the previous activity. For a consistent notation for the first period, we suppose an arbitrary preceeding activity and denote it by a_0.

3 Two-period model

In this section, we consider a two-period version of the model in order to lay out its basic mechanics. Proofs in this section are kept verbal. Formal proofs can be found in the following section on finite games which include the two-period game as a special case. The two-period game will be solved by backward induction. First, the second period will be analysed. It will turn out, that continuation payoffs only depend on first period activity of the agent and do so in a particular form which leads to the formulation of *activity-markovian* payoffs. Then, we address the first period which exhibits no end-game effect and therefore lies in the focus of our interest.

3.1 Analysis of the second period

The last decisions to be made are the decisions on activity by the agent and support by the principal First, we deal with the activity decision of the agent. Suppose the agent receives the revenue. This is the case if he has rejected the contract and is self employed, or if he has accepted the contract and the contract stipulates that he is the receiver. In either case, the agent has no benefit from working particularly hard, and chooses the low activity level L (see inequality (2)). Under a fixed salary or a freelance scheme the agent will prefer L since it is less costly for him. The same can be achieved when the contract has been accepted, the principal is the recipient, and she uses a control contract which puts sufficient weight on the signal associated with the low activity level. Note, that the optimal activity of the agent is independent of the support decision which is discussed next.

The support can depend on the agent's participation decision. For the moment, there is no hint as to which support the principal will provide in either case. However, it will become clear below that in equlibrium the principal will only support the agent if a contract was signed. If the agent rejects a contract, the principal will minimise her help and provide no support, in order to give the agent a credible incentive for joining a contractual relationship. That is the case even though support is costless for the principal, because she will have no payoff without the agent working on her behalf.

Previous to the activity-support game comes the participation decision of the agent. The agent will accept any contract which yields him a payoff that is as least as high as what he gets outside the relationship. This outside option is determined by his activity in the preceeding period, the support the principal grants after rejection, and the low activity in the current period (which we just found to be optimal for him in case he acts as self-employed).

At the beginning of the second period, when proposing a contract, the principal has to take the participation constraint into account. Thus, the principal's payoff when the agent accepts the contract is the difference between surplus generated by the contract and the payment to the agent. Clearly, this payoff can be maximised only if the payment to the agent is as small as possible: just enough to make the agent participate. Indeed, the principal can reduce the agent's payoff until it equals the latter's outside option. This is possible because the agent is risk-neutral, and thus the principal can reduce the transfers under all signal realisations by the same amount without affecting any incentives stemming from the respective contract in place.

As mentioned before, the principal can only maximise her payoff by maximising the gross revenue of the joint relation and minimising the agent's outside option. Since the surplus increases in support (inequality (1)) it is clear now, that this can only be accomplished by providing support when a contract is agreed upon, and by providing no support otherwise. Furthermore, the principal prefers contracts which implement that the agent acts normally, without exaggerating his care for the clients. The reason being that this activity creates the highest surplus (inequality (2)). As mentioned at the beginning of this subsection, the low activity can be implemented either by "selling the shop" to the agent and making him revenue recipient, by giving him a fixed salary, or by an appropriate control contract which puts sufficient weight on the low activity signal.

To sum up, in the second and last period of the game, contracts of all four types can be used by the principal to maximise her payoff: fixed pay, freelance, performance pay, and franchise. They implement a low activity level (L), leave the agent with $\Phi(0, L, a_1)$ and the principal with $\Phi(1, L, a_1) - \Phi(0, L, a_1)$.

3.2 Activity-markovian continuation payoffs

Note that the continuation payoff replacing the second period only depends on the activity of the agent in the first period. Continuation payoffs at the moment of principal's proposal which only depend on the immediately preceeding activity of the agent play an important role in the following analysis. Therefore, we define:

Definition 1 (Activity Markovian Continuation Payoffs). *Payoffs are called* activity-markovian *continuation payoffs, if the difference between continuation payoffs, given the previous activity was high (H) instead of low (L), amounts to*

$$\begin{aligned} \Phi(0,L,H) - \Phi(0,L,L) \qquad & \textit{for the agent and} \\ \Phi(1,L,H) - \Phi(0,L,H) - [\Phi(1,L,L) - \Phi(0,L,L)] \qquad & \textit{for the principal.} \end{aligned}$$

The concept of activity-markovian continuation payoffs is only relevant at the beginning of a period when the principal proposes a contract. So, conditioning on actions within a period is not precluded. After having introduced the terminology of activity-markovian continuation payoffs, we now replace the second period by the respective activity-markovian continuation payoffs and analyse the first period of the game.

3.3 Analysis of the first period

Let us consider first what happens after a rejection of the contract in the first period. Later, we will deal with what follows in case of acceptance of the contract.

The self-employed agent has to decide on his activity level. This decision depends on the current and future payoff difference between the two alternatives he has, the high and the low activity level. If the principal grants the self-employed agent support, it follows from the fact that working hard secures the agent the faith of the clients in the second period when the agent is self-employed (compare inequality (5)), that the latter will pick $a_1 = H$. Conversely, without support, the agent has no chance to alienate customers and does not bother about exerting the high activity (c.f. inequality (3) with $s = 0$). Thus the agent's activity depends on whether he is given the chance to alienate customers or not, i.e. whether the principal supports him or not.

After a rejection, the principal's support choice only influences her payoff via the activity of the agent which in turn determines the second period continuation payoff. The agent's outside option in the next period increases more in the preceeding activity than the surplus generated in the next period within the relationship (by inequality (7)). Thus, the principal wants to induce the agent to exert low activity. She can do so by credibly committing herself not to support the agent.

After having dealt with the consequences of contract rejection, we turn to contracts and the activity implemented by them. The payoff to the principal is the surplus of the current period plus the benefit from the second period minus the compensation for the forgone earnings the agent might have earned as a self-employed. Since the latter is fixed in the actions taken after contract acceptance, the principal maximises: $\Phi(s_1^{\text{in}}, a_1, a_0) + \delta(\Phi(1, L, a_1) - \Phi(0, L, a_1))$, where s_1^{in} is her support decision after contract acceptance. Since the principal has no gain from attracting new customers, the advantage of being particularly nice to customers is much lower for her which is described by inequality (7). She wants to implement L and support the agent to get the maximal possible surplus (see (2) and (1)).

A control contract can induce the agent to choose L if the transfer to the agent for the signal realisation associated with L is sufficiently large. This can be achieved by increasing the payment under this realisation and lowering the payment under the other realisation. Since the agent is risk neutral, these changes can be undertaken without affecting expected payments. Then, the agent still has the formal authority, but the real decision is made by the principal through a respective design of the contract.

In a fixed pay contract, the principal is the revenue recipient and payments are not conditioned on signals. Under these circumstances, the agent weighs the current costs of the high activity against the future gains. To generate the maximal possible surplus, the principal must grant the agent access to her customers (see (1)). Since future payoffs are activity markovian and thus reward the agent for having been particularly nice, the agent will exert high activity to improve his future outside option (see (4)) and thereby reduce the generated surplus (see (2)).

Under a freelance contract, the same problem occurs. Activity markovian continuation payoffs, i.e. the future gain from high activity, render treating the customers with too much care very attractive if the principal supports the agent by establishing contact with these customers. There is no element in a freelance contract which offsets this attractivity of the high activity reflected in (5). Incidentially, not supporting the agent is no cure to the problem. It is true, that the agent then refrains from the sub-optimal high activity since he has no opportunity to turn customers of the principal into his customers (see (3)). On the other hand, the customers cannot be served which also leads to a loss of surplus (see (1)).

Summing up, it is optimal for the principal to propose a control contract in the first period. Fix salary and freelance contracts are dismissed by the principal as they generate a lower payoff: either the agent exaggerates when choosing the activity level to improve the prospects with regard to the next contract renegotiation or the principal cannot support the agent. In the optimal control contract, the agent is "held down" by the principal to exert only low activity –despite the fact that the principal grants support which can be potentially exploited by the agent.

The surplus generated in the first period amounts to $\Phi(1, L, a_0)$ of which the agent gets $\Phi(0, L, a_0)$. This implies that the first-period continuation payoffs are activity markovian: they differ only in the previous activity a_0 such that the agent profits $\Phi(0, L, H) - \Phi(0, L, L)$ when $a_0 = H$ instead of $a_0 = L$ while the principal looses $\Phi(1, L, H) - \Phi(0, L, H) - [\Phi(1, L, L) - \Phi(0, L, L)]$.

To reach this conclusion, we only relied on the basic Assumptions 1 to 4 and on the fact that continuation payoffs replacing game of the second period were activity markovian. Thus, if the sub-game beginning in period $t + 1$ can be replaced by activity-markovian continuation payoffs, the sub-game beginning in the preceeding period t can also be replaced by such payoffs. Since the last period was characterised by activity markovian payoffs, it is thus possible to replace all sub-games by activity markovian payoffs. This induction argument will be formalised in the next section when considering a game with a finite number of periods.

4 Finite games

In this section, the two period model is extended to an arbitrary finite number of periods $\mathcal{T}$. First, the focus will be on one-period contracts, later it will be extended to multi-period contracts. The findings of the two-period example can be generalised to finite games by an induction argument. The starting point is the last period $\mathcal{T}$. From the previous section, we know that it can be replaced by activity-markovian continuation payoffs. It remains to show, that continuation payoffs are activity-markovian at t if they are activity-markovian at $t+1$. To do so, consider the game after rejection of the contract.

Proposition 1. *If continuation payoffs are activity-markovian at* $t+1$*, the principal chooses* $s_t = 0$ *and the agent* $a_t = L$ *after rejection of the contract.*

Proof. The relevant payoff to the principal after rejection is: $\delta(\Phi(1, L, a_t) - \Phi(0, L, a_t)) + c$, where c is a constant which is independent of any action taken in t. Since this payoff is independent of s_t the principal is indifferent between $s_t = 1$ and $s_t = 0$. The relevant payoff to the agent is:

$$\Phi(s_t, a_t, a_{t-1}) + \delta\Phi(s_{t+1}, a_{t+1}, a_t) + d, \tag{10}$$

where d is a constant which is independent of any action taken in t. From (5), it is clear that the best reply of the agent to $s_t = 1$ is $a_t = H$ wheras from (6) the best reply under $s_t = 0$ is $a_t = L$. So, there are two equilibria of the activity-support game: $(s_t = 1, a_t = H)$ and $(s_t = 0, a_t = L)$. By (7) the principal prefers the second equilibrium. The principal can implement this equilibrium by threatening to play $s_t = 0$ when negotiating the contract. This threat is credible, since the principal is indifferent between her two actions. □

This result is robust with respect to the sequence of decisions. In the simultanous game, the principal prefers the Nash equilibrium where she is not supporting the agent and the latter exerts low activity. If the principal has the first-mover advantage in a sequential game, she will choose this equilibrium straight away. But even if the agent can decide first, not being supported is a credible threat to which he has to react by picking the low activity level. Note, that if setting support were costly, e.g. because it is more difficult to support a self-employed agent than an agent with whom one has a contractual relationship, this would make the case for the outcome desired by the principal even stronger. Then, not supporting and choosing the low activity is the only Nash equilibrium.

Next, we turn to assessing the two player's decisions if a contract is signed. It turns out that the first best outcome can be reached by a contract, however only if the principal succeeds in retaining sufficient control over the agent's actions:

Proposition 2. *If continuation payoffs are activity-markovian at $t+1$, the efficient outcome can only be implemented at t by a control contract.*

Proof. The relevant payoff to the principal after contract acceptance is:

$$\pi(s_t, a_t, a_{t-1})1\!\!1_{\{P\}}(r_t) - \tau^H P(a_t) - \tau^L(1-P(a_t)) + \delta[\Phi(1, L, a_t) - \Phi(0, L, a_t)] + c,$$

where $1\!\!1_{\{P\}}(r_t)$ is an indicator function taking the value 1 if the principal is the revenue recipient and 0 otherwise. The continuation payoff c is independent from any action taken in t. If $r_t = A$ the principal is indifferent between $s_t = 0$ and $s_t = 1$, if $r_t = P$ the principal strictly prefers $s_t = 1$ by (1). Thus, implementing $s_t = 1$ poses no problem. The payoff to the agent after contract acceptance is:

$$\pi(s_t, a_t, a_{t-1})1\!\!1_{\{A\}}(r_t) - c(s_t, a_t, a_{t-1}) + \tau^H P(a_t) + \tau^L(1-P(a_t)) + \delta\Phi(0, L, a_t) + d, \tag{11}$$

where d is constant, i.e. it does not depend on any action taken in t. To implement the efficient outcome, the agent must prefer $a_t = L$ to $a_t = H$ while $s_t = 1$:

$$\begin{aligned} &\pi(1, H, a_{t-1})1\!\!1_{\{A\}}(r_t) - c(1, H, a_{t-1}) + \tau^H P(H) + \tau^L(1 - P(H)) + \delta\Phi(0, L, H) \\ <\ &\pi(1, L, a_{t-1})1\!\!1_{\{A\}}(r_t) - c(1, L, a_{t-1}) + \tau^H P(L) + \tau^L(1 - P(L)) + \delta\Phi(0, L, L)\,. \end{aligned}$$

This can be rewritten as:

$$\begin{aligned} &(\tau^H - \tau^L)\{P(L) - P(H)\} \\ > \ &\{\pi(1, H, a_{t-1}) - \pi(1, L, a_{t-1})\}\, 1\!\!1_{\{A\}}(r_t) - \{c(1, H, a_{t-1}) - c(1, L, a_{t-1})\} \\ + \ &\delta\{\Phi(0, L, H) - \Phi(0, L, L)\}\,. \end{aligned} \tag{12}$$

The right-hand side of this expression is larger than zero for $r_t = A$ due to (5) and for $r_t = P$ due to (4). Since $P(L) < P(H)$, it must follow that $\tau^H < \tau^L$. Thus, in the class of one-period contracts only control contracts can implement the efficient outcome. A respective control contract can be found by reducing τ^H by Δ and increasing τ^L by $\Delta\frac{P(L)}{1-P(L)}$ until (12) is valid. Such a change leaves the expected payoff to the agent unaltered and hence does not influence the participation decision. □

Observing (12), and noting that revenue $\pi(\cdot,\cdot,\cdot)$ is non-decreasing in current activity, it is easier for the principal to induce the agent to exert low activity by using a performance pay contract in comparison to a freelance contract. The reason is that the freelancer receives directly part of the revenues as $1\!\!1_{\{A\}}(A) = 1$ which provides him with even stronger incentives to exert H which have to be offset by transfers.

In order to maximise her surplus, the principal attempts to reduce her payments as much as possible. The following result shows that she can reduce these payments until the agent is just indifferent between signing the proposed contract or staying on his own:

Proposition 3. *If continuation payoffs are activity-markovian at $t+1$, the participation constraint of an optimal contract signed in t binds.*

Proof. From plugging-in the results from Proposition 1 and Proposition 2 into the payoff to the agent after contract rejection and acceptance, it is clear that the agent is willing to accept if and only if:

$$\pi(1, L, a_{t-1})1\!\!1_{\{A\}}(r_t) - c(1, L, a_{t-1}) + \tau^H P(L) + \tau^L(1 - P(L)) \geq \Phi(0, L, L). \tag{13}$$

When reducing τ^H and τ^L by a fixed amount, the incentive condition (12) remains unaltered. Hence, the principal increases her payoff by reducing transfers to the agent until inequality (13) binds. □

The preceeding findings essentially allow to deduce a characterisation of all subgame perfect equilibria of the contracting game between principal and agent:

Theorem 1. *In all subgame-perfect equilibria of the finite game, continuation payoffs are activity-markovian. Further, the agent exerts L, the principal supports after acceptance and does not support after rejection in any period t.*

Proof. For the last period $\mathcal{T}$, the claim was proven in Section 3. By Propositions 1 to 3, it follows that continuation payoffs are activity-markovian in t given they are activity-markovian in $t+1$. Further, the contract will be accepted and stipulate $s_t = 1$ and $a_t = L$ so that the principal gets $\Phi(1, L, a_{t-1}) - \Phi(0, L, a_{t-1})$ while the agent gets $\Phi(0, L, a_{t-1})$ in t. By induction, payoffs are activity-markovian in all periods. □

Taking the previous propositions together, the continuation payoffs at the proposal point of any period can be computed:

Corollary 1. *In all subgame-perfect equilibria of the finite game, continuation payoffs at the proposal stage of period t_0 are*

$$\sum_{t=t_0}^{\mathcal{T}} \delta^t \Phi(0, L, L) + \Phi(0, L, a_{t_0}) \textit{ for the agent and}$$

$$\sum_{t=t_0}^{\mathcal{T}} \delta^t \Phi(1, L, L) + \Phi(1, L, a_0) - \sum_{t=t_0}^{\mathcal{T}} \delta^t \Phi(0, L, L) - \Phi(0, L, a_{t_0}) \textit{ for the principal.}$$

Proof. The overall payoff to principal and agent can be found by adding the properly discounted periodwise payoffs from the previous theorem. □

In practice, this implies that the principal only needs to concentrate on one-period control contracts to maximise her payoff and to achieve the efficient outcome. In fact the efficient outcome can only be achieved by control contracts. But what about multi-period contracts? In particular: how do multi-period non-control contracts such as freelance and fixed salary contracts fare? The answer is given by the following proposition:

Proposition 4. *Freelance and fixed salary contracts ending in $T < \mathcal{T}$ do not implement the efficient outcome as a subgame-perfect equilibrium.*

Proof. Consider the last period of validity of the contract T. Continuation payoffs at T are activity-markovian by Theorem 1. To implement $s_T = 1$ and $a_T = L$ requires (12) to hold for $t = T$. This, however, implies that $\tau^H < \tau^L$ which is not true for fixed salary or freelance contracts. □

The consequences of this proposition are far reaching: freelance and fixed salary contracts can only be sustained in subgame perfect equilibria if they include the last period of the game $\mathcal{T}$. If we want to construct equilibria which entirely rely on freelance and fixed salary contracts, it would have to be a single contract which spans the entire duration of the game. Yet, the following proposition asserts that for freelance contracts, it is even impossible to be element of an equilibrium strategy.

Proposition 5. *In a subgame-perfect equilibrium, multi-period freelance contracts cannot be sustained.*

Proof. Take any period $t > t_0$ of a multi-period freelance contract beginning in t_0. In this period the principal can propose a new contract respecting activity-markovian continuation payoffs which implies an additional positive payment to the principal. The contract will be accepted by the agent and lead to a higher payoff to the principal. Foreseeing the renegotiation at t, the agent is not willing to pay the price in advance for the periods following t. Thus, periods following $t > t_0$ have to be excluded from the contract (for all t) and free-lance contracts can only last one period. □

With this proposition freelance contracts can be entirely ruled out. Unfortunately, a fixed pay contract which lasts the entire game (from the first to the last period) cannot be excluded. However, selling one's labour until the end of one's life has a smack of slavery. An extension of the model which would rule out optimal multi-period fix-pay contracts would be an exogenous probability that the relation breaks up so that the agent has to live without the support. If there is sufficient slack in (4), the agent will exert the high activity as a precautionary measure.

Some important results from the finite game carry over to the infinite horizon: a sequence of one-period control contracts is a subgame-perfect equilibrium. It generates the highest possible payoff which the principal can ever achieve in a subgame perfect equilibrium; so while other equilibria may exist, they cannot improve the situation of the principal. If one abstracts from self-enforcing arrangements and limits attention to formal contracts in the infinite game, the only candidate yielding the same payoff as control contracts is an everlasting fix salary contract. This imposes a strong commitment requirement which seems rather unrealistic. In general, we believe the infinite game to be less apt to the application at hand, since the number of opportunities for a worker to sign labour contracts is limited.

5 Discussion and conclusions

We have presented a hidden action model in which the agent can influence his future outside option by his present action. In such a setting, it was shown that a freelance contract (selling the shop) does not yield the optimal payoff to the principal, although the agent is risk neutral and not credit constrained.

This result implies that transferring entrepreneurial responsibilities may not always be desirable. The implications go even further. Consider the owner of a special asset which she can neither exploit on her own nor sell, for example personal contacts, networks, good will, etc. If she wants the asset to be exploited and delegates this task, then she must retain sufficient control, otherwise she will forfeit some of her surplus in the future. The reason is that intensive use of the asset by the agent allows him to improve his future bargaining position.

Thereby, this article also contributes another explanation for why firms exist without resorting to an insurance motive which is usually evoked in the principal agent context. Here, the boundary of the firm is defined through activities of the described type which cannot be outsourced and be taken care of by markets if the owner of the asset wants to optimally attend her interests.

References

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Thoery of Discounted Repeated Games with Imperfect Monitioring," *Econometrica*, 58(5), 1041–1063.

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

Baker, G., R. Gibbons, and K. J. Murphy (2002): "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics*, 117(1), 39–84.

Fernandes, A., and C. Phelan (2000): "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory*, 91, 223–247.

Fudenberg, D., B. Holmström, and P. Milgrom (1990): "Short-Term Contracts and Leng-Term Agency Relationships," *Journal of Economic Theory*, 51, 1–31.

Grossman, G. M., and E. Helpman (2002): "Integration versus Outsourcing in Industry Equilibrium," *Quarterly Journal of Economics*, 117(1), 85–120.

Holmström, B. (1999): "Manegerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66, 169–182.

Ma, C.-T. A. (1991): "Adverse Selection in Dynamic Moral Hazard," *The Quarterly Journal of Economics*, 106(1), 255–275.

Marcet, A., and R. Marimon (1998): "Recursive Contracts," mimeo, can be downloaded at the University Pompeu Fabra, Barcelona.

Nöldeke, G., and K. Schmidt (1998): "Sequential Incestments and Options to Own," CES Working Paper 160.

Posner, E. A., and G. G. Triantis (2001): "Covenants Not to Compete from an Incomplete Contracts Perspective," Working Paper 137, Law School of the University of Chicago.

Salanié, B. (1998): *Economics of contracts.* MIT Press.

A Unobservability of past effort

This appendix shows that the agent's activities need not necessarily be observable by the principal in order for the link between past activity and future outside option to work. Assume that the principal does not know the past activity of the agent when proposing a contract in period t –this implies that the principal did not exercise her authority to implement a particular activity in the previous period $t-1$. Then, the principal has some a-priori beliefs p that the agent chose the high activity level and updates these beliefs using the signals from the preceeding period Z_{t-1}. Denote by p_1 the updated belief that the agent has choosen $a_{t-1} = H$ after the signal realisation z_{t-1} was $z_{t-1} = 1$ and by p_0 the respective belief after the signal realisation $z_{t-1} = 0$. Then Bayes' rule leads to:

Lemma 3. $p_1 = \frac{P(H)p}{P(H)p+P(L)(1-p)}, \quad p_0 = \frac{(1-P(H))p}{(1-P(H))p+(1-P(L))(1-p)}.$

From this lemma and the fact that $P(H) > P(L)$, one can deduce that $p_1 > p_0$ for any a priori beliefs.

Given these beliefs, the principal has essentially the choice between offering the agent a high or a low base payment. If she offers an agent who exerted the high activity a low payment, the agent will reject but if she offers the agent who exerted a low activity the high payment, the agent will accept and will be "overpaid". The principal prefers to offer the agent the low payment if her expected payoff is larger than when offering the high payment: $(1-p_i)(\Phi(1,L,L)-\Phi(0,L,L)) > p_i\Phi(1,L,H)+(1-p_i)\Phi(1,L,L)-\Phi(0,L,H)$, where i is the realisation of the signal that the principal has observed. Thus, one can define $\hat{p}$ to be the threshold value below which the principal offers a low base payment: $\hat{p} := \frac{\Phi(0,L,H)-\Phi(0,L,L)}{\Phi(1,L,H)-\Phi(0,L,L)}$. If $p_0 < p_1 < \hat{p}$ or $\hat{p} < p_0 < p_1$, differences in the past activity exerted by the agent do not translate into different proposals by the principal. An equilibrium where future payoffs depend on previous activity requires: $p_0 < \hat{p} < p_1$. From the composition of p_0 and p_1 in Lemma 3, one can see that this condition is fulfilled when the discriminatory power of the signal is sufficiently large, i.e. $P(H)$ is close to one and $P(L)$ sufficiently small. Only then, the principal will offer the low payment after having seen the low signal and the high payment after a high signal. Otherwise payments will be constant in the signal. Having ensured that the principal reacts to different past activity levels, it is also necessary that the agent prefers high activity to low activity when being supported for the model to work. That is, we need the following analogon to inequality (4):

$$\forall\, a\,, a' :\ c(1,H,a) - c(1,L,a) < \delta(\Phi(0,a',H)P(H) - \Phi(0,a',L)P(L)).$$

This inequality is also fulfilled, if the discriminatory power of the signal is sufficiently large. Conclusively, the principal adjusts the offer to the previous period's activity of the agent and the agent reacts to this adjustment as if the principal knew the activity.

B Sufficiency of simple contracts

In the text, it is claimed that it is not necessary to consider contracts which condition the identity of the recipient on signals and that one can limit attention to contracts which condition transfers only on current signals. Such contracts are called *simple*.

Definition 2. *A contract $\mathcal{C}$ is called* simple *if and only if*

$$\forall\, \mathcal{C}_t \in \mathcal{C}:\ \tau_t(Z_t, \xi_{t-1}) = \tau_t(Z_t) \text{ and } r_t(\xi_{t-1}) = r_t.$$

The following lemma claims that focusing on the class of simple contracts is legitimate.

Lemma 4. *For any contract $\mathcal{C}$, there is a simple contract $\tilde{\mathcal{C}}$ which generates the same expected payoff to principal and agent and induces the same actions.*

Proof. To construct a simple contract, we consider any period t with a signal that determines the recipient or transfers in a later period t'. Transfers and the recipient rule in both periods are then replaced by transfers which yield the same expected payoffs to both parties and induce the same actions. Compute the expected net present value for the agent given he exerts high activity and low activity under the contract $\mathcal{C}$ in period t. Call those quantities V_t^H and V_t^L. Now, take a pair of transfers which only depend on the current signal: $(\tilde{\tau}_t^0, \tilde{\tau}_t^1)$, where $\tilde{\tau}_t^0$ is paid when the signal realisation is $z_t = 0$ and $\tilde{\tau}_t^1$ is paid when the realisation is $z_t = 1$. To replicate the incentives of the contract $\mathcal{C}$ in period t, expected payoffs under this transfer scheme must be equivalent to the respective expected net present value:

$$\begin{aligned} \tilde{\tau}_t^0(1 - P(H)) + \tilde{\tau}_t^1 P(H) &= V_t^H \\ \tilde{\tau}_t^0(1 - P(L)) + \tilde{\tau}_t^1 P(L) &= V_t^L. \end{aligned}$$

Solving for the transfers yields:

$$\tilde{\tau}_t^0 = \frac{V_t^L P(H) - V_t^H P(L)}{P(H) - P(L)} \quad \tilde{\tau}_t^1 = \frac{V_t^H(1 - P(L)) - V_t^L(1 - P(H))}{P(H) - P(L)}. \quad (14)$$

Then, subtract a fixed amount from both transfers so that the new expected payoff in period t to the agent is identical to the old expected payoff.

Transfers in period t' (which conditioned on the signal Z_t) are replaced by transfers where Z_t is integrated out given the action a_t. To get rid of the recipient being conditioned on Z_t in period t', replace $r_{t'}(\xi_{t'})$ by $\tilde{r}_{t'}(\xi_{t'}) = r_{t'}(\xi_{t'})|_{z_t=1}$ and compensate the agent by adjusting the transfer:

$\tilde{\tau}_{t'} = \tau_{t'} + \pi(s_{t'}, a_{t'}, a_{t'-1}) \cdot [P(r_{t'}(\xi_{t'}) = A) - P(\tilde{r}_{t'}(\xi_{t'}) = A)]$. Both these operations leave the expected payoff to the agent unaltered.

Repeating this exercise for all t leads to a contract $\tilde{\mathcal{C}}$ which exactly replicates the incentives of $\mathcal{C}$ and has the same expected payoffs for principal and agent.

$\square$

IZA Discussion Papers

No.	Author(s)	Title	Area	Date
448	M. P. Keane E. S. Prasad	Inequality, Transfers and Growth: New Evidence from the Economic Transition in Poland	4	03/02
449	B. R. Chiswick P. W. Miller	Do Enclaves Matter in Immigrant Adjustment?	1	03/02
450	B. R. Chiswick Y. L. Lee P. W. Miller	Schooling, Literacy, Numeracy and Labor Market Success	1	03/02
451	B. R. Chiswick P. W. Miller	The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada	1	03/02
452	B. R. Chiswick Y. Liang Lee P. W. Miller	Longitudinal Analysis of Immigrant Occupational Mobility: A Test of the Immigrant Assimilation Hypothesis	1	03/02
453	P. Carneiro K. T. Hansen J. J. Heckman	Removing the Veil of Ignorance in Assessing the Distributional Impacts of Social Policies	6	03/02
454	W. Koeniger	Defensive Innovations	5	03/02
455	J. T. Addison C. R. Belfield	Unions and Establishment Performance: Evidence from the British Workplace Industrial/ Employee Relations Surveys	3	03/02
456	T. M. Andersen	International Integration, Risk and the Welfare State	2	03/02
457	M. Lindeboom M. Kerkhofs	Health and Work of the Elderly: Subjective Health Measures, Reporting Errors and the Endogenous Relationship between Health and Work	3	03/02
458	M. Galeotti L. J. Maccini F. Schiantarelli	Inventories, Employment and Hours	5	03/02
459	M. Biewen	The Covariance Structure of East and West German Incomes and its Implications for the Persistence of Poverty and Inequality	3	03/02
460	B. R. Chiswick Y. Liang Lee P. W. Miller	Family Matters: The Role of the Family in Immigrants' Destination Language Acquisition	2	03/02
461	M. Gurgand D. N. Margolis	Welfare and Labor Earnings: An Evaluation of the Financial Gains to Work	3	03/02
462	B. R. Chiswick Y. Liang Lee P. W. Miller	The Determinants of the Geographic Concentration among Immigrants: Application to Australia	2	03/02
463	E. Wasmer	Labor Supply Dynamics, Unemployment and Human Capital Investments	5	03/02
464	W. Schnedler U. Sunde	The Hold-Down Problem and the Boundaries of the Firm: Lessons from a Hidden Action Model with Endogenous Outside Option	1	03/02