

# Agency Contracts with Long-Term Customer Relationships

Ignatius J. Horstmann\*    Frank Mathewson†    Neil Quigley‡

## Abstract

In certain types of industries, contracts for sales agents include both commission payments for sales and clawbacks of these payments if existing clients are not retained. This paper provides a model that shows that contracts with these features arise in equilibrium in environments having: i) up-front selling costs that are re-couped from on-going sales, ii) heterogeneous customers, iii) limited sales agent access to capital markets and iv) imperfect commitment by customers and agents to long-term contracts. We test the model using information on insurance sales agent contracts from New Zealand prior to and after bank entry into the insurance sales market. Increased policy lapse rates for traditional insurers post bank entry indicates that banks were cream-skimming customers. Our model predicts that, in this case, bank entry should reduce the value of both the initial commission paid for sales and the clawback for policy lapses. The data support this prediction.

---

\*Rotman School of Management and Institute for Policy Analysis, University of Toronto

†Department of Economics and Institute for Policy Analysis, University of Toronto

‡Victoria University, Wellington



## 1. INTRODUCTION

In certain industries, incentive contracts for sales agents feature not just commission payments for sales to new and on-going customers but also clawbacks of commissions paid for new customers who subsequently fail to continue with the firm. Sales agent contracts in the life insurance industry are an example. Here agents receive a commission both on initial sale and upon renewal of a term policy; agents are subject to clawbacks of the initial sales commission if the customer allows the policy to lapse within the first several years. Similar arrangements occur for independent retailers/distributors of cellular phone services. In this case, the service provider pays a “bonus” to the retailer for each new customer connection. There is a (partial) clawback of that bonus if the customer cancels the phone service within the first year or so. Provisions for clawback of profit distributions to fund managers when the fund underperforms have become prevalent in the venture capital market. More generally Tallitsch and Moynihan (1994) indicate that, in addition to traditional performance measures such as total sales, “successful companies” may use indicators such as the proportion of potential accounts that place orders, the proportion of prospects that purchase and high sales per account in determining agent bonus payments. Performance indicators that may enter negatively into bonus payments are lost accounts and order cancellations.

The appearance in incentive contracts of either commission payments or payment clawbacks can be explained by standard contracting models (see Lazear (1995)). These models do not explain the simultaneous appearance of both commissions and clawbacks nor do they explain why these features occur in agent contracts in some industries but not in others. We address these issues in this paper, providing a contracting model in which the equilibrium contract necessarily contains both commissions payments for initial and renewal sales and clawbacks of initial commissions

payments for customers who subsequently fail to renew. The model identifies four elements in the customer-sales agent-firm relation that are key to the simultaneous emergence of commissions and clawbacks: i) costs of creating the initial customer-firm relation that are recovered through future sales, ii) customer heterogeneity, iii) limited agent access to capital markets, and iv) inability of the agent and the customer to commit to long-term contracts. We test the model using data on life insurance agent contracts.

To see how the above elements interact to produce both commissions and clawbacks, consider the case of insurance sales. The up-front costs include the costs of creating, underwriting and selling the policy. These costs are recouped via the initial premium payment *and by renewal payments*. The problem for the firm is that the insurance contract does not commit the customer to on-going payments. Rather, the customer can simply let the policy lapse at any point by not making a renewal payment. A consequence of these two features (items i) and iv) above) is that, if customers are heterogeneous in some dimension that is relevant to the decision to renew (item ii)), it is in the firm's interest to induce agents to exert not just sales effort but also effort identifying customers that are more likely to renew. Unobserved agent efforts at client recruitment and client retention confront the firm with two moral hazard issues in writing the incentive contract for sales agents. A scheme of commissions for initial and renewal sales and penalties for the breakdown of the relation (policy lapses) is one potential way of resolving these moral hazard issues.

The above elements, by themselves, are insufficient to explain the appearance of both commissions and clawbacks, however. A sell-out contract to the agent induces first-best effort and so is profit maximizing for the firm in this situation. Such a contract becomes infeasible, though, if the sales agent has limited access to capital markets (item iii)). Just as importantly, lack of agent access to capital markets combined with the fact that an agent's selling costs are all up-front costs means that,

whatever alternative incentive scheme is employed, it must serve a capital market function by providing income to cover the up-front costs. Here is where the final element comes into play. If the agent were able to commit to a long-term employment relation with the firm, then a simple commission scheme with loan repayment elements would be sufficient. Without such a commitment (item iv)), and given the other elements, the profit-maximizing incentive scheme must include both commissions and clawbacks. Essentially, if the agent can quit the firm and so forgo both renewal income and loan repayments, the firm rather than the agent bears the cost of too many bad prospects. Put differently, if quitting is possible, the agent doesn't properly internalize the cost of not identifying good prospects. The clawback for policy lapses solves this problem.

Our test of the model uses data from the life insurance market in New Zealand. In 1986, financial markets there were deregulated and banks began to sell life insurance in their branches. Data on policy sales and lapses indicate that bank selling changed the composition of the traditional insurers' customer pool by cream-skimming good prospects. Our model predicts that, in response to the banks' superior ability at identifying those more likely to renew, traditional insurers should lower both initial sales agent commission rates and the level of clawbacks per lapsed policy . Examining the contracts offered by large insurance firms in New Zealand prior to and after bank entry, we find that contracts changed in precisely the way predicted by the model.

The remainder of the paper presents the model and our empirical results in more detail. The next section describes our model, while Section 3 derives the profit-maximizing agent contract. Section 4 details the application of the model to the New Zealand life insurance market and provides tests of the models predictions. Section 5 concludes. An Appendix contains some technical details of the analysis.

## 2. MODEL

Because the focus of our analysis is the contract that an individual agent obtains from the firm, we employ a highly stylized model of the firm. In this model, the firm is merely a collection of matches of individual agents with sets of potential customers. The firm sells contracts to customers via sales agents with whom it also contracts. Firms are identical and operate in competitive markets, earning zero profits. Agents are heterogeneous in their ability to generate outside income and so the more able agents earn scarcity rents under the competitively offered sales agent contract.

As to the details, consider the market for a homogeneous good sold by competitive firms in each of two periods, 1 and 2, and at per period price normalized to 1. For reasons unmodelled, the efficient way for a firm to sell the good is via a two-period contract with the customer. This contract commits the firm to delivering the good in each period at a per period price of 1. It commits the customer to purchasing the good only in the first period. The customer can withdraw from the contract in the second period at zero cost, in which case the firm does not deliver the good nor does the customer pay the second-period price. Since the competitive price is the same at all firms, a customer who withdraws from the contract withdraws from the market entirely (i.e., the customer purchases from no firm in the market). To write a contract with any given customer, the firm incurs a cost  $F$ , sunk at the time the policy is sold. The firm produces the good at a constant marginal cost normalized to 0. Firms are risk-neutral and have access to perfect capital markets.

An example of the above sort of product is a standard term life insurance policy. These policies, written for multiple years, remain in force as long as the insured individual makes the annual premium payments. The insured can choose not to renew at any time, in which case the policy lapses and the contract is no longer in force. The up-front cost to the insurance firm is the cost of creating and underwriting

the policy. The policy has some expected payout but for insurance to fit the model, this payout is normalized to 0; a positive payout is easily incorporated.

In order to sell its product (contracts), a firm employs sales agents. Each agent within a firm is matched with a continuum of potential customers of mass normalized to 1. Customers are heterogeneous both in their willingness to buy the good and in their likelihood of later withdrawing from the market. As a result of this customer heterogeneity and because the cost of writing the contract does not depend on whether it is maintained, a firm has an interest in inducing two kinds of agent activity: identifying those customers who are less likely to withdraw (the screening activity) and selling policies to these customers (the sales activity). Each of these activities requires an unobservable (to the firm) effort on the part of the agent and so exposes the firm to a moral hazard problem.

For the sales activity, the fraction of potential customers who purchase a contract initially is a random variable whose realization depends on agent sales effort,  $e_s$ . All agents are assumed *ex ante* identical in terms of sales ability and we define units so that  $e_s \in [0, 1]$  denotes the *expected* fraction of potential customers purchasing a contract. The cost to the agent of sales effort is given by  $e_s^2$ . Neither this cost nor an agent's level of sales effort is observable by the firm. The *realized* fraction of customers who purchase is observable by both firm and agent. A customer's decision to continue with the contract in period 2 is not affected by agent selling activity in either period.

The agent screening activity behaves similarly, with the fraction of customer continuations,  $\rho$ , being a random variable whose outcome depends on agent screening effort. To simplify the exposition, we assume that  $\rho$  is distributed uniformly on the interval  $[e_I, 1]$ , where  $e_I \in [0, 1]$  denotes agent screening effort. This screening technology is one in which screening removes the bad prospects from the sales pool. Again, agents are assumed *ex ante* identical in terms of screening ability and the cost of screening

is given by  $e_I^2$ . Neither the cost nor the level of screening effort is observable by the firm. The realized fraction of non-lapsed policies is observed by both the firm and agent.

Prior to the choice of either screening or sales effort, an agent and the firm enter into a long term (two-period) contract. This contract can be conditioned on both the fraction of initial purchases and the fraction of renewals. The contract is binding on the firm; however, the agent may choose to terminate the contract at any point at which it is no longer in the agent's interest to meet the contract terms. In the case of agent termination, the relation between the firm and the agent ends; in the case of no termination, the initial contractual terms are enforced. Both firms and agents are risk neutral and firms have access to perfect capital markets. Agents, by contrast, have limited access to external capital markets and are unable to borrow against future sales income to cover current costs. Because agents are liquidity constrained in this way, all contracts must have the feature that the expected payment in the first period is at least as large as an agent's expected cost of selling insurance.

The timing of activities in terms of contracting, the screening of customers, the sale of the product and the agent quit decision is summarized as follows:

1. In period 1, the firm and the agent enter into a two-period contract. The contract specifies a payment to the agent at the end of period 1 that can be conditioned on the observed fraction of customers purchasing contracts in period 1 and a payment in period 2 that can be conditioned on the observed fraction of continuing contracts in period 2.
2. Upon accepting the contract, the agent chooses effort,  $e_I$ , in screening potential customers for those more likely to continue and effort  $e_s$  in selling to this screened set. The market reveals the fraction of customers who purchase in period 1.

3. In period 2, the market reveals the fraction of contracts sold that continue (customers who exercise their option to continue define  $\rho$ ). Both the agent and the firm observe this outcome.
4. Having observed the realization of  $\rho$ , the agent decides whether to stay with the firm and to fulfill the contract terms or to exercise a costless option to leave (termination).

In undertaking sales and screening activities in period 1, an agent forgoes some alternative opportunity in that period whose utility value is denoted by  $W$ . Agents are assumed to be heterogenous in  $W$ , with the value distributed on the interval  $[\underline{W}, \overline{W}]$ . The value of  $W$  for any given agent is assumed unobservable by the firm.<sup>1</sup> An agent need not forgo  $W$  in period 2 to discover the fraction of continuing contracts. For any agent to be willing to participate in the contract, the expected two-period value of payments must be at least  $W$  (i.e., we assume no discounting). The competitive equilibrium contract maximizes an agent's expected two-period income subject to the capital market and participation constraints above and to the firm earning zero expected profits.<sup>2</sup>

### 3. THE CONTRACTING PROBLEM

In this section we derive the equilibrium contract. We show that, for a range of costs, this contract has a simple structure that features both a commission payments to the agent based on first and second period sales as well as a clawback of sec-

---

<sup>1</sup>This assumption is a simple way of capturing heterogeneity in agents' cost of engaging in sales activity. Heterogeneity of this form results in contracts that are uniform for all agents, as per observed contracts.

<sup>2</sup>If the contract failed to maximize an agent's expected income, then a new firm could enter, offer the expected income maximizing contract, compete away agents from other insurance firms and also make positive profits.



ond period payments based on contract cancellations. We also show that these two characteristics of the contract derive from four features of the environment: i) contract continuations are needed for the firm to cover initial costs of selling, ii) neither customers nor the sales agents commit to a long-term contract, iii) heterogeneity in customer continuations and iv) lack of agent access to capital markets. In the presence of moral hazard, but absent any one of these elements, an equilibrium contract would not need both commissions and clawbacks. We show, as well, that the equilibrium contract in our environment implements the full-information efficient levels of agent effort.

### 3.1 The Full Information Contract

As a bench mark, we first derive the equilibrium contract when agent effort levels are observable and contractible. In this case, the effort levels for an agent with opportunity cost  $W$  (a type  $W$  agent) are chosen to maximize total expected surplus from the agent/firm match. For the sale of the fixed good at a per period price of 1, this contracting problem is characterized as:

$$\max_{e_I, e_S} ES(e_S, e_I, W) = e_S \left[ 1 + \frac{1 + e_I}{2} - F \right] - e_I^2 - e_S^2 - W.$$

The chosen effort levels are then given as the solution to the equations:

$$\frac{e_S^*}{2} = 2e_I^* \tag{1}$$

$$1 + \frac{1 + e_I^*}{2} - F = 2e_S^*, \tag{2}$$

which yield:

$$e_I^* = \frac{3 - 2F}{15} \tag{3}$$

$$e_S^* = \frac{4(3 - 2F)}{15}. \tag{4}$$

The requirement that  $e_I^*, e_s^* > 0$  implies that  $F < 1.5$ .<sup>3</sup>

The set of individuals that become agents is given by all those agent/firm matches that generate non-negative surplus. The marginal agent is defined by the opportunity cost,  $W^*$ , such that  $ES(e_I^*, e_s^*, W^*) = 0$ . Substituting and solving defines this agent by the equation

$$W^* = \frac{(3 - 2F)^2}{15}. \quad (5)$$

For example, if  $F = .5$ , then  $W^* = .26$  so that individuals facing an opportunity cost  $W \in [0, .26]$  become agents under this contract (with inframarginal agents earning scarcity rents) while those with  $W > .26$  do not.

Finally, the up-front cost of creating and selling a contract is  $F + e_s^{*2} + e_I^{*2}$ . For capital market imperfections to play a role (specifically, for the firm to be unable to implement the full-information outcome with a sell-out contract to the agent), first-period expected sales revenues must fall short of costs. This condition is

$$e_s^*(1 - F) - e_s^{*2} - e_I^{*2} < 0. \quad (6)$$

Substitution from (3) and (4) above reveals that the condition is satisfied as long as  $F > .35$ . In what follows, we maintain this further restriction on  $F$ . The evidence on insurance agent contracts supports such a restriction. Specifically, the evidence is that agents receive more than their contribution to firm profits in early periods and less in subsequent periods (i.e., the contract provides a loan to agents based on future expected renewal income).

### 3.2 Equilibrium Contract with Unobservable Efforts

We now show that with agent efforts unobservable and so uncontractible, a contract linear in sales – the commission payments – and subsequent renewals – the clawback – can induce the full-information effort levels and levels of agent participation. With

---

<sup>3</sup>Still further restrictions on  $F$  appear below.

agent types unobservable, this contract provides the same expected payment to all agents. Since the full-information levels maximize two-period expected surplus, this linear contract will be the equilibrium contract if it yields zero expected profit for the firm and satisfies the first-period capital market constraint. We show that, for a broad range of  $F$  values, the linear contract satisfies these two conditions. We also discuss what contract is required for those values of  $F$  for which our contract fails.

The specific elements of the agent's contracts are as follows:

1. a linear sharing rule associated with first-period revenues, with the agent's share of revenues denoted by  $b \in (0, 1)$ .
2. a linear payment rule associated with cancelled contracts, with the payment per cancelled contract denoted by  $a$ . A positive value for  $a$  indicates a payment to the agent from the firm, while a negative value indicates a payment from the agent to the firm.
3. the agent is the residual claimant on any revenue from contract continuation.

Three features of this contract are worth noting. First, the details of the contract are independent of agent characteristics (like experience, for instance). This feature is not only what we observe for insurance agent contracts but also sufficient to implement full-information effort levels. Second, the contract is linear and all agent income comes from commissions. Again, this feature is consistent with observed insurance agent contracts which have little in the way of non-sales related income.<sup>4</sup> Linear contracts

---

<sup>4</sup>In contrast to the contract here in which the agent receives all second-period income, the typical contract gives the agent only a share of this income. Although not detailed here, one can show that there exists a continuum of contracts, indexed by a real number  $m \in [0, \bar{m} < 1]$  that generate the full-information effort levels. These contracts provide the agent with a share  $(1 - m)$  of each continuing contract and have a clawback of  $(a - m)$  on each laosed contract. The agent's share of initial sales,  $b$ , is the same in all of these contracts. This would give a contract more clearly aligned

are also standard in moral hazard problems when both the agent and the firm are risk neutral. Finally, if  $a < 0$ , the contract contains both commissions (items 1 and 3) and clawbacks (item 2).

This final feature merits some elaboration. Recall that, since first-period revenue fails to cover costs in the full-information contract and since an agent has no access to external capital markets, a sell-out contract is not feasible in this setting: an individual agent cannot afford to operate as a single proprietorship. Instead, the firm provides liquidity services through the contract. Given that the firm must borrow against expected second-period revenue to cover first period costs, it must recover some of the revenue in the second period to break even as it operates in competitive output markets. The firm has three possible avenues for doing so: i) a sharing arrangement with the agent based on second-period income, ii) a lump-sum charge against the agent's second-period income and iii) a charge against the agent based on contract lapses. Not surprisingly, a sharing arrangement reduces agent incentives for both selling and screening and so cannot, by itself, implement the full-information effort levels. Although less immediately obvious, the same is true of the lump-sum charge. Rather than lay out the argument in this case, we proceed to the analysis of the clawback on cancelled contracts and show that it is sufficient both to achieve loan repayment and to provide incentives for effort provision. From this analysis it will be clear that a lump-sum payment fails in this setting, so that a clawback is the only contract provision that can implement the full-information effort levels.

To show that the above contract both provides for loan repayment and implements the full-information outcomes, we first examine an agent's second-period decision problem. Recall that the contract is not binding on the agent so that, having observed with the insurance data. Also, in the model, neither the firm nor the agent discount the future. If the agent were to discount the future more heavily than the firm, the contract would again set the agent's share of second-period revenue at less than 1.

the cancellation outcome, the agent can always choose to terminate the employment relation. In this case, the agent forgoes any second period revenues as well as any payments based on lapses. Since the agent incurs no opportunity cost to observe the cancellation outcome, the agent terminates only if second period income is negative.

From the agent's quit decision we can define a critical fraction of renewals,  $\rho^0$ , such that the agent's second-period income is exactly zero:  $\rho^0 + (1 - \rho^0)a = 0$ . Since the firm must recover a portion of second period renewal income to satisfy loan repayment,  $a$  must be negative and so  $\rho^0 > 0$ .<sup>5</sup> Its value is given by the expression:

$$\rho^0 = -\frac{a}{1 - a} \quad (7)$$

The agent terminates the employment relationship if  $\rho < \rho^0$  and continues it if  $\rho \geq \rho^0$ . As long as  $\rho^0 > e_I$ , the agent terminates with positive probability under the contract.<sup>6</sup> A similar termination outcome would occur if the contract specified a lump-sum payment.

Now, consider the agent's screening decision. Given that the agent remains with the firm if  $\rho \geq \rho^0$ , the agent's *ex ante* expected pay-off from screening is:

$$EV_I = e_S ER - e_I^2,$$

where  $ER$  is the agent's second period expected income per contract renewal and is given by

$$ER = \frac{1 - \rho^0}{1 - e_I} \left[ a + (1 - a) \frac{1 + \rho^0}{2} \right].$$

In this expression, the term in brackets is the expected agent income net of clawbacks should the agent continue; the other term is the probability of the agent continuing.

---

<sup>5</sup>Were  $a \geq 0$ , then  $\rho^0 = 0$ . In this case, the agent never terminates the contract, keeps all renewal income and receives a payment from the firm on lapsed policies. The firm recovers no income in the second period, an outcome inconsistent with loan repayment.

<sup>6</sup>In what follows, we maintain this assumption and show later that this assumption holds for the optimal contract.

The agent's screening effort level solves  $\max_{e_I} EV_I$  and, with (7) substituted for  $\rho^0$ , is given by:

$$\frac{e_S}{2(1-a)} \left( \frac{1}{1-e_I} \right)^2 = 2e_I, \quad (8)$$

From (1), we have that  $e_I^* = e_S^*/4$ . Equation (8) then implies that, if the agent exerts the full-information level of sales effort, the agent also exerts the full-information level of screening effort as long as  $(1-a)(1-e_I^*)^2 = 1$ ; that is, as long as

$$a = \frac{e_I^*(e_I^* - 2)}{(1-e_I^*)^2}. \quad (9)$$

Since  $e_I^* < 1$ ,  $a < 0$ , implying that the full-information screening effort is implemented with a clawback. From (7), this clawback results in a value of  $\rho^0$  equal to  $e_I^*(2 - e_I^*) > e_I^*$ , so that the agent terminates the contract with positive probability. Finally, the larger the screening effort (higher  $e_I^*$ ), the larger are the clawbacks required to implement it ( $a$  more negative).

How do we understand these results? Because loan re-payment induces the agent to terminate the contract for some renewal realizations, the agent is only the residual claimant on effort for large enough renewal realizations. This limited liability presents potential incentive problems in implementing effort choices. By making cancellations costly for the agent in those cases in which the agent continues the contract, the clawback induces the agent to work hard at screening to avoid clawback payments. For a sufficiently large clawback, the agent screens at the full-information level. Note that a lump-sum scheme would have no such beneficial incentive effects.

What remains is to check that full-information sales effort and agent participation are implemented and that loan repayment is satisfied. As to the first, the agent's



pay-off from sales effort is defined by

$$EV_S = e_S(b + ER^*) - e_S^2, \quad (10)$$

where  $ER^*$  gives the value of second-period agent revenues at the screening effort  $e_I^*$ . The agent's choice of sales effort solves  $\max_{e_S} EV_S$  and is defined by the equation

$$b + ER^* = 2e_S^*.$$

Substitution for  $\rho^0$  from (7), for  $a$  from (9) and for  $ER^*$  reveals that the agent chooses the the full-information level of sales effort (see (2)) if  $b = 1 - F + e_I^*$ . Note from (3) that for  $b \geq 0$ ,  $F$  must be less than or equal to 1.4.<sup>7</sup>

We have then that the equilibrium contract satisfies the following two conditions:

$$\begin{aligned} a^* &= \frac{e_I^*(e_I^* - 2)}{(1 - e_I^*)^2} < 0 \\ b^* &= 1 - F + e_I^* > 0. \end{aligned}$$

This contract generates full-information agent participation and satisfies an agent's need for liquidity in the first period if the following two conditions are met:

$$EU_1^* = b^* e_S^* - (e_S^*)^2 - (e_I^*)^2 \geq 0 \quad (11)$$

and

$$EU_1^* + e_S^* ER^* = W^*. \quad (12)$$

Substitution of the equilibrium contract values into (12) yields the condition:

$$e_S^* \left[ 1 + \frac{1 + e_I^*}{2} - F \right] - e_I^{*2} - e_S^{*2} - W^* = 0.$$

---

<sup>7</sup>The contract specified can also implement the full-information efficient efforts for  $F \in [1.4, 1.5]$ ; however it will require  $b < 0$ , an outcome not observed in practice.

The left-hand side of this expression is the value of the firm in the full information case and so must be 0 for the marginal agent at  $W^*$ . Therefore, the contract implements the full-information agent participation rate. Note also that, since the expected payment to the agent is the two-period expected revenues *net* of the expected underwriting costs ( $e_S^*F$ ), the contract also satisfies loan repayment (the firm covers its underwriting costs).<sup>8</sup> As the firm's output market is competitive, superior inframarginal agents collect the full-information scarcity rents.

Finally, we should emphasize that the commission and clawback features of the contract rely both on the agent's lack of access to capital markets and on the agent's inability to commit to a long term contract. The former has already been discussed as it relates to the infeasibility of sell-out contracts. The agent's need for liquidity also means, though, that the firm can't simply backload the entire contract and pay the agent only in the second period based on observed outcomes. The result is that the firm must pay the agent in the first period and then implement some income recovery scheme in the second period. Provision of liquidity still wouldn't be sufficient to create the above contract if the agent couldn't terminate the contract in the second period. With commitment, the agent could always be made the residual claimant on all screening effort. Loan repayment could be achieved with a lump-sum transfer and so there would be no need for clawbacks. This combination of features therefore generates both commissions and clawbacks.

#### **4. TESTING THE MODEL: THE NEW ZEALAND LIFE INSURANCE MARKET**

As a test of the model, we examine the contracts for insurance sales agents employed by five large, multinational insurance firms in New Zealand. These five companies

---

<sup>8</sup>A simple calculation also shows that, for  $F \leq .6$ , the agent liquidity constraint (11) is satisfied. For  $F > .6$ , the firm must also make a lump-sum payment to the agent in the first period.

have a combined market share in excess of 50% as measured by premium revenue. We chose New Zealand because it presents something of a natural experiment for our model. Specifically, up until 1986, banks in New Zealand were effectively prohibited from selling insurance through their branches. Life insurance policies were sold either by independent agents or agents employed by insurance companies. Financial services deregulation in 1986 opened up this market for bank selling of life insurance. Bank entry had a major impact on the traditional life insurance companies: by 1996 the banks had a 25 percent share of the term life market and insurance companies had been forced to undertake a major restructuring of their operations, including their sales force remuneration practices (Horstmann, Mathewson and Quigley 1996).

We obtained data on the contracts for traditional insurance company sales agents both prior to and post bank entry. These data allow us to compare our model's predictions about the impact of bank entry on agent's contract with the actual changes in these contracts. To execute this comparison, we need to model the effect of bank entry on the traditional insurance market. In almost all instances, banks have demonstrated an interest in selling but not underwriting insurance. This suggests that a bank's advantage lies not in creating policies/pricing risks but in identifying and selling to good prospects. An example illustrates this point. A young couple making the decision to purchase their first home may seek a mortgage through a bank. To assess the accompanying mortgage risk, the bank would evaluate the income and existing debt obligations of the candidate household. This information lets banks more easily identify those households who need insurance and are more likely to maintain it. With superior prospects that emerge as a by-product of another financial service function, insurance sales and screening become a routine function that may be undertaken by bank employees.

If banks possess a superior insurance sales technology, then bank entry may affect traditional insurance sales in three possible ways. First, banks may siphon-off client

prospects across the entire distribution of client renewals. We refer to this effect as the scale effect. Second, banks may face a favorable client selection bias; that is, banks can easily sort on their clients to identify potential insurance purchasers who are more likely to continue their insurance coverage. A favorable selection bias for banks means an unfavorable client selection bias for traditional insurance sales agents. We refer to this effect as the cream-skimming effect. Third, with a less favorable set of potential customers from which to select, agent screening may become more critical to traditional insurance activity. Essentially, if the pool of potential customers consists predominantly of those who are likely to renew – the situation prior to bank entry – then screening provides a small marginal gain over simple random selection. As the fraction of non-renewers in the potential customer pool increases because of bank entry, the marginal return to screening (relative to random selection) may well increase.<sup>9</sup> We call this effect the screening-efficiency effect.

How do these effects alter agent contracts? In our model, a scale effect simply alters the size of the insurance company’s overall client pool and so the number of agents the company wishes to hire. Scale changes have no effect on any given agent’s contract as long as it is efficient for the company to assign each agent a potential client pool of measure 1. We maintain this assumption in what follows and so scale effects are neutral.

---

<sup>9</sup>As an example of a screening technology that has this feature, consider the following. Any given customer either renews with probability 1 or with probability 0. An agent can either draw customers at random from a finite pool of potential customers and sell to them or screen first. Screening involves drawing a subset of customers at random and conducting an interview with each that reveals with certainty the customer’s renewal probability (either 0 or 1). The agent sells to those customers having a 1. Those customers having a 0 are removed from the pool and the agent fills his client list by drawing at random from the remaining customer pool. In this case, screening has very little value over random drawing if the fraction of 1’s in the pool is large. If the fraction of 1’s is .5, then screening has much higher value.

We capture both of the other effects locally via a linear transformation of the lower bound of the distribution of renewals. Specifically, we assume that, for  $e_I$  in the neighborhood of  $e_I^*$ , this lower bound is given by  $e_I/\theta - \gamma$ , with  $1 \geq \theta > 0$  and  $1 > \gamma > 0$ .<sup>10</sup> The constant term  $\gamma$  captures the cream-skimming effect of bank entry. The assumption  $\gamma > 0$  implies that, all else equal, bank entry reduces the expected fraction of renewals. The  $\theta$  term captures the screening-efficiency effect of bank entry. In the case of  $\theta < 1$ , a marginal increment in screening effort raises the (now lower) average fraction of renewals more quickly than previously. Maintaining the assumption of a uniform distribution of customer renewals, we have that the net effect of bank entry is that the expected fraction of renewals becomes  $\frac{1}{2}(1 - \gamma + e_I/\theta)$ . We assume that bank entry has no direct effect on the existing sales technology of insurers.<sup>11</sup>

Given this specification, we proceed as before to define the expected joint surplus under full-information contracting (for a type  $W$  agent):

$$ES(e_S, \hat{e}_I) = e_S \left[ 1 + \frac{1 - \gamma + \hat{e}_I}{2} - F \right] - (\theta \hat{e}_I)^2 - e_S^2 - W$$

where  $\hat{e}_I = e_I/\theta$  measures agent screening effort in efficiency units and  $(\theta \hat{e}_I)^2$  defines the cost of these efficiency units. The full-information ‘revised’ effort levels are given by:

$$\begin{aligned} \frac{e_S^{**}}{2} &= 2\theta \hat{e}_I^{**} \\ \left[ 1 + \frac{1 - \gamma + \hat{e}_I^{**}}{2} - F \right] &= 2e_S^{**} \end{aligned}$$

---

<sup>10</sup>This transformation applies only locally in the sense that, for  $e_I$  near zero, it would imply the possibility of negative renewal fractions. One could make the approximation global by assuming that the original distribution of renewals is on the interval  $[e_I + \varepsilon, 1]$ ,  $\varepsilon > \gamma > 0$ . The normalization to  $\varepsilon = 0$  saves superfluous notation but requires either that the transformation adopted here not apply in the neighborhood of  $e_I = 0$  or that one allow for negative renewal fractions.

<sup>11</sup>As will be seen, bank entry has an effect on the equilibrium level of sales effort because of the change in the renewal prospects.

Solving these equations, we obtain effort levels of:

$$\begin{aligned} e_S^{**} &= \frac{4\theta(3 - 2F - \gamma)}{16\theta - 1} \\ \hat{e}_I^{**} &= \frac{3 - 2F - \gamma}{16\theta - 1}, \end{aligned}$$

with  $\theta > 1/16$  and  $3 - 2F > \gamma$  sufficient to guarantee that  $e_S^{**}, \hat{e}_I^{**} > 0$ .

The impact of the client selection bias on agent screening effort depends on the relative magnitudes of the cream-skimming effect (the size of  $\gamma$ ) versus the screening-efficiency effect (the size of  $\theta$ ). Specifically, if the screening-efficiency effect dominates, then  $\hat{e}_I^{**} - \gamma$  is larger than  $e_I^*$ ; conversely, if the cream-skimming effect dominates, then  $\hat{e}_I^{**} - \gamma$  is smaller than  $e_I^*$ . The same is true for  $e_S^{**}$  versus  $e_S^*$ .<sup>12</sup> In an Appendix, we show that the cream-skimming effect must dominate since, otherwise, it would pay the insurance firm to engage *ex ante* in cream skimming. We have, then, that  $\hat{e}_I^{**} - \gamma < e_I^*$  and  $e_S^{**} < e_S^*$ : both agent sales and screening effort decline. Correspondingly, the fraction of sold policies that lapse increases.

Our assumption that bank entry has an adverse selection bias on the client base for traditional insurers is supported by the data on policy lapses for banks versus lapses for traditional insurance companies. We examined quarterly data on lapsed policies sold by both banks and traditional insurance sales agents from the fourth quarter of 1987 to the fourth quarter of 1994. This period is the one during which most bank entry occurred.<sup>13</sup> Lapsed policies are term insurance contracts where the insurance customer ceased to remit premium payments before the end of the contract term.<sup>14</sup> If cream-skimming is dominant, then the lapse rate for term policies sold through banks

---

<sup>12</sup>In particular,  $\hat{e}_I^{**} - \gamma \geq e_I^*$  and  $e_S^{**} \geq e_S^*$  as  $(3 - 2F)(1 - \theta) \geq 15\theta\gamma$ .

<sup>13</sup>Our lapse data are compiled from the reports of insurance companies to the Life Office Association of New Zealand (now the Investment, Savings and Insurance Association).

<sup>14</sup>The data differentiate between on-going policies where premium payments cease (lapsed policies) and term insurance that reaches its termination date and is not renewed.



should be lower than the corresponding lapse rate for policies sold through traditional insurers. Furthermore, as banks begin with a new stock of policies, bank lapse experience should be constant over the sample period. Traditional insurers, however, will experience an adverse selection bias on new sales and so the corresponding lapse data over all contracts in force should reveal a deterioration over time. No such trends should be observed if the scale effect is dominant.

Looking first at mean lapse rates over the entire period, we find that the mean lapse rate is 1.30 (a 1.30% average lapse rate for term policies in effect in the 29 quarters of data) for bank sales and 1.77 for traditional insurer sales. A ‘t’ test on the difference between the means reveals a significantly higher average lapse rate for traditional insurers than for banks. The time series behavior reveals a similar pattern. In particular, a regression of the traditional insurers’ lapse rate against time for the 29 quarters of observations yields a regression coefficient on time of .011 with a corresponding ‘t’ statistic of 2.19, significant at the 5% level. A similar regression for lapse rate for bank sales of insurance over the same 29 quarters of observations yields a regression coefficient of .005 with a corresponding ‘t’ statistic of .87, insignificant at the 5% level. That is, the lapse rates for traditional insurers were on average higher than banks and increased at a rate of .011% per quarter. The lapse rate for banks revealed no significant increase over the 29 quarters of data. Both of these outcomes are consistent with bank cream-skimming.

What is our model’s prediction on the impact of bank cream-skimming on the agent’s contract? In the post-bank-entry world, the qualitative features of the equilibrium contract remain; we obtain a symmetric clawback (defined by  $\hat{a}^{**}$ ) as  $\hat{a}^{**} = 1 - 1/[1 - (\hat{e}_I^{**} - \gamma)]^2$ . The first-period revenue sharing parameter,  $\hat{b}^{**}$ , is given by  $\hat{b}^{**} = 1 - F + \hat{e}_I^{**} - \gamma$ . Comparing  $\hat{b}^{**}$  with  $b^*$  we see that whether or not cream skimming by banks leads to more or less high powered contracts depends on whether  $\hat{e}_I^{**} - \gamma \geq e_I^*$ . From above, we have that  $\hat{e}_I^{**} - \gamma < e_I^*$ , so the clawback (a negative

number) is less negative (smaller in absolute value) than the clawback in the absence of the competing technology. Similarly, the first-period commission is smaller.

To test this prediction, we collected data on the average commission and clawback rates for the period 1980-84 (prior to bank entry) and in 1996.<sup>15</sup> Here the test is on the parameters of the sales contracts for the traditional insurers between the two periods, prior to and after bank entry. We let the total commission rate over the first two years correspond to the initial commission  $b$  since one of the five companies providing us with contracts spread their initial commission payment over two years. For the 1980-84 period, agents selling multi-year term life policies had initial commissions over the first two years of a policy's life that aggregate to 60% of the annual premiums. Commissions for year 3 and beyond were 2% of the annual premium for the life of the policy. By 1996, commissions in the first two years of a policy's life aggregated to 47% of the annual premium, but commissions from year 3 and for the remainder of the life of the policy were 3.8% of the annual premium. Consistent with the model, initial commission rates fell with bank entry.<sup>16</sup>

Clawback provisions in the pre-bank entry period were 100% for policy lapses in the first 6 months of the policy's life. This clawback was reduced by 5.65% per month, falling to zero at 24 months. These provisions were unchanged after bank entry.

---

<sup>15</sup>The data are the average commission and clawback rates for 5 multinational traditional insurance companies operating in New Zealand. These companies have a current combined market share (as measured by premium revenue) in excess of 50 percent of the insurance market. Each company supplied us with agent commission schedules operative in the period 1980 - 84 (prior to bank entry) and in 1996.

<sup>16</sup>Our model predicts that renewal commissions (at 100%) are independent of bank entry. In the data, renewal commissions are considerably less than 100% and, as noted above, rise from 2% pre-bank entry to 3.8% post-bank entry. As mentioned previously, if the agent discounts the future more heavily than the firm, then the efficient contract has renewal commissions of less than 100%. Further, a shift to smaller up-front commissions and larger renewal commissions post bank entry induces the desired reduction in agent effort.

While these contract features remained constant, the value of initial commissions fell and so the amount of the clawback for lapsed policies also fell. Again, this outcome is consistent with the model.

Finally, we note that the model also makes a prediction about the fraction of agents that quit after the initial period. Specifically, it predicts that the quit fraction should fall. To see this note that the probability that the agent quits the firm is given by  $[\rho^0 - (\hat{e}_I^{**} - \gamma)] / [1 - (\hat{e}_I^{**} - \gamma)]$ . Substitution for  $\rho^0$  defines this probability as  $(\hat{e}_I^{**} - \gamma)$  so that, if the clawback decreases, so does the quit probability. Unfortunately, there are no data available for New Zealand that allow for a test of this proposition. Interviews with insurance company managers in 1996<sup>17</sup> confirmed that in the future each traditional insurer intended to hire fewer new agents and to retain more of the agents that they hired as a response to the entry of the banks.

## 5. CONCLUSIONS

This paper provides an explanation for sales contracts that contain both commission payments and clawbacks. In a competitive environment, contracts with these features arise as a response to significant up-front selling costs that make customer retention valuable, imperfect commitment by customers, incomplete contracts with agents who may put too little effort into selecting and retaining customers, and the agent's limited access to external capital markets. Commissions induce agent sales effort, while penalties provide incentives for the agent to identify customers who will continue to pay premiums over the life of the policy. Thus, the equilibrium contract with agents uses a combination of commissions to reward sales and clawbacks to sanction failure to retain the business of the customers who are recruited.

Data on life insurance lapse rates and agent sales contracts before and after banks

---

<sup>17</sup>Interviews were conducted with managers of the 5 multinational traditional insurer who provided us with data on agents' commissions and with the Chief Executive of the Life Office Association.

began to sell life insurance in New Zealand offer support for our assumptions and allow a test of our model. Policies sold by banks had lower lapse rates while lapse rates on policies sold by the traditional life insurers increased after bank entry. The entry of banks with a superior screening technology for the identification of prospective purchasers of life insurance results in the traditional life insurance companies changing their contracts with sales agents to make both the commissions for initial sales and the clawbacks for policy lapses smaller.

While our empirical results apply specifically to life insurance agent sales contract, we believe that the insights from our analysis apply more generally. That is to say, firms whose products and services require client-specific initial investments will seek a return through future sales from these same clients. When selling agents are the key to assembling and retaining a client base, our analysis illustrates why and how a profit-maximizing compensation mechanism exhibits commissions for initial success at client recruitment and penalties for subsequent failure at client retention. In a setting where on-going client relationships are critical to success, the introduction of competing distribution technologies may not eliminate the incumbent sale technology but will impact the structure of the existing contract for the traditional agents. Our predictions of smaller initial commissions and reduced subsequent penalties for these traditional agent contracts should generalize.

## REFERENCES

- [1] Rajiv Banker, Seok-Young Lee, Gordon Potter and Dhinu Srinivasan (1996) “Contextual Analysis of Performance Impacts of Outcome-Based Incentive Compensation,” *Academy of Management Journal* 39(4): 920 – 948.
- [2] Ignatius Horstmann, Frank Mathewson and Neil Quigley (1996) *Ensuring Competition: Bank Entry to the Insurance Industry* (Toronto: C D Howe Institute).
- [3] Edward Lazear (1995) *Personnel Economics*, Cambridge: MIT Press.
- [4] Edward Lazear (1998) *Personnel Economics for Managers*, New York: John Wiley & Sons, Inc.
- [5] Edward Lazear (1999) “Output-Based Pay: Incentives or Sorting?” Working Paper 7419, *National Bureau of Economic Research*.
- [6] John Tallitsch and John Moynahan (1994) “Fine-Tuning Sales Compensation Programs,” *Compensation and Benefits Review*, 26(2): 34 – 37.

## APPENDIX

Define the Kronecker  $\delta = 1$  in the presence of the alternative bank sales technology and  $\delta = 0$  in its absence and rewrite the expected joint surplus for the firm and the sales agent under full-information contracting (for a type  $W$  agent) as:

$$ES(e_S, e_I) = e_S \left[ 1 + \frac{1 - \delta\gamma + e_I + \delta e_I(1 - \theta)/\theta}{2} - F \right] - (e_I)^2 - e_S^2 - W$$

As a first-order effect with the sales contract implementing first-best agent sales and screening efforts, we have

$$\frac{dES(e_S, e_I)}{d\delta} \Big|_{e_S^*, e_I^*} = e_S^* \frac{-\gamma + e_I^*(1 - \theta)/\theta}{2}$$

The presence of the alternative sales technology must decrease expected surplus; otherwise expected joint surplus could be enhanced by the firm's *ex ante* truncating the upper end of the client prospect distribution by  $\gamma$  and invoking enhanced agent screening by  $\theta$ . Therefore

$$\begin{aligned} \frac{dES(e_S, e_I)}{d\delta} \Big|_{e_S^*, e_I^*} &< 0 \\ &\Leftrightarrow \\ e_I^*(1 - \theta) &< \theta\gamma \end{aligned}$$

Starting with the first-best screening effort absent the competing technology (i.e., with  $\delta = 0$ ), we substitute for  $e_I^*$  to obtain

$$(3 - 2F)(1 - \theta) < 15\theta\gamma$$

Thus from this and the above conditions

$$\begin{aligned} e_S^{**} &< e_S^* \\ \hat{e}_I^{**} - \gamma &< e_I^*. \end{aligned}$$