# EXISTENCE OF OPTIMAL NASH CONTRACTS
# OF A PRINCIPAL-AGENT MODEL
# SINGLE AND MULTIPERIOD CONSIDERATIONS

By

Peter Cheng

A DISSERTATION

Submitted to
Michigan State University
ı partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Accounting

1982

ABSTRACT

EXISTENCE OF OPTIMAL NASH CONTRACTS
OF A PRINCIPAL-AGENT MODEL
SINGLE AND MULTIPERIOD CONSIDERATIONS

BY

Peter Cheng

This dissertation attempts (1) to construct an abstract, general yet analytically rigorous agency model of a business entity, and (2) to study the conditions or circumstances such that the principal can design an optimal contract with his agent in both the single and multiperiod settings.

In the single period model, it is shown that if the contract to be negotiated belongs to the class of totally bounded functions or to that of monotonic increasing functions, the principal can be successful in arriving at an optimal contract.

The multiperiod model is constructed as a sequential decision process on a discrete time basis. Dynamic programming algorithm is a suggested solution procedure to the problem. Under various assumptions, it is shown that the proposed model meets the hypothesis

of the algorithm. The questions of the existence of optimal or nearly optimal contracts, uniformly optimal contracts and stationary optimal contracts as well as the convergence of the algorithm are investigated.

In the multiperiod model, the basic agency problem is first considered on a "most well-behaved" setting. This means that the payoff outcomes are observable by both the principal and the agent. Analysis is carried on both a finite and an infinite time horizon. Then, it is assumed that the payoff is observable by the agent alone while the principal receives a signal on the payoff. The signal is chosen by the agent and there is no restriction on how the agent should report the payoff. This leads to an imperfect state information model.

The imperfect state information model is analogous to the reporting and auditing functions of an entity. If appropriate measurability restrictions are imposed on the various functions, it is shown that if the auditor is acting in the best interest of the principal, the audited financial statements are adequate for the derivation of a long run optimal contract through the dynamic programming algorithm.

To
Christina, Nicholas and Cecilia

## ACKNOWLEDGMENTS

I wish to thank my committee members, Dr. Ronald Marshall, Dr. Kenneth Janson and Dr. Peter Lappan, Jr., for their continual help and guidence throughout the years of research on the dissertation. Special thanks are given to Dr. Lappan who patiently taught me mathematics and showed me how to apply mathematical analysis to study accounting and economic problems. Also, I have to thank Mary Reynolds for her excellent typing of the manuscript. Finally, I must acknowledge the patience and understanding of my wife Christina and children, Nicholas and Cecilia, from whom I took time normally available for family activities to work on this project.

TABLE OF CONTENTS

| CHAPTER | | | PAGE |
|---|---|---|---|

## CHAPTER I

## PRELIMINARIES

### 1.1: Introduction

An agency can be defined as an economic arrangement in which two or more individuals share the outcome produced by an action or a sequence of actions and the occurrence of some random event(s). The individual who wishes to delegate the action responsibility while receiving a share of the outcome is called the principal. The party who makes the action decisions is called the agent.

The study of the economic effects of various contractual arrangements between the principal and his agents has been of increasing interest in both the accounting and economic literatures. Amershi and Butterworth [1979] have identified three research lines of inquiry on the subject. One branch of study focuses on the welfare effects in terms of expected utilities of contractual aggrangements among economic agents. A number of writers have contributed to the literature of this approach: Demski and Feltham [1978], Harris and Raviv [1979], Holmstrom [1977,1979], Kobayashi [1980], Mirrlees [1974,1976], Ross [1973, 1974], Shavell [1979], Spence and Zeckhauser [1971], and Wilson [1968], among others. The question of market efficiency with respect to the ability of markets to absorb and transmit information about the activities of economic agents was studied by Akerloff [1977], Hirschleifer [1977], Spence [1974]. The third direction focuses on investigating the interactive effects of markets and contractual decentralization: Alchian and Demetz [1972], Fama [1980], Jensen and Meckling [1976].

This study concentrates on the first category. Focus is placed on identifying the set of sufficient conditions to guarantee the existence of optimal contracts between the principal and his agent in both a single period and multiperiod settings. In the multiperiod model, emphasis is placed on the existence of long run contracts. Although the formulation of the model does not preclude the consideration of a two period agency, the thrust of the research is to investigate contracts over the life of the entity. The multiperiod model is developed as a sequential decision process on a discrete time bases. Contracts are negotiated at the end of each period after the payoffs are observed or reported by the agent. There is no restriction that contracts for subsequent periods be of the same form as those of the previous periods. Under such a scenario, I would like to study the circumstances, both economic and behaviorial, under which the principal would be successful in designing a contract which would maximize his expected utility on the net return while at the same time allowing the agent to maximize his own expected utility by selecting an optimal action choice.

## 1.2: Scope and Objective

The purposes of this research are: (1) to construct an abstract, general yet analytically rigorous agency model of a business entity, (2) to study the conditions or circumstances such that the principal can design an optimal contract with his agent in both the single and multiperiod settings. In a sole proprietorship or a partnership, the principals are the owners and the managers are the agents. For corporations, the stockholders as represented by the board of directors are the principals while the managers are the agents. The auditors are assumed to act always in the best interest of the principal; that is, they act cooperatively with the principal. It is also assumed that all principals within the company act congruously and have common utility preferences. Likewise, the agents within the entity also have common goals and utility preferences. The model in the subsequence chapters effectively focuses on the relationship between one principal and one agent.

Ng and Stoeckenius [1979] point out that, in general, not only is the manager's action not observable by the owner, the firm's payoff is also unobservable by him as well. Since this asymmetry of information exists between the principal and the agent, the latter is required to issue financial reports periodically to convey information concerning the firm to the owner. If the manager's remuneration is directly a function of his reported performance, then the manager will have incentive to misrepresent in his report. If auditing is effective in detecting the misrepresentation, then it is of value to the owner, provided that the audit cost is small enough.

Fama [1980] notes that the contracting literature is almost uniformly concerned with one-period models. He further argued that in a single period world, there can be no enforcement of contracts through a wage revision process imposed by the managerial labor market. Radner [1981] shows that the sequential observation of the agent's performance over time is by itself an effective monitoring device. Since any real world contracting process is dynamic, an agency model is incomplete if its dynamics are not studied. A complete dynamic general equilibrium analysis including the external labor market is beyond the scope of the research. I plan to study the partial equilibrium effects on the contracting behaviors between the principal and the agent. This means that the external labor market opportunities to the agent are represented by a constant exogeneously determined. The principal must pay him in such a way that the agent's expected utility on his compensation is greater than the given outside opportunity set.

In the single period model, it will be shown, in Chapter II, that if the contract to be negotiated is bounded or if it is nondecreasing (monotonic increasing) with respect to payoff, the principal can be successful in searching for the optimal one within the above two mentioned classes of contracts. Bounded contracts are reasonable since no principal would pay his agents an unlimited amount of compensation in excess of the payoff. Monotonic increasing contracts are those of the bonus type. Bonus has been an extremely common form of remuneration. Agents are paid according to the payoff outcomes. These two classes of contracts clearly describe some common contracting behavior of economic agents.

For the multiperiod model, there are no assumptions nor restrictions placed on the form and types of the long run contracts. All ad hoc assumptions and conditions are imposed either on the net payoff function (the payoff less the compensation paid to the agent), the total expected discounted return function (the net payoffs to the principal over the planning horizon discounted to period 0) and the probability distribution of uncertain events (the period payoffs and the reported payoffs). The first two are restrictions on the economic behavior of the entity, how its payoffs, per period and total, relate to each other and to the environment in which the entity operates. The conditions on probability distributions are more behaviorial in nature. The distributions capture the principal's beliefs of his expectation on the agent's performance and the manner the agent reports, that is, how truthful the latter is. These are discussed in details in Chapter III and subsequent chapters. It is the hope of this research that the general agency model developed will describe, both normatively and positively, the behavior of the principal and his agent in the above context. It is a positive model because if the conditions or assumptions are met and optimal contracts are found, the principal can predict the behavior of the agent in terms of his performance.

The imperfect state information model proposed in Chapter VII will attempt to describe the reporting and auditing functions in terms of a multiperiod agency model. The reduction of the imperfect state model to a perfect state model through the auditing process and the conclusion that the audited financial statement is sufficient to design a long term optimal contract confirms the current belief about the value of auditing. Hopefully, such a model will enhance some understanding of the process of financial reporting and auditing as they are related to the contracting procedures of the company.

Most of the current researches in the agency area are conducted by imposing certain ad hoc assumptions on the model. Some of these assumptions are descriptions of the economic environment of the entity and some are behaviorial in nature while others are for mathematical tractability of the model. Assuming a solution to the problem, that is an optimal contract, exists under the proposed assumptions, the researcher carries on either to characterize the assumed optimal contract or to draw implications from the optimal contract. There is not any documentation about the kind of environment and conditions under which a researcher can appropriately make such an assumption, the existence of an optimal solution. This project attempts to lay the foundation and investigate the fundamentals of the agency problem.

Every attempt in the research is made to impose the minimal amount of restrictions on the behavior of the various functions. The model will be described and analyzed in its most possible generality. It is hoped that any results derived under such hypothesis are applicable to a greater variety of situations. Due to the mathematical and technical nature of the analysis, discussions in the following chapters will be separated into a general and non-technical description of the analysis and its results in the non-starred chapters. The full technical model, with all theorems and proofs is presented in the starred chapters. The two appendices contain materials which are well-known in the literature but are crucial to the development of the model. They are collected there for completeness.

However, by developing the model in its most general form, it becomes extremely difficult to give a direct and elaborate characterization of the optimal contract even after showing its existence. Characterization of the optimal contract requires additional hypothesis on the model. I do not intend to carry the analysis to such an extent. Also, this research will only discuss the conditions under which the principal can design an optimal contract and suggest, wherever possible, some procedure to arrive at or approximate the solution. It will not discuss the specific numerical aspects of the actual search for solution.

## 1.3: Organization of the Study and Summary of Results

The organization of the study is as follows. A general agency model with one principal and one agent is analyzed in Chapter II. A set of sufficient conditions for the existence of an optimal contract will be derived. Financial reporting and auditing are then introduced into the model. As indicated earlier, the auditor is assumed to act cooperatively with the principal. This implies that the auditor's decision variables are inputted into the model exogeneously and the auditor is acting as a "surrogate" for the principal. Auditing is essentially a dynamic process;

analysis of the auditing function is deferred until Chapter VII in the multiperiod model.

In the multiperiod model, the basic agency model is first investigated (Chapter III). The setting is the most "well-behaved" one in the sense that payoffs are observable by both the principal and the agent. The model is well-defined stochastically at time 0. This means that the sequence of payoffs over the entire planning horizon is defined stochastically given an initial payoff at the initial period. A probability distribution on this sequence of payoffs exists with a well-defined probability belief on the initial payoff. Analysis of the model is carried on a finite time horizon (Chapter IV) and infinite time horizon (Chapter V).

In the second part of the multiperiod analysis, it is assumed that the actual payoff is observable only by the agent while the principal will receive a signal on the payoff. The signal is chosen by the agent and there is no restriction on how the agent should "report" the payoff. This leads to an imperfect state information model (Chapter VII). The principal must attempt to assess the likelihood of the actual payoff given the signal produced by the agent. Hopefully, he can achieve such a task through the auditing and contracting processes. Serious mathematical problems arise under such a setting. These problems and their suggested solutions are discussed in Chapter VI.

The model constructed in the subsequent chapters imposes no differentiability restriction on the various functions. Heterogeneous probability beliefs between the principal and the agent are allowed. In the single period setting, it is first shown that under very mild conditions on the agent's choice set and utility function, the Nash constraint that the contract is incentive compatible is met for all incentive functions. Optimal contracts are shown to exist under two classes of functions, totally bounded and monotonic increasing functions.

The multiperiod model represents a sequential decision process. Long run optimal contracts are determined on the basis of maximizing the principal's total expected discounted net return (period payoffs less the agent's remuneration) over the planning time span. The model is considered under a finite and an infinite horizon settings.

Dynamic programming algorithm is an iteration procedure over time through which (1) it computes a conditional expectation; (2) the objective function in two variables (state and incentive) is optimized over one of these variables (incentive); (3) if an optimal contract is to be constructed, a "selector" which maps each state to a contract which achieves the maximum in the second step is to be chosen. It is shown that under certain circumstances such an algorithm is valid to solve the multiperiod agency problem in both the finite and infinite horizon models.

In the finite horizon model, optimal contract exists if the principal's total expected discounted net return is continuous and behaves "somewhat" linearly with respect to his period net return. In addition, the dynamic programming algorithm provides a stronger result. The optimal contract thus obtained maximizes the principal's period net return as well as his total expected discounted net return.

The infinite horizon model is discussed under two separate sets of assumptions. The first set includes conditions that the discount factor is less than unity and that the total discounted expected net return is bounded for all contracts. Existence of optimal contracts is established. Under the dynamic programming algorithm, the optimal contract derived is stationary, that is, the form of the contract remains the same throughout time. The second set of assumptions is that the total discounted expected net return to the principal is monotonic (either increasing or decreasing) with respect to time. Under both sets of assumptions, the infinite horizon solution is shown to equal to the limit of that of the finite horizon model when time is allowed to tend to infinity. Also, it is shown the dynamic programming algorithm implemented under these sets of conditions converges.

If appropriate measurability restrictions are imposed on the various functions, the dynamic programming algorithm can be implemented to the imperfect state information model. In this part of the analysis it is shown that if the auditor is acting strictly in the best interest of the principal and the audit is performed in the best possible manner, the audited financial statement is adequate for the derivation of a long run optimal contract through the dynamic programming algorithm.

CHAPTER II

SINGLE PERIOD MODELS

## 2.1: A General Agency Model

This section considers the situation in which there is a principal with one agent. The agent is entrusted with the task of selecting an action from among a set of alternatives. Then some random event occurs and the outcome of the action is some payoff which is assumed to be observable by both parties. The action choice, however, is not observable by the principal. A problem is then to decide on some "best" sharing scheme between the principal and the agent. Under the assumption that the agent's utility for wealth is positive and his utility for effort is negative, it has been shown that a pure wage contract is incentive incompatible, that is, the agent will always choose the act which requires the minimum level of effort (Ng and Stoeckenius [1979], Shavell [1979]). On the other hand, since the action choice is not observable by the principal, an incentive contract based on the outcome alone may cause moral hazard problems. Under such circumstances, Harris and Raviv [1979] have shown that some information about the agent's action by the principal is desirable even at some cost, namely, some monitoring is "good". The solution to the general agency problem is the determination of the action by the agent, an incentive function and a monitoring system by the principal to motivate the behavior of the agent.

The usual criteria adopted by writers in agency literature for the choice of optimal incentive contracts are Pareto optimality and Nash optimality. The Pareto criterion provides the principal with an expected utility at least as great as that which is obtainable from among alternatives which satisfy some minimal requirement (minimum security level) imposed by the agent. This implies that the principal and the agent decide cooperatively on the contract and the action choice such that the expected utility of the principal is maximized subject to the minimal requirement. If the principal can observe the agent's action choice as well as the payoff, or if there is full and truthful communications between the principal and the agent on the action and payoff, it would be sufficient to guarantee that the agent will act cooperatively. Some authors call this the first best solution (Holmstrom [1979], Shavell [1979]). However, a solution in the cooperative game does not necessarily imply that there is full and truthful communication of the action and payoff. The first best solution implies that the agent always acts in the best interest of the principal. He does not have a conflict of interest over effort with the principal either by the observability of the action or other monitoring means.

The Nash condition assumes noncooperative behavior between the principal and the agent. It requires an additional restriction, that the contract must enable the agent to choose an action from among the alternatives which maximizes his expected utility under the contract. This allows the agent to choose his actions freely based on his own choice criteria. A typical situation of such a phenomenon would be the asymetry of information between the principal and the agent. The principal does not observe the agent's action even after the payoff is observed. There is no way of enforcing any action choice on the agent. The best thing the principal can do then is to design the contract based on his "best" guess on the agent's action. The solution set of the Nash problem is a subset of that of the Pareto one. Hence there exist solutions which are Pareto efficient but are not Nash solutions.

The following is a general agency model. Define the expected utility of the principal as

$\phi_1(I,a) = E_1U_1(\omega(a,s) - I(\omega(a,s)))$, where

| | |
|---|---|
| $a \in A$ | set of possible actions |
| $s \in S$ | set of states of nature |
| $\Omega : A \times S \rightarrow R$ | payoff function $\omega(a,s) = \omega$ units of wealth. |
| $I : \Omega \rightarrow R$ | Incentive function $I(\omega(a,s))$. |

Throughout this paper, the same notation will be used to denote a function as an element of a functional space and the value of the function in the range. For example, $\omega$ denotes a payoff function as well as the units of wealth payoff for a certain action $a$ and state of nature $s$. The actual reference should be clear from the context.

Under most common economic circumstances, it will not be unreasonable to assume that the agent has only a finite number of mutually exclusive action choices. Then, without loss of generality, $A$ is assumed to be a finite subset of $R^n$. The set $S$ is defined with its usual statistical meaning in a decision theory context (Savage [1954]) with probability beliefs $P_i$ on $S$, where $i = 1$ denotes the principal and $i = 2$

denotes the agent. The usual homogeneity of beliefs between the principal and the agent is not assumed in this model. Let $E_i$ denote the expectation operators with respect to the beliefs of the decision maker.

It is assumed that the utility function of the principal, $U_1$, is non-negative, concave and monotonic increasing with respect to wealth. It can then be shown that such a utility function is continuous on its domain (Theorem 2.7).

For the agent, define his expected utility as

$$\phi_2(I,a) = E_2U_2(I(\omega(a,s)),a).$$

His utility function is also non-negative, concave, and monotonic increasing in wealth, but non-negative, concave, monotonic decreasing in effort. Assuming that both the principal and the agent are utility maximizers, one can formulate the problem as follows.

Maximize

$$E_1U_1(\omega(a,s) - I(\omega(a,s))$$

$I \in \{I\}$

Subject to: $\phi_2(I,a) \geq \nu$ (1)

$$a \in \underset{a \in A}{\operatorname{argmax}}\ E_2U_2(I(\omega(a,s)),a) \qquad (2)$$

$\nu$ represents the minimum levels of security for the agent to remain in the company. Solving the above program subject to constraint (1) will yield the Pareto solution. Constraint (2) represents the additional incentive compatibility condition that the Nash solution requires.

## 2.2: The Pareto Optimal Contract

Under the conditions of Pareto optimality, all parties are assumed to act cooperatively in the best interest of the company. The existence of a Pareto optimal contract is usually not too difficult to guarantee. Given sufficient regularity conditions, at least one solution is guaranteed. Wilson [1968] has demonstrated the existence of a Pareto optimal sharing rule and characterized the behavior of this sharing rule in a syndicate setting. Kobayashi [1980] analyzed the role of private information of an individual in the syndicate by redefining the core. He showed that an equilibrium contract belongs to the core as well as the existence of such equilibrium contracts.

Amershi and Butterworth [1979] investigated the problem assuming diverse beliefs among the principal and the agents. They analyzed the conditions for existence of the optimal contract and the characteristics of such an optimal contract. Most of their findings on the Pareto optimal contract are consistent with the results of previous studies.

## 2.3: The Nash Optimal Contract

In the formulation of the Pareto problem, one can assume some kind of convexity property for the objective function and the constraints are generally well behaved mathematically. Such assumptions will, to some extent, simplify the solution techniques. Under the Nash criterion, the additional constraint that the agent is able to maximize his expected utility removes any reason to impose the well behavior of the functions of the model. Also, one cannot be sure that the optimal contracts are differentiable. Differentiability is a requirement in most optimization techniques. The problem is compounded if the principal and the agent are allowed to hold diverse beliefs about the state of the world.

One common method in showing existence is by analyzing the first and second order conditions. Under the assumptions of Kuhn-Tucker in the nonlinear programming literature, these conditions are, in general, satisfactory. Another usual method is to use the Euler-Lagrange equation in the calculus of variation literature. In the agency setting, however, both of these techniques may not be appropriate. Both methods require differentiability of the incentive contracts. The principal's decision variable, incentive contract, is a function belonging to some functional space. A space is of finite dimension if its elements can be expressed as some finite number of vectors (or basis). There are only very few finite dimensional spaces that are of interest. In fact, the only one which is of any use is the space of polynomials of finite degree. The principal is effectively maximizing his expect utilities over infinite dimensional spaces. Under such handicapped conditions, most of the common calculus methods fail.

Both Mirrless [1974] and Gjesdal [1976] have correctly shown that the differentiability assumption may be too restrictive. Holmstrom [1977] constructed a counter-example, the optimal solution of which can be attained by a nondifferentiable sharing rule and no differentiable rule can precisely attain this solution.

One of the more extensive works in demonstrating existence is also done by Holmstrom [1977]. He proved the existence of two classes of incentive contracts under the Nash conditions. His work relied heavily on the assumption of homogeneity of beliefs between the principal and the agent. He also put some additional restrictions on the behavior of the functions to arrive at his results.

The first step of the analysis is to generate conditions on the agent's problem such that given any contract I, his expected utility function will always achieve a maximum. This will satisfy the Nash additional condition that the contract is incentive compatible.

To show incentive compatibility, the Weierstrass Maximum Theorem (Theorem 2.3) is used. Loosely stated, an upper semicontinuous function on a compact set achieves a maximum. Upper semicontinuity is a milder condition than continuity. When applied to the agent's expected utility function with respect to the action set, this means that if the agent shirks a very little bit, that is, if a is allowed to change by a small amount, then the agent's expected utility will not be increased by a large amount. Hence, upper semi-continuity simply means that the function is continuous from above. The agent should not expect a substantial gain if he changes his effort level slightly. This is reasonable in terms of expected utility when all possible states are considered.

Recall that A is assumed to be a finite subset of $R^n$, it is bounded by definition. Compactness on the real line means closed and bounded. All one needs for A to be compact is that it is a closed subset. Upper semicontinuity on $\phi_2$ requires some additional condition on the utility function of the agent. It can be shown that if $U_2$ is upper semicontinuous with respect to a, the action choice, then the resulting expected utility is also upper semicontinuous (Theorem 2.5).

With the above construction, indeed only very mild conditions, the Nash constraint is satisfied. Then the principal's problem is considered, again applying the Weierstrass Theorem on his expected utility function.

To guarantee upper semicontinuity on $\phi_1$ is easy. The principal's utility is monotonic increasing and concave with respect to wealth, which is generally the residual of the payoff after the manager's compensation is deducted. A concave function is always continuous (Theorem 2.7), and the integral, in our case, the expectation operator, is continuous if the integrand is continuous, which in turn implies upper semicontinuity.

Compactness of the space of incentive contracts is more difficult to show. The space of incentive contracts is of infinite dimensions. It would require stronger conditions on the contracts than just being closed and bounded. The mathematical requirement is that the contracts are totally bounded. Total boundedness means that no matter how the elements of the set (in this case set of functions) are grouped, they are always enclosed in a ball of some given radius. Here the radius has to be the same for all possible groupings. It is a kind of uniform bound for all possible payoff outcomes. Thus, if the incentive contracts are totally bounded, then the Weierstrass Theorem will guarantee the existence of a maximum for the principal.

In most situations, one would expect that the contracts should be increasing with respect to the payoffs or a bonus as it is more commonly called. The class of monotonic increasing contract is considered next. Each contract of this type is assumed to be bounded below. As the principal obtains the residual of the outcome after he pays the agent and his utility is increasing with wealth, given any wealth position from the outcome, he would choose the incentive function that would pay the agent the least amount. That is, the optimal contract in the principal's viewpoint can be defined as $I_n^* = \inf \{I_1, \ldots, I_n\}$ for each $\omega \in \Omega$ where $n$ denotes the number of possible contracts given $\omega$. As the problem is formulated, the agent must maintain a minimum security level in terms of expected utility to remain in the company, the incentive function should be bounded below for each $\omega \in \Omega$.

Now, consider the sequence $\{I_n^*\}$, the elements of which are defined earlier. This is a decreasing sequence and it is shown that it converges to a limit $I^*$ which is also a monotonic increasing function (Theorem 2.12). As mentioned earlier, the upper semicontinuity on $\phi_2$ on $A$ and the compactness of the action set $A$ guarantee the existence of an optimal action for the agent given any incentive contract. Thus, for each $I_n^*$, there exists a corresponding $a_n^*$ for the agent. The final step is to show that the sequence $\{a_n^*\}$ also converges to a limit $a^*$ which is then the optimal action for the agent if he is given the optimal contract $I^*$ (Theorem 2.13).

Up to this point, it has been shown that if the agent's action choice set $A$ is a closed subset of $R^n$, and if his expected utility function is upper semicontinuous with respect to $A$, there always exists an optimal $a^*$ such that his expected utility is maximized. This will satisfy the Nash constraint of incentive compatibility for all possible incentive contracts.

Two incentive spaces are then identified such that by choosing the appropriated contracts from these two spaces, it will be guaranteed that the principal's expected utility will achieve a maximum.

The first of these spaces is the class of totally bounded contracts. This is a very large class of functions which includes the differentiable functions and the equicontinuous functions which Holmstrom [1977] has demonstrated to be optimal under homogeneous belief assumptions. Functions from the totally bounded class are compact. The principal's expected utility function is continuous with respect to I. These two conditions will guarantee, by the Weierstrass Maximum Theorem, the existence of an optimal contract.

The second class of functions is the monotonic increasing incentive functions. A typical example of these functions is the bonus arrangement. Indeed it is shown that one can always pick an optimal monotonic increasing contract to maximize the principal's expected utility.

So far, an agency model has been described under the usual decision choice theoretical context. The assumptions imposed on the model, which are summarized above, describe a very reasonable and general economic environment. The restrictions on the utility function are in accordance with the usual von Morgenstern utility preference assumptions. If the contracts are totally bounded or they are monotonic increasing, the principal will always be able to negotiate a Nash equilibrium contract such that both his and the agent's expected utilities are maximized. In other words, if the agent exhibits a von Morgenstern utility preference with respect to wealth and effort, then the principal can always design a contract which is either totally bounded or monotonic increasing to achieve a Nash equilibrium.

## 2.4: Financial Reporting and Auditing

Generally, financial reporting means the conveyance of information to the principal (owner) about the outcomes of the actions taken by the agent (manager) in a period of time. When such a function is introduced into an agency model, it implies as well that the outcomes of the events are not observable by the principal. An additional objective of the principal is then to ensure that the agent is reporting the outcome truthfully. Ng and Stoeckenius [1979] have demonstrated that without monitoring, an incentive compatible contract always induces nontruthful reporting. This suggests that when the actual payoff and the agent's effort are not observable, information about the agent's performance is always valuable to the principal in terms of increasing his expected utility. Of course, the benefits derived from such information should be sufficient to justify its cost. Harris and Raviv [1979] showed that even imperfect information is beneficial.

As Baiman [1979] points out, there is a distinct difference between monitoring and auditing. The former means to verify the action taken by the agent while auditing is taken as the verification of the report of the outcome produced by the agent. In this paper, the role of the auditor is defined as another agent of the owner who "monitors" the reporting function of the manager, while the incentive contract is designed to minitor the actions of the manager. Since the objective of the principal is to control the reporting as well as the action choice by the manager, the incentive contract and the auditor can then together be considered as the monitoring system.

The "black" box of unobservables increases with the addition of the auditing function into the model. In most common situations, the original report submitted by the manager is not observable by the principal. The manager's report goes directly to the auditor who performs the necessary tasks on the report, suggests appropriate changes and adjustments and then attests the report. The principal will receive the "audited" report after all adjustments have been made or a "qualified" or "disclaimed" report if the manager refuses to make the suggested alterations. Under such a scenario, the only variable which is observable by all parties is the audited report. Any contract that is enforceable has to be based on variables which are observable by both the principal and the agent. Therefore, it is not unreasonable to suggest an incentive contract based on audited outcome. In fact, if the auditor performs all his audit tasks in the best possible manner and in the best interest of the principal, the audited outcome is shown in Chapter VII to be sufficient to design an optimal contract. The discussion of optimal contracts under a reporting and auditing environment will be deferred until Chapter VII when an imperfect state information model in a multiperiod setting is discussed.

# CHAPTER II*

## SINGLE PERIOD MODELS

### 2.1* The Model

Define the expected utility of the principal as $\varpi_1(I,a) = E_1U_1(\omega(a,s) - I(\omega(a,s)))$, where

| | |
|---|---|
| $a \in A$ | set of possible actions |
| $s \in S$ | set of states of the world |
| $\omega : A \times S \to R$ | payoff function $\omega(a,s) = \omega$ units of wealth |
| $I : \Omega \to R$ | incentive contract $I(\omega) = k_2$. |

The following are some standard assumptions on decision choice theory.

Assumption S.1: Let the action set $A$ be a closed, bounded subset of $\mathcal{A}$, where $\mathcal{A}$ is a normed linear space of dimension $n < \infty$.

Assumption S.2: The set $S$ is non-empty with a sigma-algebra $\sigma(s)$ defining the events in $S$, $P_i$, $i = 1$ (principal), $2$ (agent), are probability measures on $\sigma(s)$. It is assumed that all functions are measurable with respect to $\sigma(s)$.

Assumption S.3: (i) The utility function of the principal $U_1$, is assumed to be non-negative, concave, and monotonically increasing with respect to wealth.

(ii) The utility function of the agent $U_2$, is also nonnegative, concave, and monotonically increasing in wealth, but decreasing in effort.

Assumption S.4: There exists a $\omega_o$, $-\infty < \omega_o < \infty$, such that $|\omega(a,s)| \leq |\omega_o|$ for each $a \in A$ and $s \in S$.

Assumption S.1 describes the action choice set. More details about its topological structure will be derived in the later part of this section. Since focus is put on a single period model, at this stage of analysis, S can be viewed as deterministic. Hence no serious problem should arise on the measurability of the functions, which are assumed to be measurable. The assumptions on the utility function simply imply that the preference relation of the individuals is convex, complete and transitive. Assumption S.4 imposes a bound on the payoff function. In any economic situation, it will be very unlikely that the payoff is unbounded. One inherent consequence of Assumption S.4 is that the incentive function is also bounded as no principal can pay his agent an unspecified amount of compensation in excess of his return.

Thus, the expected utility of the agent can be defined as

$$\varphi_2(I,a) = E_2U_2(I(\omega(a,s)),a)$$

and the typical principal-agent problem represented by the following program:

$$\text{Maximize} \quad E_1U_1(\omega(a',s) - I(\omega(a',s)))$$
$$I \in C$$
$$\text{Subject to:} \quad \varphi_2(I,a') \geqq V \qquad (1)$$
$$a' \in \underset{a \in A}{\text{argmax}}\ E_2U_2(I(\omega(a,s)),a) \qquad (2)$$

C represents the incentive space and V represents the minimum level of security for the agent to remain in the company or the opportunity set he could attain outside the company. Solving the program subject to constraint (1) yields the Pareto optimal solution or commonly known as the first-best solution. Constraint (2) represents the additional condition that a Nash equilibrium requires.

<u>Theorem 2.1</u>: Let $(\mathcal{Q}, \|\cdot\|_1)$ be an n-dimensional normed linear space over the real field and $n < \infty$. Then $(\mathcal{Q}, \|\cdot\|_1)$ is topologically isomorphic to $(R^n, \|\cdot\|_2)$.

Proof: Refer to Larsen [1973].

Definition: A class of sets in a topological space is said to cover a given set X if and only if each point of X lies in at least one of the sets. If the diameter of each set in a cover of X is not greater than $\epsilon$, the class is called an $\epsilon$-cover of X.

Definition: Let X be a subset of a topological space. X is said to be compact if and only if every class of open sets which covers X has a finite subclass which also covers X.

Definition: A subset X of a complete normed space is said to be sequentially compact if and only if every sequence in X contains a convergent subsequence with limit in X.

It can be shown that in a Banach space, compactness and sequential compactness are equivalent. The following is a well-known and useful result of a finite dimensional space.

Theorem 2.2: Let $\mathcal{A}$ be an n-dimensional normed linear space over the real field with $n < \infty$. Then

(i) $\mathcal{A}$ is a Banach space

(ii) If $A \subset \mathcal{A}$ is a closed bounded set, then A is compact.

Proof: Refer to Rudin [1973].

Theorems 2.1 and 2.2 establish the structure of the action choice set and guarantee its compactness. By Theorem 2.1 and without loss of generality, A is assumed, for the rest of this paper, to be a closed subset of $R^n$ with an appropriate norm and all results can then be generalized to other topological spaces.

## 2.2* Totally Bounded Incentive Functions

In this section, by considering the requirements of the Weierstrass Maximum Theorem, a set of sufficient conditions is derived for the existence of an optimal contract under the conditions established in Section 2.1*.

Definition: A real-valued function f defined on a normed space X is said to be upper semicontinuous at $x_o$ if and only if $\limsup_{x \to x_o} f(x) \leqq f(x_o)$.

Theorem 2.3 (Weierstrass Maximum Theorem): An upper semicontinuous function on a compact subset X of a normed linear space achieves a maximum on X.

Proof: Refer to Luenberger [1969].

First consider the agent's problem. The idea is that given any incentive contract, we want the agent to be able to select $a \in A$ such that his expected utility function will achieve a maximum. This will satisfy the Nash additional condition that the contract is incentive compatible. Since A is a closed and bounded subset of $R^n$, it is compact by Theorem 2.2 (ii). All that remains to be shown is that $\varpi_2$ is upper semicontinuous on A.

Assumption S.5: For each fixed $s \in S$ and $\epsilon > 0$, there exist a $\delta(s) > 0$ such that $\delta : S \to R$ is measurable and $U_2(I(\omega(a,s)),a) < U_2(I(\omega(a_o,s)),a_o) + \varepsilon$ whenever $\|a - a_o\| < \delta(s)$.

<u>Theorem 2.5</u>: Let $I(\omega(a,s))$, $\omega(a,s)$ be bounded measurable functions and $\mu$ be the measure of S. Suppose for each fixed $s \in S$ and each $\epsilon > 0$, there exists a $\delta(s) > 0$ such that $\delta : S \to R^+$ is measurable and $U_2(I(\omega(a,s)),a) < U_2(I(\omega(a_o,s)),a_o) + \epsilon/2$ whenever $\|a - a_o\| < \delta(s)$. Then $\varpi_2 = \int_S U_2(I(\omega(a,s)),a)P_2(s)ds$ is upper semicontinuous on A.

Proof: Let $\varepsilon > 0$ be given. For each $s \in S$ define

$$\delta_o(s) = \sup \delta(s).$$

Clearly $\delta_o(s)$ is measurable since $\delta(s)$ is measurable. Let

$$S_o = \{s \in S : \delta_o(s) > 1\}$$

and

$$S_n = \{s \in S : \delta_o(s) < \frac{1}{n},\ n = 1,\ldots\}.$$

Since $\delta_o(s) > 1$ for every $s \in S$

$$S = \bigcup_{n=0}^{\infty} S_n$$

$\therefore$ For $\gamma > 0$ to be chosen later, there exist a subset $S'$ of $S$ such that $\mu(S - S') < \gamma$.

Hence there is a $k$ such that $S' = \bigcup_{n=0}^{k} S_n$ and

$$\delta_o(s) < \frac{1}{k} \quad \forall\ s \in S'.$$

Let $\delta' = \frac{1}{k}$. Then

$$U_2(I(\omega(a,s)),a) < U_2(I(\omega(a_o,s)),a_o) + \epsilon/2$$

whenever $\|a - a_o\| < \delta'$ and $s \in S'$. This implies

$$\int_{S'} U_2(I(\omega(a,s)),a)P_2(s)ds < \int_{S'} U_2(I(\omega(a_o,s)),a_o)P_2(s)ds + \int_{S'} \epsilon/2\ P_2(s)ds.$$

By Assumption S.4, $\omega(a,s)$ is bounded for each $a \in A$, $s \in S$. This implies that $I$ and consequently $U_2$ are both bounded. Thus for each $\epsilon > 0$, there exists a $\gamma > 0$ such that

$$\int_{S-S'} U_2(I(\omega(a,s)),a)P_2(s)ds < \epsilon/2$$

whenever $\mu(s - s') < \gamma$

$$\therefore \int_S U_2(I(\omega(a,s)),a)P_2(s)ds$$

$$= \int_{S'} U_2(I(\omega(a,s)),a)P_2(s)ds$$

$$+ \int_{S-S'} U_2(I(\omega(a,s)),a)P_2(s)ds$$

$$< \int_S U_2(I(\omega(a_o,s)),a_o)P_2(s)ds + \varepsilon/2 + \int_S \varepsilon/2 \; P_2(s)ds$$

$$= \int_S U_2(I(\omega(a_o,s)),a_o)P_2(s)ds + \varepsilon. \quad \text{QED}$$

Assumption S.5 says that locally, when $a$ approaches a limit point, the increase in utility to the manager given a very small change of effort will not be large. This requirement seems reasonable, because with a small increase in the agent's effort he should not expect a large increase in the outcome as well as his compensation, or else he will be remunerated for very trivial efforts.

Given an incentive contract $I$, Theorems 2.2 (ii) and 2.6 guarantee the compliance of the requirements of Weierstrass' Theorem and hence the Nash criterion is satisfied. The problem is now reduced to finding an incentive contract such that the principal's expected utility function achieves a maximum.

It is easy to verify that space of incentive functions satisfies the properties of a vector space. Since any linear combination of two contracts is also a contract, i.e., for every $I_1$ and $I_2 \in C$, $\alpha I_1 + (1-\alpha)I_2 \in C$, $0 \leqq \alpha \leqq 1$, $C$ is locally convex. Formally, the above idea can be stated in the following theorem.

<u>Theorem 2.6</u>: The space of incentive functions C is a locally convex vector space.

Assumption S.6: C is complete.

Hence, by the above construction, C is a locally convex Banach space.

Now, let us digress for a moment and investigate the behavior of the utility function of the principal under Assumption S.3 (ii).

<u>Theorem 2.7</u>: Let $U: R \to R$ be a concave mapping on $(a,b)$. Then U is continuous on $(a,b)$.

Proof: The following proof is due to Rudin [1974]. Suppose $a < s < x < y < t < b$. Write S for the point $(s,U(s))$ in the plane and do the same thing with S,Y, and T. Then X is on or above the line SY; hence Y is on or below the line through S and X; also, Y is on or above XT. As $Y \to X$, it follows that $Y \to X$, i.e., $U(Y) \to U(X)$. Left-hand limits can be obtained in the same manner. Then continuity of U follows. QED

The above theorem establishes the continuity of the utility function on its domain. The integral of a continuous function is also continuous.

<u>Theorem 2.8</u>: For a given $a \in A$, $\omega_1(I,a) = \int_S U_1(\omega(a,s) - I(\omega(a,s)))P_1(s)ds$ is continuous on C.

If $C$ is of finite dimension, then by Theorem 2 (ii), a closed subset of $C$ is compact. However, if $C$ is of infinite dimension the compactness of $\{I\}$ is not easy to guarantee. This leads to the following.

Definition: A subset $X$ of a normed space is called totally bounded if and only if, for each $\varepsilon > 0$, there is a finite set of open balls which form an $\varepsilon$-cover of $X$.

Theorem 2.9: If a closed subset $I$ of a complete normed space $C$ is totally bounded, then it is sequentially compact.

Proof: Let $\{I_n\}$ be any infinite sequence in $I$. As $I$ is totally bounded, there is a finite 1-cover of $I$, say $\{I(f_1,\frac{1}{2}),\ldots,I(f_k,\frac{1}{2})\}$ where $f_j \in I$, $j = 1,\ldots,k$. At least one of these balls contains an infinite subsequence, $\{I_{n,1}\}$ say, of $\{I_n\}$. Take next a finite $\frac{1}{2}$-cover of $I$ and as before extract an infinite subsequence, $\{I_{n,2}\}$ say, of $\{I_{n,1}\}$ contained in one of these balls. Proceed in this way to find for each $m$ an infinite subsequence $\{I_{n,m}\}$ say, of $\{I_{n,m-1}\}$ such that $\{I_{n,m}\}$ is contained in a ball of diameter $m^{-1}$. Now, consider the diagonal sequence $\{I_{n,n}\}$. Then $\{I_{j,j}\}_{j=n}^{\infty}$ is a subsequence of $\{I_{n,n}\}_{j=n}^{\infty}$, and so is contained in a ball of diameter $n^{-1}$. Therefore,

$$|I_{n,n} - I_{m,m}| \leqq 1/\mathrm{Min}(n,m).$$

Thus $\{I_{n,n}\}$ is Cauchy, and hence is convergent in $I$ as $C$ is complete and $I$ is closed.

This shows that $\{I_n\}$ has a convergent subsequence, and proves sequential compactness. QED

In a Banach space, compactness and sequential compactness are equivalent. The requirements of the Weierstrass Theorem are satisfied to guarantee the existence of an optimal incentive contract such that the expected utility of the principal is maximized. The following gives a useful example of the class of totally bounded functions.

Definition: Suppose $\Omega$ is a subset of $R^n$, $C(\Omega)$ is the sup-normed Banach space of all continuous real-valued functions on $\Omega$. A set $C \subset C(\Omega)$ is said to be equicontinuous if and only if, for each $\epsilon > 0$, there exists a $\delta > 0$ such that $|I(\omega_1) - I(\omega_2)| < \epsilon$ for all $\omega_1, \omega_2 \in \Omega$ with $|\omega_1 - \omega_2| < \delta$ and for all $I \in C$.

Notice that for a given $\epsilon$, the same $\delta$ can be chosen for every I in $\mathcal{J}$. Equicontinuity means roughly that the degree of continuity is independent both of the position in the set $\Omega$ and of the functions in the space C.

<u>Theorem 2.10</u> (Ascoli): Suppose $\Omega$ is a bounded subset of $R^n$, $C(\Omega)$ is the sup-normed Banach space

of all continuous real-valued functions on $\Omega$, and $C \subset C(\Omega)$ is pointwise bounded and equicontinuous. Then $C$ is totally bounded in $C(\Omega)$.

Proof: Refer to Rudin [1973].

## 2.3* Monotonic Incentives

Definition: $I(\Omega)$ is monotonic increasing if $I(\omega_1) \leqq I(\omega_2)$ whenever $\omega_1 < \omega_2$ for all $\omega_1, \omega_2 \in \Omega$.

In this section, the class of incentive contracts which are monotonic increasing as defined above is considered. Recall that the principal's objective is to find an incentive function such that his expected utility is maximized. As the principal obtains the residual of the outcome after he pays the agent and his utility is increasing with wealth, given any wealth position from the outcome, he would choose the incentive function that would pay the manager the least amount. That is, in the principal's viewpoint, the optimal contract can be defined as $I_n^* = \inf\{I_1, \ldots, I_n\}$ for each $\omega \in \Omega$, $I_n \in C$, and $n$ denotes the possible contracts given $\omega$. It is clear that $I_n^*$ is a decreasing sequence. As the problem is formulated, the agent must maintain a minimum security level in terms of expected utility to remain in the company.

Thus, the incentive function should be bounded below at least pointwise. More explicitly, there exists a $I_o$ such that $\inf\{I_n^*\} \geqq I_o$ for every $\omega \in \Omega$.

Theorem 2.11 (Egoroff): Let the measure of a set $X$ be $\mu(X) < \infty$. If $\{f_n\}$ is a sequence of measurable functions which converges pointwise at every point of $x$, and if $\epsilon > 0$, there is a measurable set $E \subset X$, with $\mu(x - E) < \epsilon$, such that $\{f_n\}$ converges uniformly on $E$.

Proof: Refer to Rudin [1973].

Theorem 2.12: Suppose $I \subset C$, where $C$ is the class of monotonic increasing functions, each $I$ is pointwise bounded below for each $x \in \Omega$, and $\Omega$ is a subset of $R^n$. Let $I_n^* = \inf\{I_1, \ldots, I_n\}$ where $\{I_n\} \subset C$. Then $I_n^* \to I^*$ and the limit $I^* \in C$.

Proof: Let $\omega_1 < \omega_2$ and $\omega_1, \omega_2 \in \Omega$. Consider a subsequence $\{I_{i,k}\}$ of $\{I_i\}$, where $1 \leqq i \leqq n$ and $k = 1, 2, \cdots$. Let

$$I_n^*(\omega_1) = I_{i,j}(\omega_2)$$

and

$$I_n^*(\omega_2) = I_{i,m}(\omega_2).$$

Then

$$I_n^*(\omega_1) = I_{i,j}(\omega_1) \leqq I_{i,m}(\omega_1) \leqq I_{i,m}(\omega_2) = I_n^*(\omega_2).$$

Thus, $I_n^*$ is a monotonic increasing function of $\omega$ for each $n$. Clearly, $I_1^*(\omega) \geqq \cdots \geqq I_o(\omega)$ for every $\omega \in \Omega$. Therefore, $I_n^*(\omega) \to I^*(\omega)$ pointwise for each $\omega \in \Omega$. By Theorem 2.11, $I_n^* \to I^*$. What remains to be shown is that $I^* \in C$. Let $\epsilon > 0$ be arbitrary. Let $\omega_1$ and $\omega_2$ be chosen as above. There exists an $n$ such that

$$I_n^*(\omega_2) < I^*(\omega_2) + \epsilon.$$

Then

$$I^*(\omega_1) \leqq I_n^*(\omega_1) \leqq I_n^*(\omega_2) < I^*(\omega_2) + \epsilon.$$

This implies that $I^*(\omega_1) < I^*(\omega_2) + \epsilon$. Since $\epsilon$ is arbitrary, thus $I^* \in C$. QED

Theorem 2.12 provides a scheme to pick an optimal contract from a class of nondecreasing functions and it also establishes the existence of such a contract. If the principal were to enforce $I^*$, we still have to ensure that the agent will be induced to choose the correct action.

By Theorem 2.6, the upper semicontinuity of $\varphi_2$ on $A$ and the compactness of the set $A$ have guaranteed the existence of an optimal strategy for the agent given any incentive contract. Possibly there can be more than one optimal strategies for an incentive contract. Under such circumstances, it is

assumed that the agent will choose among the possible optimal strategies one which would maximize the principal's expected utility. Let $a^*$ denote such strategy. Thus, for $\{I_n^*\}$, there exists a corresponding sequence of optimal strategies $\{a_n^*\}$ by the agent. Clearly $\{a_n^*\} \subset A$.

Theorem 2.13: Suppose $I \subset C$ where $C$ is the class of monotonic increasing functions, each $I$ is pointwise bounded below for each $\omega \in \Omega$ and $\Omega$ is a subset of $R^n$. Let $I_n^* = \inf\{I_1,\ldots,I_n\}$ and $I_n^* \to I$ on $\Omega$. Suppose the set $A$ is bounded below. Then there exists an $a^*$ such that $\varphi_2(I^*,a^*)$ is maximized.

Proof: Since $\{a_n^*\} \subset A$ which is a closed and bounded subset of $R^n$, there exists a subsequence $\{a_{n_m}^*\}$ such that $a_{n_m}^* \to a^*$ where $a^* \in A$. We claim that $a^*$ is the optimal action choice. Let $\epsilon > 0$ be arbitrary. By the convergence of $\{I_n^*\}$ on $\Omega$,

$$\varphi_2(I_n^*,a_n^*) < \varphi_2(I^*,a_n^*) + \epsilon/2.$$

By Theorem 2.5 $\varphi_2$ is semicontinuous on $A$.

$$\varphi_2(I^*,a_n^*) \leqq \limsup_{n\to\infty} \varphi_2(I^*,a_n^*)$$

$$\leqq \varphi_2(I^*,a^*).$$

Suppose there exists an $a_o \in A$ such that

$$\varphi_2(I^*,a^*) < \varphi_2(I^*,a_o).$$

Then, by the convergence of $\{I_n^*\}$

$$\varphi_2(I_n^*,a_o) > \varphi_2(I^*,a_o) - \epsilon/2$$

$$> \varphi_2(I^*,a^*) - \epsilon/2$$

$$> \varphi_2(I^*,a_n^*) - \epsilon/2$$

$$> \varphi_2(I_n^*,a_n^*) - \epsilon.$$

Since $\epsilon$ is arbitrary, we have

$$\varphi_2(I_n^*,a_o) > \varphi_2(I_n^*,a_n^*).$$

But this is impossible, because $a_n^*$, by construction, is the action which maximizes $\varphi_2$ given $I_n^*$. Therefore, $a^*$ is the action which maximizes $\varphi_2$ given the optimal contract $I^*$. QED

# CHAPTER III

# MULTIPERIOD MODELS - PRELIMINARIES

## 3.1: Introduction

There are two major considerations in the construction of the multiperiod model: mathematical rigor and economic as well as behavior implications. Obviously, there is never a sharp distinction between the two, for they are not mutually exclusive. The thrust of discussions in this and subsequent chapters will be devoted to the development of a rigorous analytic model for the problem. No specific restrictions are imposed on the incentive contracts or its parameters. The main interest of the analysis is to investigate under what conditions would the principal be able to negotiate an optimal contract with the agent such that the expected utilities of the outcomes for both parties would be maximized over the long run. Most of the conditions which guarantee the existence of long run optimal contracts are imposed either on the net wealth return to the principal per period, the total expected discounted return to the principal, or the probability distributions of the outcomes. Conditions on the return functions, both per period and in total, are descriptions of the economic enrironments of the entity. For example, one of the conditions required in the infinite horizon model (Chapter V), is that the return to the principal per period is bounded and the discount factor has to be less than one. These are reasonable descriptions of common characteristics of an economic enterprise. Conditions on probability distributions are mainly behaviorial in nature. The type of "Bayesian" update as proposed in the imperfect state information model (Chapter VII) is similar to that imposed in a typical behaviorial research.

Multiperiod models have not been extensively dealt with in the agency literature. Authors of recent works attempted to extend their models to their multiperiod counterpart at the end of their respective works, for example, Baiman and Demski [1980]. Lambert [1981] views the multiperiod model in terms of a two-period agency and proposed a delayed payment scheme for the agent. Radner [1981] investigated the behavior of contracts in a repeated agency setting and concluded that if the process was repeated for long enough period of time, one can approach the first-best solution (a Pareto optimal contract) within some predetermined error bounds. His analysis imply an infinite horizon model.

This and subsequent chapters will attempt to formulate the multiperiod agency problem and obtain sets of conditions which guarantee the existence of optimal or nearly optimal contracts. The problem is first formulated in a typical decision choice model setting in terms of the most basic and rudimentary generic elements of the decision processes of the principal and the agent. Virtually no assumptions are imposed on the functions other than those which are required on the action choice set and the utility function of the agent to guarantee the compliance of the Nash incentive compatibility requirement. Any analysis cannot be carried too far under such a general formulation of the model. Dynamic programming algorithm is a procedure for searching solutions in dynamic problems and is well documented in both the operation research and stochastic optimal control literature. By reformulating the multiperiod agency problem under the dynamic programming hypothesis, I am able to adapt the recent findings of the above mentioned two bulks of researches to investigate the existence of long run optimal contracts.

The adaptation of the dynamic programming algorithm does not cause the model to lose any of its intended generality. The algorithm provides a structural and systematic procedure to search for an optimal or nearly optimal contract. It is through this systematic procedure that one can find the conditions under which an optimal contract can be found. This research is interested in the procedure of the search and the adaptation of such a procedure to search for a solution. It does not intend to go on to investigate the actual numerical and computational aspects of the algorithm.

The multiperiod agency problem is developed as a sequential decision process on a discrete time basis. Contracts are negotiated at the end of each period and there are no restrictions on the forms of the contract. There is no requirement that the contracts negotiated are to have the same form throughout the entire time span. The agent is free to leave the company any time within the horizon the model considers. When the agent leaves the company and a new one is hired, then the model will revert back to time 0 and the optimization process will restart all over again. The principal will attempt to maximize his total expected return over the time span and seek the corresponding optimal contracts to achieve his goal.

In the first part of the analysis, a most basic multiperiod agency model will be considered. Both the principal and the agent observe the payoff at the end of each period before the contract is negotiated. Payoffs in any particular period depend stochastically on the payoffs in the previous period, the incentive contract of the current period and the total accumulated wealth of the company. Such a transition function is assumed to be well-defined stochastically. Should there be any disturbance arising from the above description of the payoff transition, there always exists a probability distribution on the disturbance given the payoff and the incentive contract. In other words, the entire sequence of payoffs over the planning horizon can be stochastically defined at time 0 once an initial payoff $\omega_0$ is specified. The maximization process is then reduced to finding a sequence of contracts corresponding to the payoffs over all possible initial payoffs.

Considerations of the basic model will be given on finite horizon (Chapter IV) and infinite horizon (Chapter V) assumptions. The finite horizon model considered will be a long run model. Although a two-period agency is not excluded in the analysis, the focus is that of a much longer period, the economic life of the entity.

The development of this part of the analysis dwells heavily on writings in the optimal contral literature by Blackwell [1965a, 1965b, 1974, 1978], Denardo [1967], Strauch [1966], and particularly, Bertsekas and Shreve [1978].

The second part of the research centers on the imperfect state information model. Payoffs are observable only by the agent who reports a signal to the principal concerning the payoffs. Since the agent can choose any information signal he pleases, it becomes difficult if not impossible for the principal to assess the likelihood of the actual payoffs conditioned on the signal received. He can no longer project the sequence of expected payoffs such that he can optimize his expected utility as in the first part of the analysis. In the imperfect state model, necessary steps must be taken to guarantee the existence of a probability distribution on the discrepancy between the actual payoff and the signal reported (Chapters VI and VII). This is important because the principal is maximizing his total expected discounted wealth and there is no way that he can assess his expected wealth without some degree of control on the distribution of the disturbance which is the discrepancy in the current context.

Works in the literature concerning the imperfect state information models are sparse. The more notable ones are Åstrom [1965], Juskevic [1976], Sawaragi and Yoshikawa [1970], and Striebel [1975]. The proposed model will draw materials from all these papers and monographs freely, in particular that by Striebel whose work is also documented by Bertsekas and Shreve [1978].

### 3.2: The Basic Multiperiod Model

It was proved in the single period model that under very mild conditions on the agent's action choice set and his utility function, the Nash constraint of incentive compatibility is satisfied for all incentive contracts. These conditions can be trivially carried over to the multiperiod model. The condition on the action choice set is that it must be compact. In a multiperiod setting, the agent is choosing his sequence of optimal actions from the possible action set which is in fact the Cartesian or cross product of the sets $A_o, A_1, \ldots, A_{N-1}$. Each of the $A_i$, $i = 0, 1, \ldots, N-1$, is compact by assumption. The Cartesian product of compact sets is also a compact set. The multiperiod model is formulated in such a way that both the principal and the agent is maximizing their expected discounted utilities on the outcomes. The agent will discount all his expected remuneration over the time periods to time 0 and apply his utility preferences on the discounted compensation. The only utility function in question is the one at time 0 which is assumed to be upper semicontinuous with respect to the action choice set. With these two assumptions on the action choice set and the agent's utility function, the Weierstrass Maximum Theorem will guarantee that the agent will be able to select a sequence of optimal actions such that his expected utility is maximized.

In this and subsequent chapters, it is assumed that all $A_i$'s are closed subsets of $R^n$, and hence compact (Theorems 2.1 and 2.2(ii)), and the agent's utility function is upper semicontinuous with respect to the actions (Assumption S.5). Thus all contracts under consideration are incetive compatible. It is also assumed that if there are more than one optimal action choice for a particular contract, that is, the solution to the agent's problem does not have a unique solution, the agent will choose among the possible "optimal" solutions one which maximizes the principal's expected utility.

Now, given any contract, the agent is guaranteed that he can select an optimal action $a^*$ corresponding to the contract. He will execute $a^*$ with payoff $\omega(a^*,s)$. The payoff will effectively capture all the randomness of the state of nature. This will allow the following formulation of the multiperiod model to suppress the state parameter and treat the payoff as a random variable while assuming the existence of a probability distrubition on all possible payoff wealth positions.

The following definitions and conventions will be adopted:

(1) $\Omega$ and $C$ are two given nonempty spaces referred to as the payoff and incentive spaces respectively.

(2) For each $\omega \in \Omega$, there is given a nonempty subset $U(\omega)$ of $C$ referred to as the incentive constraint set at $\omega$. The incentive constraint set will exclude all contracts which give rise to an expected utility to the agent which is less than the outside labor market opportunity.

(3) Denote by $M$ the set of all functions $I : \Omega \to C$ such that $I(\omega) \in U(\omega)$ for every $\omega \in \Omega$. Denote by $\Pi$ the set of all sequences $\pi = (I_o, I_1, \ldots)$ such that $I_k \in M$ for every $k$. Elements of $\Pi$ are referred to as contracts. Elements of $\Pi$ of the form $\pi = (I, I, \ldots)$ where $I \in M$ are referred to as stationary contracts.

Recall that the principal's utility function is assumed to be monotonic increasing with respect to wealth which is its only argument. As his total expected return is maximized, his expected utility on the return will also be maximized simultaneously. To simplify the analysis that follow, it is assumed that the principal is risk neutral. All results in subsequent chapters will hold if the principal is risk adverse although the proofs to various theorems have to be modified to accommodate the risk adverseness.

Define the function

$$J_{N,\pi}(\omega_o) = E\{\sum_{k=0}^{N-1} \alpha^k g(\omega_k, I_k, y_k)\}$$

and

$$J_\pi(\omega_o) = \lim_{N\to\infty} J_{N,\pi}(\omega_o)$$

where $\alpha$ discount factor, a positive real number

$g$ the principal's net return function, $g : \Omega CY \to R^*$

$y$ a disturbance term. For each fixed $(\omega, I) \in \Omega C$, a probability $p(y|\omega, I)$ on the disturbance is given and $E\{\cdot|\omega, I\}$ denotes the expectation operator with respect to that probability.

$R$ is used to denote the real line and $R^*$ to denote the extended real line, that is, $R^* = R \cup \{-\infty, \infty\}$. The Cartesian product of sets $X_1, X_2, \ldots, X_n$ is denoted by $X_1X_2\ldots X_n$.

J, as defined, is then the expected total net return to the principal for N periods. It is sometimes called the N-stage net return. Depending on the values of $\alpha$, that is, whether $\alpha$ is greater or less than one, J represents the expected total future or discounted wealth to the principal respectively. The objective of the model is then to find an incentive contract which will maximize $J_N$.

To complete the model, the manner in which the payoff changes from period to period has to be specified. The following transition function is hypothesized.

$$\omega_{k+1} = f(\omega_k, I_k, J_k, y_k).$$

The payoff in period k then depends on the payoff in the previous period, the incentive contract which directs the agent's action in the period and the accumulated wealth of the company with a disturbance term.

The basic multiperiod problem can now be formulated as follows. For each $\omega_o \in \Omega$,

$$\underset{\pi \in \Pi}{\text{Maximize}} \quad J_{N,\pi}(\omega_o)$$

$$\text{Subject to} \quad \omega_{k+1} = f(\omega_k, I_k, J_k, y_k).$$

## 3.3: Optimality Concepts

For a fixed $\omega \in \Omega$, let $J_N^*(\omega)$ and $J^*(\omega)$ denote the optimal net return function to the principal in the finite and infinite horizon models respectively. Notationally,

$$J_N^*(\omega) = \sup_{\pi \in \Pi} J_{N,\pi}(\omega) \quad \text{for every} \quad \omega \in \Omega$$

$$J^*(\omega) = \sup_{\pi \in \Pi} J_{\pi}(\omega) \quad \text{for every} \quad \omega \in \Omega.$$

Regarding the optimal contracts, $\pi^*$ denotes the optimal contract in a sense that such a contract will give rise to $J_N^*$ or $J^*$. However, $\pi^*$ needs not always exist. A weaker form of optimality is also considered. A $\pi_\varepsilon$ contract is one which enables the principal to get an expected total net return of at least $J_N^* - \epsilon$ or $J^* - \epsilon$, where $\epsilon$ is a predetermined positive number. Such a contract is called a nearly optimal contract. By definition, the contract $\pi^*$ taken as a whole sequence of incentive function $\{I_1, I_2, \ldots\}$ will give $J_N^*$ or $J^*$ as the total expected net return to the principal for the N periods. A stronger form of optimality implies that at each stage, $k \leqq N-1$, $I_k$ can guarantee that the principal will obtain wealth position $J_k^*$. Such a contract is called uniformly N-stage optimal contract.

Sometimes, it may become necessary to approximate $J_N^*$ or $J^*$ by considering a sequence of contracts $\{\varepsilon_n\}$. In doing so, the rate of convergence to optimality or near optimality is always in important issue. Given a sequence of positive real numbers $\{\varepsilon_n\}$ converges to zero as $n$ tends to infinity, the sequence of contracts $\{\varepsilon_n\}$ is said to exhibit $\{\varepsilon_n\}$-dominated convergence to optimality if $J_{N,\pi_n}$ converges to $J_N^*$ and $J_{N,\pi_n} \geqq J_N^*(\omega) - \varepsilon_n$, for $n = 1, 2, \cdots$. Thus, if $\{\varepsilon_n\}$ converges to zero quickly, so does $J_{N,\pi_n}$ to $J_N^*$. The existence of an $\{\varepsilon_n\}$-dominated convergence sequence of contracts to an optimal contract will provide an iterative scheme such that an optimal or nearly optimal contract results.

Up to this point, the multiperiod basic agency model has been structured primarily on the usual decision-making economic model in its most general terms. Unfortunately, the solution to such a model is extremely difficult to arrive at or even to guarantee its existence. This is again, like the single period model, due to the fact that the decision variable for the principal, namely the incentive contract, is a function which belongs to an infinite dimension space. Unless one is willing to impose a number of restrictions on the behavior of the various functions, it appears impossible to solve the problem directly by using any known analytical skills. Because of the unique nature of the problem, a sequential decision process on a discrete time basis, it can be reformulated under a dynamic programming algorithm hypothesis. This will describe a procedure which, under appropriate conditions, guarantees an optimal or nearly optimal contract can be derived. With the various numerical methods and computational algorithms developed by the engineering discipline, more results can be generated on the optimal contract.

### 3.4: Dynamic Programming Algorithm

In the optimal control literature, multiperiod models or dynamics are analyzed under two separate branches of studies. The first one is that the functions under consideration changes continuously with respect to time, a continuous time model. A natural solution procedure for such a model is to formulate it as a system of differential equations. The second branch is that the system changes at discrete time intervals. A control function (the incentive function for agency model) is selected at each of these intervals to maximize or minimize the objective function. Dynamic programming algorithm is the common procedure used to search for solutions of this type of problem. In all appearance, the multiperiod agency model as it is applied to economic entities naturally belongs to the latter group of discrete time models. The following will give a discussion of the dynamic programming algorithm as a pausible solution procedure to the problem.

A dynamic programming algorithm is a systematic sequential search for a solution to the problem of either maximizing or minimizing an objective function in a dynamic setting. The search can go forward or backward in time. There are three operations performed repetitively. These operations will be discussed in greater detail in Chapter VI under the Borel models. The operations are as follows. First, there is the evaluation of a conditional expectation. Second, the objective function in two variables (payoff and incentive) is optimized over one of these variables (incentive). Finally, if an optimal or nearly optimal contract is to be constructed, a "selector" which maps each state to a contract which achieves or nearly achieves the maximum in the second step must be chosen.

There are two stages in the analysis of a problem under the dynamic programming algorithm. The first stage involved in the investigation of whether the three operations as outlined above are feasible under the assumptions of the problem. By feasibility, it would mean whether or not one can execute these operations. This is more of an existence phrase of the procedure than the implementation steps. However, there is a little more to be said than just mere existence. At this stage, one is concerned with not only the feasibility of the algorithm but also whether or not there are iterative schemes to perform the operations. The second stage is the actual finding of the solution procedure under the dynamic programming algorithm hypothesis to arrive at a solution. This includes the actual programming algorithm, the various numerical aspects of the computation and maybe the coding of the algorithm in terms of computer software.

In Chapter I, it was mentioned that the objective of this research is to build a rigorous analytic model in its most general terms to study the conditions such that the principal can design an optimal contract. Consistent with that objective, analysis in the subsequent chapters on the multiperiod model are focused on the first stage of the dynamic programming algorithm. Given the complexity of the problem, it should not be misled to believe that the second stage of actual implementation is easy. The techniques used will depend on the various behaviors of the functions of the model. If the functions behaves in a "nice" way, an analytic solution may be feasible. If they behave in some pattern, there are standard numerical methods which exist in the literature to solve the problem or to approximate the solution, for example, Marchuk [1982], Glowinski, Lions and Tremilieres [1981] and Dahlquist and Bjork [1974]. It is not the intention of this research to investigate and develop the numerical aspects of the solution to the problem.

Dynamic programming algorithm will be adopted to search for long run optimal or nearly optimal contracts in a multiperiod agency model. The model as proposed in the previous section will be reformulated under the dynamic programming algorithm hypothesis. Analysis will be conducted to investigate (i) whether the ad hoc assumptions imposed on the different models are met by the algorithm, (ii) whether a $J_N^*$ or $J^*$ can be achieved under the assumptions, (iii) if a $J_N^*$ or $J^*$ exists, then whether or not there are optimal or nearly optimal contracts that would give rise to the $J_N^*$ or $J^*$, and (iv) at each stage of the iteration, are there any numerical search procedures that can product the optimum.

To construct a dynamic programming model for the multiperiod agency problem, first consider the mapping $H$ defined by

$$H(\omega, I, J) = E\{g(\omega, I, J, y) + \alpha J(\omega, I, J, y)) | I, \omega\}$$

with the time subscript on the variables suppressed and the expectation operation is taken with respect to the probability distribution of $y$ given $I$ and $\omega$. Hence $H_k$ represents the total expected discounted net return to the principal from period $k$ through $N$ given an incentive contract and a payoff outcome. The dynamic programming algorithm will evaluate $H_k$ sequentially for $k = 0,1,\ldots,N-1$ and search for an $I_k^*$ that will maximize the corresponding $H_k$. Effectively, the algorithm is maximizing $H_k$ recursively $N$ times with respect to $I_k$ given $\omega$ and an initial value of $J, J_o$. In order to describe the operation for all $N$ periods together, define for each $I \in M$ and every $\omega \in \Omega$ the mappings $T_I$ and $T$ by

$$T_I(J)(\omega) = H(\omega, I, J)$$

$$T(J)(\omega) = \sup_{I \in U(\omega)} H(\omega, I, J)$$

Same interpretation is given to $T_I$ as $H$, whereas $T$ represents the solution of the dynamic programming algorithm per period: the total discounted expected return for $I_k^*$.

Let $T^k$, $k = 1,2,\ldots$ be the composition of $T$ with itself $k$ times and $T^o(J) = J$ for every possible $J$. $J_{N,\pi}(\omega)$ and $J_\pi(\omega)$, the total expected net return function can now be defined in the dynamic programming context as follows.

$$J_{N,\pi}(\omega) = (T_{I_o} T_{I_1} \cdots T_{I_{N-1}})(J_o)(\omega)$$

$$J_{\pi}(\omega) = \lim_{N\to\infty} (T_{I_o} T_{I_1} \cdots T_{I_{N-1}})(J_o)(\omega).$$

The solution under the dynamic programming algorithm can then be expressed as follows.

$$J_N(\omega) = T^N(J_o)(\omega)$$

$$J(\omega) = \lim_{N\to\infty} T^N(J_o)(\omega)$$

for every $\omega \in \Omega$ for the finite and infinite horizon models respectively.

The multiperiod model has been formulated using two different approaches. A model based on decision making process is first constructed. The same model is given in a dynamic programming structure. The natural question that arises is whether the two approaches are equivalent. In particular, one would be interested to know whether $J_N^* = J_N = T^N(J_o)$. In the following chapters, assumptions are hypothesized under each of the finite and infinite horizon models. It will be shown that the function $H$ satisfies the assumptions, and, under these assumptions, $J_N^* = T^N(J_o)$. This will establish the validity of using dynamic programming in generating solutions for the model. Next, the question of the existence of optimal or near optimal contracts will be investigated. Finally, the requirements to arrive at the optimal or near optimal contracts will be considered.

CHAPTER III*

MULTIPERIOD MODELS - PRELIMINARIES

## 3.1* Notations and Assumptions

The mathematical notations used in this work are fairly standard. R is used to denote the real line and R* to denote the extended real line, i.e., $R^* = R \cup \{-\infty, \infty\}$. It is assumed throughout that R is equipped with the usual topology generated by the open intervals (a,b), where $a,b \in R$, and with the Borel $\sigma$-algebra generated by this topology. Similarly R* is equipped with the topology generated by the open intervals (a,b), $a,b \in R$, together with the sets $(c,\infty]$, $[-\infty,c)$, $c \in R$, and with the $\sigma$-algebra generated by this topology. The Cartesian product of sets $X_1, X_2, \ldots, X_n$ is denoted by $X_1X_2\ldots X_n$.

It was proved in the single period model that the Nash constraint is satisfied for all incentive contract if the action space A is of finite dimension and the utility function of the agent $U_2$ is upper semicontinuous with respect to his action choice a. In this and subsequent parts of the analysis, it is assumed that the contracts under consideration are incentive compatible. Therefore, given any contract, the payoff function is well-defined with respect to the state space. This enables the following analysis to suppress the state parameter and treat the payoff as a random variable and assume the existence of a probability measure on all possible payoff wealth positions.

The following definitions and conventions will be adopted:

(1) $\Omega$ and $C$ are two given sets referred to as the payoff space and incentive space respectively.

(2) For each $\omega \in \Omega$, there is given a nonempty subset $U(\omega)$ of $C$ referred to as the incentive constraint set at $\omega$.

(3) Denote by $M$ the set of all functions $I : \Omega \to C$ such that $I(\omega) \in U(\omega)$ for every $\omega \in \Omega$. Denote by $\Pi$ the set of all sequences $\pi = (I_o, I_1, \ldots)$ such that $I_k \in M$ for every $k$. Elements of $\Pi$ are referred to as contracts. Elements of $\Pi$ of the form $\pi = (I, I, \ldots)$ where $I \in M$ are referred to as stationary contracts.

(4) Denote by

$F$ the set of all extended real-valued functions $J : \Omega \to R^*$. (The exact form of $J$ will be defined in the next section.)

B the Banach space of all bounded real-valued functions $J : \Omega \to R$ with the sup norm, i.e., $\|J\| = \sup_{\omega \in \Omega} |J(\omega)|$ for every $J \in B$.

(5) For all $J, J'$ write

$$J = J' \text{ if } J(\omega) = J'(\omega) \text{ for every } \omega \in \Omega$$

$$J \leq J' \text{ if } J(\omega) \leq J'(\omega) \text{ for every } \omega \in \Omega.$$

For all $J \in F$ and $\epsilon \in R$, denote $J + \epsilon$ the function taking the value $J(\omega) + \epsilon$ at each $\omega \in \Omega$, i.e., $(J + \epsilon)(\omega) = J(\omega) + \epsilon$ for every $\omega \in \Omega$.

(6) The usual arithmetic for $R^*$ is adopted except $\infty - \infty = -\infty + \infty = \infty$ and $0 \cdot \infty = 0$.

Since the principal's utility function is assumed to be monotonic increasing with respect to wealth which is its only argument, it is assumed for the rest of this analysis that the principal is risk neutral. Such an assumption will not cause any loss in generality because as the principal's total expected return is maximized, so is his expected utility on the return.

## 3.2* The Basic Multiperiod Model

The expected utility of the principal or the agent is defined as the expectation taken as an integral of their respective utility functions with respect to the probability measure on the state space. Such an integral may not always exist as a Reimann integral.

An outer integral is adopted as the expectation operator.

Definition: If $f \geq 0$, the outer integral of $f$ with respect to a probability measure $p$ is defined by

$$\int^{*} f dp = \inf \{\int g dp : f \leq g, \ g \text{ is Borel-measurable}\}.$$

If $f$ is arbitrary, define

$$\int^{*} f dp = \int^{*} f^{+} dp - \int^{*} f^{-} dp$$

where $f^{+} = \max\{0, f\}$, $f^{-} = \max\{0, -f\}$.

Define the total net return to the principal for $N$ periods as

$$J_{N,\pi}(\omega_0) = E^{*}\{\sum_{k=0}^{N-1} \alpha^{k} g(\omega_k, I_k, y_k)\}$$

where $\alpha$ discount factor, a positive real number

$g$ the principal's net return function $g : \Omega C Y \to R^{*}$

$y$ a disturbance which takes values in a measurable space $(Y, \mathcal{Y})$. For each fixed $(\omega, I) \in \Omega C$, a probability measure $p(dy|\omega, I)$ on $(Y, \mathcal{Y})$ is given and $E^{*}\{\cdot|\omega, I\}$ denotes the expectation in terms of the outer integral with respect to that measure.

To complete the model, the following transition function is hypothesized for the payoff function.

$$\omega_{k+1} = f(\omega_k, I_k, J_k, y_k).$$

The basic multiperiod problem can now be formulated as follows. For each $\omega_o \in \Omega$,

$$\underset{\pi \in \Pi}{\text{Maximize}}\ J_{N,\pi}(\omega_o)$$

$$\text{Subject to}\quad \omega_{k+1} = f(\omega_k, I_k, J_k, y_k).$$

## 3.3* Optimality Concepts

For a fixed $\omega \in \Omega$, let $J_N^*(\omega)$ denote the optimal net return functions to the principal in the finite and infinite horizon models respectively, that is,

$$J_N^*(\omega) = \sup_{\pi \in \Pi} J_{N,\pi}(\omega) \quad \text{for every} \quad \omega \in \Omega$$

$$J^*(\omega) = \sup_{\pi \in \Pi} J_\pi(\omega) \quad \text{for every} \quad \omega \in \Omega.$$

Regarding the contract, the following optimality concepts are adopted.

Definition: A contract $\pi^* \in \Pi$ is said to be N-stage optimal at $\omega \in \Omega$ if $J_{N,\pi^*}(\omega) = J_N^*(\omega)$ and optimal at $\omega \in \Omega$ if $J_{\pi^*}(\omega) = J^*(\omega)$.

A contract $\pi^* \in \Pi$ is said to be N-stage optimal if $J_{N,\pi^*} = J_N^*$ and optimal if $J_{\pi^*} = J^*$.

Definition: A contract $\pi^* = (I_o^*, I_1^*, \ldots)$ is called uniformly N-stage optimal if the contracts $(I_i^*, I_{i+1}^*, \ldots)$ is $(N - i)$-stage optimal for all $i = 0, 1, \ldots, N-1$.

Definition: Given $\epsilon > 0$, a contract $\pi_\epsilon \in \Pi$ is N-stage $\epsilon$-optimal if

$$J_{N,\pi_\epsilon}(\omega) \geq \begin{cases} J_N^*(\omega) - \epsilon & \text{if } J_N^*(\omega) < \infty \\ 1/\epsilon & \text{if } J_N^*(\omega) = \infty \end{cases}$$

The contract $\pi_\epsilon \in \Pi$ is said to be $\epsilon$-optimal if

$$J_{\pi_\epsilon}(\omega) \geq \begin{cases} J^*(\omega) - \epsilon & \text{if } J^*(\omega) < \infty \\ 1/\epsilon & \text{if } J^*(\omega) = \infty \end{cases}$$

Definition: If $\{\epsilon_n\}$ is a sequence of positive numbers with $\epsilon_n \to 0$, we say a sequence of contracts $\{\pi_\epsilon\}$ is said to exhibit $\{\epsilon_n\}$-dominated convergence to optimality if

$$\lim_{n\to\infty} J_{N,\pi_n} = J_N^*$$

and for $n = 2,3,\ldots$

$$J_{N,\pi_n}(\omega) \geq \begin{cases} J_N^*(\omega) - \epsilon_n & \text{if } J_N^*(\omega) < \infty \\ J_{N,\pi_{n-1}}(\omega) - \epsilon_n & \text{if } J_N^*(\omega) = \infty \end{cases}$$

## 3.4* The Dynamic Programming Algorithm

The discrete time sequential decision process provides a natural framework for adopting the dynamic programming algorithm in generating an optimal contract or nearly optimal contracts. This section gives a definition of $J_{N,\pi}(\omega)$ and $J_\pi(\omega)$ in a dynamic programming framework.

Consider the mapping $H : \Omega CF \to R^*$ defined by

$$H(\omega, I, J) = E^*\{g(\omega, I, J, y) + \alpha J(f(\omega, I, J, y)) \mid I, \omega\}.$$

$H_k$ represents the total expected discounted net return to the principal from period $k$ through $N-1$ given an incentive contract and a payoff outcome. To describe the operations for all $N$ periods together, define for each $I \in M$ the mappings $T_I : F \to F$ by

$$T_I(J)(\omega) = H(\omega, I, J) \quad \text{for every} \quad \omega \in \Omega$$

and $T : F \to F$ by

$$T(J)(\omega) = \sup_{I \in U(\omega)} H(\omega, I, J) \quad \text{for every} \quad \omega \in \Omega.$$

Let $T^k$, $k = 1, 2, \ldots$ be composition of $T$ with itself $k$ times and $T^o(J) = J$ for every $J \in F$. $J_{N,\pi}(\omega)$ and $J_\pi(\omega)$ can now be defined in the dynamic programming context.

$$J_{N,\pi}(\omega) = (T_{I_o} T_{I_1} \cdots T_{I_{N-1}})(J_o)(\omega)$$

$$J_\pi(\omega) = \lim_{N \to \infty} (T_{I_o} T_{I_1} \cdots T_{I_{N-1}})(J_o)(\omega)$$

The solution can be expressed as

$$J_N(\omega) = T^N(J_o)(\omega)$$

$$J(\omega) = \lim_{N \to \infty} T^N(J_o)(\omega)$$

for every $\omega \in \Omega$ for the finite and infinite horizon models respectively.

# CHAPTER IV

# FINITE HORIZON MODEL

## 4.1: Introduction and Assumptions

The finite horizon model describes the long term planning behavior of an entity. It assumes the fact that the entity will dissolve at the end of $N$ periods. The model seeks optimal contracts such that the total expected discounted net return to the owner(s) over the $N$ periods is maximized. Sole proprietorship and partnership are the two prime examples of entities which dissolve in a finite period of time.

In considering the N-stage optimization problem, as indicated towards the end of the last chapter, the central question is whether $J_N^* = T^N(J_o)$ in which the optimal $J_N^*$ is obtained by successively computing $T(J_o), T^2(J_o), \ldots$ via the dynamic programming algorithm. Another issue is the existence of optimal or nearly optimal contracts. In order to respond to the above questions affirmatively, some assumptions have to be imposed on the function H which describes the transition of the total net return to the principal from one period to another.

The first assumption concerns the behavior of H in response to small changes in the value of J. For every payoff wealth positions and every possible contracts within the constraint set, given a small change in J, the corresponding change in H is also small, in fact within some predetermined bounds. This is a weak statement to say that H is continuous with respect to J.

The second assumption imposed a kind of linear behavior on H from below with respect to J. It says that if J is reduced by some positive value, say r, then the value of H is reduced by no more than r times the discount factor $\alpha$. This assumption restricts the behavior of H when a large change in J occurs.

The third assumption is admittedly somewhat more complicated. It allows one to get stronger results on the existence of nearly optimal contracts than can be obtained under assumption two. It says that if one were to approximate the entity's total net return function J by a sequence of return functions, $\{J_n\}$ then each $J_n$ can in turn be approximated by a sequence of contracts within some predetermined bounds.

The first two assumptions seem reasonable in most economic environments in which business entities operate. There should be no drastic changes in the total net return (or total wealth) of a company in any period given any changes in the total wealth of the previous period. The third assumption simply implies that there always exists some computational procedures to approximate the return function to a desired accuracy. Although the current work does not include the analysis of the actual numerical methods to be used to solve the problem, I will attempt at least to guarantee that some computation procedures exist and they are implementable under the various assumptions studied.

## 4.2: The Finite Horizon Model

The first step in the analysis of the finite horizon model is to establish the validity of the assumptions in the model, that is, whether the function $H$ as defined in the dynamic programming formulation of the problem satisfies the assumptions. If one were to examine the assumptions one and two carefully, one would notice that assumption one is really a special case of assumption two. Since the real number in assumption two can be any positive number, it can be made arbitrary small which is exactly what assumption one requires. All that is needed to be shown is that $H$ satisfies assumptions two and three (Theorem 4.1) and extend the result to conclude that assumption one is also satisfied.

The fact that $J_N^* = T^N(J_o)$ can be shown to be true if $J_{k,\pi}(\omega) > -\infty$ under assumptions one or three. This is also the case if $J_N^* < \infty$ under assumption two (Theorems 4.2 and 4.3). $J_{k,\pi}(\omega)$ being negatively finite is trivially satisfied in any economic situations. No company can survive infinite losses for any one period, not to mention that $J_k$ is the total net return for $k$ periods. The same situation applies for $J_N^*$ being finite. It is unrealistic to have $J_N^*$ infinite in any common business environment. With the above facts established, one can say that dynamic programming algorithm is appropriate as a solution procedure for the finite horizon models.

Next, the question of whether an optimal or nearly optimal contract would exist under the assumptions is investigated. Under assumption three, it is shown that there always exists an $\{\epsilon_n\}$-dominated convergence to optimality sequence of contracts (Theorem 4.3). Recall that assumption three says that if $J$ is to be approximated by a sequence of functions $\{J_n\}$ then each $J_n$ can in turn be approximated by a sequence of contracts within some predetermined bounds. Theorem 4.3 guarantees that such sequence of contracts is an $\{\epsilon_n\}$-dominated convergence to optimality contracts. This implies that assumption three not only guarantees the existence of a nearly optimal contract, it ensures that one can find or adopt an iteration procedure to approximate the nearly optimal contract.

In the development of the model, there are no restrictions placed on the individual functions and their parameters with the exception of $U_2$, the agent's utility function being continuous from above with respect to the action choice to guarantee the compliance of the Nash constraint and H under the above three assumptions. The model is drawn in its maximum generality. A trade-off the researcher has always had to make under such general setting is the inability to make specific characterization of the optimal or nearly optimal contracts. Such comments can only be made if more information about the behavior of individual functions are known, that is, more assumptions and restrictions on the functions are needed. The only characterization of the optimal contracts, under the current general formulation, that can be made is that how the contracts are related to each stage in the iteration procedures of the dynamic programming algorithm or how they behave each period in the process of arriving at the optimal. This may sound disturbing to the application-oriented reader. This research is carried on in the interplay of the economic model and the dynamic programming model. The correspondence in the two models, once established, will allow the dynamic programming algorithm, which is documented in both the operation research and optimal control literatures, to take care of the various numerical and computational aspects of the actual search for solution to the model.

An uniformly N-stage optimal contract is always the more desirable contract because it maximizes the total net return at any given point of time over the planning horizon rather than just at the end of N periods. Such a contract can be shown to exist if at each period and for each payoff outcome, there is a contract that would maximize $H_k$, that is the supremum is attained in the relation $T^{k+1}(J_o)(\omega) = \sup_{I \in U(\omega)} H[\omega, I, T^k(J_o)]$. And if a uniformly N-stage optimal contract exist, $J_N^* = T^N(J_o)$ (Theorem 4.4 and its corollaries).

In the dynamic programming algorithm, each $T^k(J_o)$ is computed sequentially over all possible contracts. Under the structure of the model proposed, this is equivalent to saying that a sequential decision process is adopted and the total wealth return is maximized at each stage of the process. This entails the question of whether an optimal can be achieved each period. If the incentive constraint set is compact, which is a typical requirement for optimization problems, an optimal contract can always be achieved (Theorem 4.5). Together with the results of Theorem 4.4, it can be concluded that under the appropriate assumptions of the finite horizon model, an uniformly N-stage optimal contract is guaranteed to exist under the dynamic programming algorithm.

## CHAPTER IV*

## FINITE HORIZON MODELS

### 4.1* Introduction and Assumptions

In considering the N-stage optimization problem, the central question is whether $J_N^* = \sup_{\pi \in \Pi} J_{N,\pi} = T^N(J_o)$ such that the optimal $J_N^*$ can be obtained by successively computing $T(J_o), T^2(J_o), \ldots$ via the dynamic programming algorithm. Another issue is the existence of optimal or nearly optimal contracts. In order to respond to the above questions affirmatively, some assumptions have to be imposed on the function H.

Assumption F.1: If $\{J_k\} \subset F$ is a sequence satisfying $J_{k+1} \geq J_k$ for all $k$ and $H(\omega, I, J_1) > -\infty$ for all $\omega \in \Omega$, $I \in U(\omega)$, then

$$\lim_{k\to\infty} H(\omega, I, J_k) = H(\omega, I, \lim_{k\to\infty} J_k) \quad \omega \in \Omega,\ I \in U(\omega).$$

Assumption F.2: There exists a scalar $\alpha \in (0, \infty)$ such that for all scalars $r \in (0, \infty)$ and functions $J \in F$, we have

$$H(\omega, I, J) \geq H(\omega, I, J - r) \geq H(\omega, I, J) - \alpha r \quad \omega \in \Omega,\ I \in U(\omega).$$

Assumption F.3: There is a scalar $\beta \in (0,\infty)$ such that if $J \in F$, $\{J_n\} \subset F$, and $\{\epsilon_n\} \subset R$ satisfy

$$\sum_{n=1}^{\infty} \epsilon_n < \infty \qquad \begin{cases} \epsilon_n > 0 & n = 1,2,\ldots \\ J \geqq J_n & n = 1,2,\ldots \end{cases}$$

$$J = \lim_{n\to\infty} J_n$$

$$J_n(\omega) \geqq \begin{cases} J(\omega) - \epsilon_n & n = 1,2,\ldots \quad \omega \in \Omega \quad J(\omega) < \infty \\ J_{n-1}(\omega) - \epsilon_n & n = 1,2,\ldots \quad \omega \in \Omega \quad J(\omega) = \infty \end{cases}$$

$H(\omega, I, J_1) > -\infty$ for every $\omega \in \Omega$, $I \in U(\omega)$

then there exists a sequence $\{I_n\} \subset M$ such that

$$\lim_{n\to\infty} T_{I_n}(J_n) = T(J)$$

and

$$T_{I_n}(J_n)(\omega) \geqq \begin{cases} T(J)(\omega) - \beta\epsilon_n & n = 1,2,\ldots \quad \omega \in \Omega, T(J)(\omega) < \infty \\ T_{I_{n-1}}(J_{n-1})(\omega) - \beta\epsilon_n & n = 1,2,\ldots \quad \omega \in \Omega, T(J)(\omega) = \infty \end{cases}$$

Actually, Assumption F.1 is a special case of F.2. Careful examination of the two assumptions will show that if Assumption F.2 is met, F.1 will be satisfied. Assumption F.3 is somewhat more complicated. It allows one to get stronger results on the existence of nearly optimal contracts than can be obtained under F.2 (Theorem 4.3).

## 4.2* The Finite Horizon Model

The first step in the development of the model is to show that the function H as defined in the previous chapter satisfies the three assumptions on the finite horizon model. The proof of such a statement requires four technical lemmas on outer integrals and probability measures. These results are very standard in mathematical analysis and probability theory. For completeness, they are stated without proof as follows.

Lemma 4.1: If $\epsilon > 0$ and $f \leqq g \leqq f + \epsilon$, then

$$\int^* f dp \leqq \int^* g dp \leqq \int^* f dp + 2\epsilon.$$

Lemma 4.2: Let $(X, \mathcal{B}, p)$ be a probability space. Let $\{\epsilon_n\}$ be a sequence of positive numbers with $\sum_{n=1}^{\infty} \epsilon_n < \infty$. Let $\{f_n\}$ be a sequence with $\lim_{n\to\infty} f_n = f$ pointwise $f \geqq f_n$ for $n = 1, 2, \cdots$. Let

$$f_n(x) \geqq \begin{cases} f(x) - \epsilon_n & \text{if } f(x) < \infty \\ f_{n-1}(x) - \epsilon_n & \text{if } f(x) = \infty \end{cases}$$

and $\int^* f_1 dp < \infty$. Then $\lim_{n\to\infty} \int^* f_n dp = \int^* f dp$.

Lemma 4.3: Let $(X, \mathcal{B}, p)$ be a probability space. If $p^*(\{x: f(x) = \infty\}) > 0$, then for every $g$, $g: X \to [-\infty, \infty]$ and $\mathcal{B}$-measurable, $\int^* (g + f) dp = \infty$.

<u>Lemma 4.4</u>: Let $(X,\mathcal{B},p)$ be a probability space. If $E \subset X$ satisfies $p^*(E) = 0$, then for any $f: X \to R^*$

$$\int^* f dp = \int^* \chi_{X-E} f dp.$$

<u>Theorem 4.1</u>: The mapping

$$H(\omega,I,J) = E^*\{g(\omega,I,J,y) + \alpha J[f(\omega,I,J,y)] \mid \omega,I\}$$

satisfies Assumptions F.2 and F.3.

Proof: For $r \in (0,\infty)$

$$H(\omega,I,J) \geq H(\omega,I,J-r).$$

By Lemma 4.1,

$$\begin{aligned} H(\omega,I,J) &= \int^* (g(\omega,I,J,y) + \alpha J[f(\omega,I,J,y)])p(dy \mid \omega,I) \\ &\geq H(\omega,I,J-r) \\ &\geq H(\omega,I,J) - 2\alpha r. \end{aligned}$$

Thus F.2 is satisfied. Let $J \in F$, $\{J_n\} \subset F$, $\{\epsilon_n\} \subset R$ satisfy $\sum_{n=1}^{\infty} \epsilon_n < \infty$, $\epsilon_n > 0$ and for all $n$

$$J = \lim_{n\to\infty} J_n \qquad J \geq J_n$$

$$J_n(\omega) \geq \begin{cases} J(\omega) - \epsilon_n & \text{if } J(\omega) < \infty \\ J_{n-1}(\omega) - \epsilon_n & \text{if } J(\omega) = \infty \end{cases}$$

$$H(\omega,I,J_1) > -\infty \quad \forall\ \omega \in \Omega,\ I \in U(\omega).$$

Let $\{\bar{I}_n\} \subset M$ be such that for all $n$

$$T_{\bar{I}_n}(J)(\omega) \geqq \begin{cases} T(J)(\omega) - \epsilon_n & \text{if } T(J)(\omega) < \infty \\ 1/\epsilon_n & \text{if } T(J)(\omega) = \infty \end{cases}$$

$$T_{\bar{I}_n}(J) \geqq T_{\bar{I}_{n-1}}(J).$$

Consider the set

$$A(J) = \{\omega \in \Omega \mid \exists\ I \in U(\omega) \text{ with } p^*(\{y \mid J[f(\omega,I,J,y)] = \infty\} \mid \omega, I) > 0\}$$

where $p^*$ denotes p-outer measure. Let $\bar{I} \in M$ be such that

$$p^*(\{y \mid J[f(\omega,\bar{I}(\omega),J,y)] = \infty\} \mid \omega, \bar{I}(\omega)) > 0 \quad \forall\ \omega \in A(J).$$

Define for all $n$

$$I_n(\omega) = \begin{cases} \bar{I}(\omega) & \text{if } \omega \in A(J) \\ \bar{I}_n(\omega) & \text{if } \omega \notin A(J) \end{cases}$$

Claim that $\{I_n\}$ thus defined satisfies the requirement of F.3 with $\beta = 1 + 2\alpha$. For $\omega \in A(J)$, by Lemma 4.2

$$\liminf_{n\to\infty} T_{\bar{I}}(J_n)(\omega)$$

$$= \liminf_{n\to\infty} \int^* \{g[\omega,\bar{I}(\omega),J,y] + \alpha J_n[f(\omega,\bar{I}(\omega),J,y)]\} p(dy \mid \omega, \bar{I}(\omega))$$

$$= \int^* \{g[\omega,\bar{I}(\omega),J,y] + \alpha J[f(\omega,\bar{I}(\omega),J,y)]\} p(dy \mid \omega, \bar{I}(w)).$$

Since $T_{\bar{I}}(J)(\omega) > -\infty$, by Lemma 4.3

$$\liminf_{n\to\infty} T_{\bar{I}_n}(J_n)(\omega) = \infty \geqq T(J)(\omega).$$

For $\omega \notin A(J)$, for all $n$

$$p^*(\{y|J[f(\omega,I_n(\omega),J,y)] = \infty\}|\omega,I_n(\omega)) = 0.$$

Let $B_n \in \mathcal{Y}$ $\{y|J[f(\omega,I_n(\omega),J,y)[ = \infty\} \subset B_n$ and

$$p(B_n|\omega,I_n(\omega)) = 0 \quad \forall\, n.$$

By Lemmas 4.1 and 4.4

$$\begin{aligned} &T_{I_n}(J_n)(\omega) \\ &= \int^* \chi_{\mathcal{Y}-B_n}(\omega)\{g[\omega,I_n(\omega),J,y] + \alpha J_n[f(\omega,I_n(\omega),J,y)\}p(dy|\omega,I_n(\omega)) \\ &\geqq \int^* \chi_{\mathcal{Y}-B_n}(\omega)\{g[\omega,I_n(\omega),J,y] + \alpha J[f(\omega,I_n(\omega),J,y)]\}p(dy|\omega,I_n(\omega)) \\ &\quad - 2\alpha\epsilon_n \\ &= T_{I_n}(J)(\omega) - 2\alpha\varepsilon_n \end{aligned}$$

$$\therefore \quad \forall\, \omega \notin A(J), \ \liminf_{n\to\infty} T_{I_n}(J_n)(\omega) \geqq \liminf_{n\to\infty} T_{I_n}(J)(\omega)$$

$$= T(J)(\omega)$$

$$\therefore \quad \liminf_{n\to\infty} T_{I_n}(J_n)(\omega) > T(J)(\omega) \quad \forall\, \omega \in \Omega.$$

But $T_{I_n}(J_n) \leqq T(J)$ $\forall\, n = 1,2,...$ by hypothesis

$$\therefore \quad \lim_{n\to\infty} T_{I_n}(J_n) = T(J).$$

If $\omega$ is such that $T(J)(\omega) < \infty$, then $\omega \notin A(J)$. Thus, by Lemma 4.1,

$$\begin{aligned} T_{I_n}(J_n)(\omega) &\geqq T_{I_n}(J)(\omega) - 2\alpha\epsilon_n \\ &\geqq T(J)(\omega) - (1 + 2\alpha)\epsilon_n. \end{aligned}$$

If $\omega$ is such that $T(J)(\omega) = \infty$, then either

(i) $\omega \notin A(J)$ or (ii) $\omega \in A(J)$.

(i) Let $\omega \notin A(J)$

$$\begin{aligned} T_{I_n}(J_n)(\omega) &\geqq T_{I_n}(J)(\omega) - 2\alpha\varepsilon_n \\ &\geqq T_{I_{n-1}}(J)(\omega) - 2\alpha\varepsilon_n \\ &\geqq T_{I_{n-1}}(J_{n-1})(\omega) - 2\alpha\varepsilon_n . \end{aligned}$$

(ii) Let $\omega \in A(J)$

$$\begin{aligned} T_{I_n}(J_n)(\omega) &= \int^{*} \{g[\omega,\bar{I}(\omega),J,y] + \alpha J_n[f(\omega,\bar{I}(\omega),J,y)]\} \\ &\qquad p(dy|\omega,\bar{I}(\omega)) \\ &\geqq \int^{*} \{g[\omega,\bar{I}(\omega),J,y] + \alpha J_{n-1}[f(\omega,\bar{I}(\omega),J,y)]\} \\ &\qquad p(dy|\omega,\bar{I}(\omega)) - 2\alpha\varepsilon_n \\ &= T_{I_{n-1}}(J_{n-1})(\omega) - 2\alpha\varepsilon_n \end{aligned}$$

$\therefore$ The claim is proved by letting $\beta = 1 + 2\alpha$. QED

Next, the correspondence of the solution values between the economic model and the dynamic programming model, that is, $J_N^* = T^N(J_o)$, is established under the assumptions. Theorems 4.2 and 4.3 also show the existence of a nearly optimal contract under Assumptions F.2 and F.3.

Theorem 4.2: (a) Let F.1 hold. Let $J_{k,\pi}(\omega) > -\infty$ for all $\omega \in \Omega$, $\pi \in \Pi$, and $k = 1,2,\ldots,N$. Then

$$J_N^* = T^N(J_o).$$

(b) Let F.2 hold. Let $J_k^*(\omega) < \infty$ for all $\omega \in \Omega$ and $k = 1,2,\ldots,N$. Then

$$J_N^* = T^N(J_o)$$

and for every $\varepsilon > 0$, there exists an N-stage $\epsilon$-optimal contract, i.e., a $\pi_\epsilon \in \Pi$ such that

$$J_N^* \geqq J_{N,\pi_\epsilon} \geqq J_N^* - \varepsilon.$$

Proof: (a) For each $k = 0,1,\ldots,N-1$, consider a sequence $\{I_k^i\} \subset M$ such that

$$\lim_{i\to\infty} T_{I_k^i}[T^{N-k-1}(J_o)] = T^{N-k}(J_o) \qquad k = 0,1,\ldots,N-1$$

$$T_{I_k^i}[T^{N-k-1}(J_o)] \leqq T_{I_k^{i+1}}[T^{N-k-1}(J_o)] \quad \begin{array}{l} k = 0,1,\ldots,N-1 \\ i = 0,1,\ldots\ . \end{array}$$

By F.1 and $J_{k,\pi}(\omega) > -\infty$, we have

$$\begin{aligned} J_N^* &\geqq \sup_{i_o}\ldots\sup_{i_{N-1}} \left(T_{I_o^{i_o}} \cdots T_{I_{N-1}^{i_{N-1}}}\right)(J_o) \\ &= \sup_{i_o}\ldots\sup_{i_{N-2}} \left(T_{I_o^{i_o}} \cdots T_{I_{N-2}^{i_{N-2}}}\right)\left(\sup_{i_{N-1}} T_{I_{N-1}^{i_{N-1}}}\right)(J_o) \\ &= \sup_{i_o}\ldots\sup_{i_{N-2}} \left(T_{I_o^{i_o}} \cdots T_{I_{N-2}^{i_{N-2}}}\right)[T(J_o)] \\ &= T^N(J_o). \end{aligned}$$

Since $J_N^* = \sup_{\pi\in\Pi} J_{N,\pi}(\omega) \quad \forall\ \omega \in \Omega$

$$= \sup_{\pi\in\Pi} (T_{I_o} T_{I_1} \cdots T_{I_{N-1}})(J_o)$$

$$\leqq T^N(J_o)$$

$$\therefore\ J_N^* = T^N(J_o)$$

(b) The result clearly holds for $N = 1$. Suppose it holds for $N = k$. Then $J_k^* = T^k(J_o)$ and for a $\epsilon > 0\ \ \exists\ \pi_\epsilon \in \Pi$ such that $J_{k,\pi_\epsilon} \geqq J_k^* - \epsilon$. By F.2, $\forall\ I \in M$

$$J_{k+1}^* \geqq T_I(J_{k,\pi_\epsilon}) \geqq T_I(J_k^*) - \alpha\epsilon$$

$$\therefore\ J_{k+1}^* \geqq T(J_k^*).$$

By induction, $J_{k+1}^* \geqq T^{k+1}(J_o)$. By definition $T^{k+1}(J_o) \geqq J_{k+1}^*$

$$\therefore\ T^{k+1}(J_o) = J_{k+1}^*.$$

Let $\bar{\epsilon} > 0$ be given and let $\bar{\pi} = (\bar{I}_o, \bar{I}_1, \ldots)$ be such that

$$J_{k,\bar{I}} \geqq J_k^* - \bar{\epsilon}/2\alpha.$$

Let $\bar{I} \in M$ be such that

$$T_{\bar{I}}(J_k^*) \geqq T(J_k^*) - \bar{\epsilon}/2.$$

Consider the contract $\pi_{\bar{\epsilon}} = (\bar{I}, \bar{I}_o, \bar{I}_1, \ldots)$. Then

$$J_{k+1,\pi_{\bar{\epsilon}}} = T_{\bar{I}}(J_{k,\bar{\pi}}) \geqq T_{\bar{I}}(J_k^*) - \bar{\epsilon}/2 \geqq T(J_k^*) - \bar{\epsilon}$$

$$= J_{k+1}^* - \bar{\epsilon}. \quad \text{QED}$$

If Assumption F.1 were to be replaced by the following assumption, the results of Theorem 4.2 (a) can be strengthened.

Assumption F.1$'$: The function $J_o$ satisfies $J_o(\omega) \leqq H(\omega, I, J_o)$ $\forall\ \omega \in \Omega$, $I \in U(\omega)$ and if $\{J_k\} \subset F$ is a sequence such that $J_{k+1} \geqq J_k \geqq J_o$ for all $k$, then

$$\lim_{k\to\infty} H(\omega, I, J_k) = H(\omega, I, \lim_{k\to\infty} J_k) \quad \forall\ \omega \in \Omega,\ I \in U(\omega).$$

Corollary 4.2.1: Let F.1$'$ hold. Then $J_N^* = T^N(J_o)$.

Theorem 4.3: Let F.3 hold. Let $J_{k,\pi}(\omega) > -\infty$ $\forall\ \omega \in \Omega$, $\pi \in \Pi$ and $k = 1,2,\ldots,N$. Then $J_N^* = T^N(J_o)$. Furthermore, if $\{\varepsilon_n\}$ is a sequence of positive numbers with $\varepsilon_n \to 0$, then there exists a sequence of contracts $\{\pi_n\}$ exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality. If, in addition $J_N^*(\omega) < \infty$ $\forall\ \omega \in \Omega$, then for every $\varepsilon > 0$, there exists an $\varepsilon$-optimal contract.

Proof: Let $K = 1$.

$$\begin{aligned} J_1^*(\omega) &= \sup_{\pi\in\Pi} J_{1,\pi}(\omega) \\ &= \sup_{I\in M} H[\omega, I(\omega), J_o] \\ &= T(J_o)(\omega) \quad \forall\ \omega \in \Omega. \end{aligned}$$

Given $\{\epsilon_n\}$, by F.3, it is clear that there exists a sequence $\{\pi_n\} \subset \Pi$ satisfying

$$\lim_{n\to\infty} J_{1,\pi_n} = J_1^*$$

$$J_{1,\pi_n}(\omega) \geqq \begin{cases} J_1^*(\omega) - \epsilon_n & \forall\ \omega \in \cap\ J_1^* < \infty \\ J_{1,\pi_{n-1}}(\omega) - \epsilon_n & \forall\ \omega \in \Omega\ J_1^* = \infty \end{cases}$$

Suppose the result is true for $K = N-1$. Let $\beta$ be as specified in F.3. Consider a sequence $\{\epsilon_n\} \subset R$ with $\epsilon_n > 0\ \forall\ n$

$$\lim_{n\to\infty} \epsilon_n = 0 \quad \text{and} \quad \sum_{n=1}^{\infty} \epsilon_n < \infty.$$

Let $\{\hat{\pi}_n\} \subset \Pi$, where $\hat{\pi}_n = (I_1^n, I_2^n, \ldots)$ be such that

$$\lim_{n\to\infty} J_{N-1,\hat{\pi}_n} = J_{N-1}^*$$

$$J_{N-1,\hat{\pi}_n}(\omega) \geqq \begin{cases} J_{N-1}^*(\omega) - \beta^{-1}\epsilon_n & \forall\ \omega \in \Omega\ J_{N-1}^*(\omega) < \infty \\ J_{N-1,\hat{\pi}_{n-1}}(\omega) - \beta^{-1}\epsilon_n & \forall\ \omega \in \Omega\ J_{N-1}^*(\omega) = \infty \end{cases}$$

Since $J_{k,\pi}(\omega) > -\infty\ \forall\ \omega \in \Omega,\ k = 1,2,\ldots,N,\ \pi \in \Pi$ then $H(\omega, I, J_{N-1,\hat{\pi}_n}) > -\infty\ \forall\ \omega \in \Omega,\ I \in U(\omega)$. Since

$$\lim_{n\to\infty} J_{N-1,\hat{\pi}_n} = J_{N-1}^*.$$

By F.3, it then implies $\exists\ \{I_o^n\} \subset M$ such that $\forall\ n$

$$\lim_{n\to\infty} T_{I_o^n}(J_{N-1,\hat{\pi}_n}) = T(J_{N-1}^*)$$

$$T_{I_o^n}(J_{N-1,\hat{\pi}_n})(\omega) \geq \begin{cases} T(J^*_{N-1})(\omega) - \varepsilon_n & \text{if } T(J^*_{N-1})(\omega) < \infty \\ T_{I_o^{n-1}}(J_{N-1,\hat{\pi}_{n-1}})(\omega) - \varepsilon_n & \text{if } T(J^*_{N-1})(\omega) = \infty \end{cases}$$

By induction $J^*_{N-1} = T^{N-1}(J_o)$ and by definition of $J^*_N$, $T^N(J_o) \geqq J^*_N$. Hence

$$J(J^*_{N-1}) = T^N(J_o) \geqq J^*_N .$$

But

$$J^*_N \geqq \lim_{n\to\infty} T_{I_o^n}(J_{N-1,\hat{\pi}_n}) = T(J^*_{N-1})$$

$$\therefore \; J^*_N = T(J^*_{N-1}) = T^N(J_o).$$

Let $\pi_n = (I_o^n, I_1^n, I_2^n, \ldots)$. Then for all $n$

$$\lim_{n\to\infty} J_{N,\pi_n} = J^*_N$$

$$J_{N,\pi_n}(\omega) \geqq \begin{cases} J^*_N(\omega) - \varepsilon_n & \forall\ \omega \in \Omega \text{ with } J^*_N(\omega) < \infty \\ J^*_{N,\pi_{n-1}}(\omega) - \varepsilon_n & \forall\ \omega \in \Omega \text{ with } J^*_N(\omega) = \infty \end{cases}$$

Obviously, if $J^*_N(\omega) < \infty$, $\pi_n$ is $\varepsilon$-optimal. QED

Theorems 4.1 through 4.3 established the validity of adopting dynamic programming algorithm as solution procedures to the basic multiperiod agency problem. The existence of nearly optimal contracts has also been demonstrated. In fact, the dynamic programming algorithm can provide much stronger results than those as stated.

Under conditions provided in Theorems 4.4 and 4.5, the algorithm is actually attempting to arrive at a uniformly optimal contract.

Theorem 4.4: A contract $\pi^* = (I_o^*, I_1^*, \ldots)$ is uniformly N-stage optimal if and only if

$$(T_{I_k^*} T^{N-k-1})(J_o) = T^{N-k}(J_o) \qquad k = 0, \ldots, N-1.$$

Proof: Let $(T_{I_k^*} T^{N-k-1})(J_o) = T^{N-k}(J_o)$, $k = 0, \ldots, N-1$.

Then $(T_{I_k^*} \cdots T_{I_{N-1}^*})(J_o) = T^{N-k}(J_o)$. But

$$J_{N-k}^* \geqq (T_{I_k^*} \cdots T_{I_{N-1}^*})(J_o)$$

and

$$T^{N-k}(J_o) \geqq J_{N-k}^*$$

$$\therefore\ J_{N-k}^* = (T_{I_k^*} \cdots T_{I_{N-1}^*})(J_o)$$

$\therefore\ \pi^*$ is uniformly N-stage optimal.

Conversely, let $\pi^*$ be uniformly N-stage optimal. Then by definition, $T(J_o) = J_1^* = T_{I_{N-1}^*}(J_o)$. For every $I \in M$, $(T_I T)(J_o) = (T_I T_{I_{N-1}^*})(J_o)$. This implies

$$T^2(J_o) = \sup_{I \in M} (T_I T)(J_o)$$

$$= \sup_{I \in M} (T_I T_{I^*_{N-1}})(J_o)$$

$$\leqq J^*_2 = (T_{I^*_{N-2}} T_{I^*_{N-1}})(J_o) \leqq T^2(J_o)$$

$$\therefore T^2(J_o) = J^*_2 = (T_{I^*_{N-2}} T_{I^*_{N-1}})(J_o) = (T_{I^*_{N-1}} T)(J_o).$$

By induction, the result can easily be extended for all $k$.

QED

<u>Corollary 4.4.1</u>: There exists a uniformly N-stage optimal policy if and only if the supremum is attained in the relation

$$T^{k+1}(J_o)(\omega) = \sup_{I \in U(\omega)} H[\omega, I, T^k(J_o)]$$

for each $\omega \in \Omega$ and $k = 0,1,\ldots,N-1$.

<u>Corollary 4.4.2</u>: If there exists a uniformly N-stage optimal policy, then

$$J^*_N = T^N(J_o).$$

Theorem 4.4 and its corollaries state that if the supremum can be achieved at each stage of the optimization process, the resulting contract is uniformly optimal. The next theorem states that if the incentive constraint set is compact, the existence of a supremum at each stage is guaranteed.

Lemma 4.5: Let $C$ be a Hausdorff space, $f : C \to R^*$ and $U$ a subset of $C$. Let the set $U(\lambda) = \{I \in U | f(I) \geqq \lambda\}$ is compact for each $\lambda \in R$. Then $f$ attains a maximum over $U$.

Proof: If $f(I) = \infty$ for all $I \in U$, then every $I \in U$ attains the maximum. If $f^* = \sup\{f(I) | I \in U\} < \infty$, let $\{\lambda_n\}$ be an increasing sequence such that $\lambda_n < \lambda_{n+1}$ for all $n$ and

$$\lambda_n \to f^*.$$

Thus $U(\lambda_n) \subset U(\lambda_{n+1})$ for all $n$ and the sets $U(\lambda_n)$ are nonempty and compact, $\bigcap_{n=1}^{\infty} U(\lambda_n) = U(\lambda_o)$ is nonempty and compact. Let $I^* \in \bigcap_{n=1}^{\infty} U(\lambda_n) \subset U$ and $f(I^*) \geqq \lambda_n$ for all $n$

$$\therefore f(I^*) \geqq f^*.$$

But $f(I^*) \leqq f^*$

$$\therefore f(I^*) = f^*. \quad \text{QED}$$

Theorem 4.5: Let the incentive space $C$ be a Hausdorff space and assume that for each $\omega \in \Omega$, $\lambda \in R$ and $k = 0, 1, \ldots, N-1$. The set $U_k(\omega, \lambda) = \{I \in U(\omega) | H[\omega, I, T^k(J_o)] \geqq \lambda\}$ is compact. Then $J_N^* = T^N(J_o)$ and there exists a uniformly N-stage optimal policy.

Proof: Direct application of Corollary 4.4.1 and 4.4. and Lemma 4.5.

# CHAPTER V

# INFINITE HORIZON MODELS

## 5.1: Introduction

The construction of the infinite horizon models requires a few more considerations than those in the finite horizon models. Recall that in the economic formulation of the model $J_\pi(\omega) = \lim_{N\to\infty} J_{N,\pi}(\omega)$, that is, $J_\pi(\omega)$ is the limiting value of the total expected discounted net return to the principal as $N$ gets large and $J_\pi(\omega) = \lim_{N\to\infty} (T_{I_o} \cdots T_{I_{N-1}})(J_o)(\omega)$ in the dynamic programming model. The natural initial question would be whether such limits exist and if they do, would they be equal. In addition, an immediate concern would be the convergence of the dynamic programming algorithm. Another important matter to resolve is the question of whether or not optimal or nearly optimal contracts exist.

This chapter will address the above problems on two separate sets of assumptions: contraction and monotonicity. Both assumptions are reasonable descriptions of the economic environments of a business enterprise.

The contraction assumption says that payoffs and returns that are receivable at a far distant future have very little significance in the economic decisions that are made today. In computing present values, it is well-known that with a discount factor of less than unity, the present value of any finite amount for a long period of time is close to zero. A discount factor of less than unity means that interest rate is always greater than the inflation rate.

The second set of assumptions describe a monotonicity behavior on the total net return function $J$. This assumption can be set up in two distinct scenarios. First, $J$ can be monotonic increasing with respect to time. Since $J$ is the total accumulated expected net return to the principal, the assumption that $J$ is monotonic increasing implies that the net payoff each period to the principal is greater or equal to zero. If one were to consider the net return for each period in terms of expected value over all possible payoffs, the requirement that the expected net payoff to the principal is positive makes economic sense. The entity will not survive if the expected payoff is negative. On the other hand, if the total net return function is monotonic decreasing, the situation considered is still not totally unrealistic. Consider the following situation. An entity has been suffering losses consistently from period to period. This would mean that the payoff for each period is negative and consequently $J$ is monotonic decreasing. The objective of the model is to maximize expected total return which means that it will search for some incentive contracts that would make the magnitude of period loss diminish over time with the expectation that the loss will become identically zero and eventually swing over to the positive side. Once the period payoff becomes positive, $J$ will be monotonic increasing and the modeling will continue under the earlier set of assumptions. Of course, such a situation will occur if the return becomes positive before the company goes bankrupt.

### 5.2: The Contraction Assumption

The contraction assumption is motivated by the contraction property of the mapping $H$ associated with discounted stochastic optimal total return with bounded net return per period. As mentioned earlier, this assumption will be satisfied with the discount factor, $\alpha < 1$ and the net return function $g$ uniformly bounded above and below. As always is the case, bounded return is an extremely reasonable assumption in most economic settings. Theorem 5.1 will show that indeed H as proposed in Chapter III satisfies the contraction assumption.

The next set of questions to be considered involve whether the iteration process will generate contracts that lead to an optimal total return function, that is, whether $J_N^* = T^N(J)$, and when the number of periods N is allowed to tend to infinity, whether the dynamic programming algorithm converges to any kind of limit. If a finite limit exists, what is that limit equal to? It can be shown that such a limit does exist and is equal to the infinite horizon total return function for any incentive contract (Theorem 5.2(a)). Also, it is shown that $J_N^*$ equals to $T^N(J)$ for all N. These two statements together imply that J*, the infinite horizon optimal return function is equal to $\lim_{N\to\infty} T^N(J)$ (Theorem 5.2(b)). Also the mappings $T^m$ and $T_1^m$, where $m \leq N$ are contraction mappings (Theorem 5.2(c)).

With the aid of the Fixed Point Theorem which says that a contraction mapping will converge to a unique fixed point, J* is shown to be the unique fixed point. Therefore, the validity of the dynamic programming algorithm is established for the contraction assumptions.

Next, some attempts to describe the optimal contract are made. If for each financial position, involving both the total return to date and the last net payoff, there exists an optimal contract at that position, then the optimal contract is stationary (Theorem 5.5). This means that the same form of contract is optimal over the entire time span. This result can further be strengthened by showing that the maximum is attained for each financial position if the incentive constraint set is compact. Compactness means that the set is closed and bounded and is the usual requirement for optimization problems. Also, since $T$ is a contraction mapping, the sequence $T^m(J)$ does converge to a limit as $m$ tends to infinity. Hence the optimal contract may be obtained in the limit from finite horizon optimal contracts by successively computing $T(J), T^2(J), \ldots$ (Theorem 5.6). The convergence of $T^m(J)$ implies the convergence of the dynamic programming algorithm for the infinite horizon model under the contraction assumption.

As it has been stated, the contraction assumption yields very desirable results. The optimal total return $J^*$ can be computed through a finite horizon mode $T^N(J)$ by letting $N$ go to infinity. It is unique. If the maximum is attained at each stage of the iteration process, then the resulting optimal contract is stationary.

## 5.3: The Monotonicity Assumption

In this section, two separate and parallel sets of monotonicity assumptions are considered. It is assumed that $J_k$, the total return function to the principal at the end of time period $k$ is a monotonic increasing or monotonic decreasing function over, that is $J_o(\omega) \leqq H(\omega, I, J_o)$ (Assumption I) or $J_o(\omega) \geqq H(\omega, I, J_o)$ (Assumption D) for all $\omega \in \Omega$ and all $I \in U(\omega)$. The economic motivation and interpretation of these two assumptions have been discussed in the introduction of this chapter. In the following analysis under Assumptions I or D, two additional assumptions are required on the behavior of the function H. The first Assumption I.1 or D.1, describes a continuity property on $H$. Similar to Assumption One of the finite horizon models, this assumption guarantees that for small changes in the value of $J$, there will be only small corresponding changes in the value of $H$. The second Assumption, I.2 or D.2, imposes a kind of linearity on $H$ from below. This guarantees that $H$ cannot decrease more than a fixed multiple with a decrease in $J$. Again, this assumption is similar to Assumption Two under the finite horizon model. The economic implications of these two assumptions will be identical to that under the finite horizon models.

As expected, Assumption I will be satisfied if $g$, the net return to the principal per period is positive and Assumption D is satisfied if $g$ is negative (Theorem 5.7). Hence, the analysis can be carried on to investigate the equivalence of the economic and dynamic programming model under these two assumptions. This is done in three stages: (1) the optimality equation $J^* = T(J^*)$ is investigated, (2) the question of existence of optimal or nearly optimal contracts is settled, and (3) the convergence of the dynamic programming algorithm is shown.

5.4: The Optimality Equation

Before considering whether the optimality equation $J^* = T(J^*)$ holds, the existence of a nearly optimal contract for Assumption D is first established. If Assumptions D, D.1 and D.2 all hold, and the optimal return $J^*$ is finite, the existence of nearly optimal contracts is guaranteed (Theorem 5.8). In addition, if the discount factor $\alpha$ is less than one, then the nearly optimal contract is stationary.

The optimality equation can be established under all Assumptions D, D.1 and D.2 (Theorem 5.9 and Corollary 5.9.1). However, the optimality equation holds under Assumption I only when Assumption I is accompanied by one of the additional conditions I.1 or I.2 (Theorem 5.10 and Corollary 5.10.1). A consequence of the results of this section is that the validity of adopting the dynamic programming algorithm for the monotonicity assumptions is established.

## 5.5: Convergence to Optimality and Existence of Optimal Contracts

Under Assumptions D, D.1 and D.2, it can be shown that if the supremum is attained in the optimality equation

$$J^*(\omega) = \sup_{I \in U(\omega)} H(\omega, I, J^*)$$

for every $\omega \in \Omega$, then the resulting contract is an optimal stationary contract. In fact, the conditions are both necessary and sufficient (Theorem 5.11). The same results can be obtained under Assumption I (Theorem 5.12). However, to arrive at the conclusion, one additional assumption, Assumption I.1, is required.

If the supremum cannot be attained for some $\omega \in \Omega$, it is shown (Theorem 5.13) that, under Assumptions D, D.1 and D.2, the optimality equation can still be used to construct a nearly optimal contract. In addition, if the discount factor is less than one, the nearly optimal contract as constructed is stationary.

Under Assumption I, only a weak counterpart of the above results can be given. If Assumptions I and I.2 hold and if the optimal total return function $J^*$ is finite, then a nearly optimal contract exists (Theorem 5.15). However, I am unable to give a counterpart under Assumption I on conditions for existence of a stationary nearly optimal contract.

The dynamic programming model requires successive generation of the functions $T(J_o), T^2(J_o), \ldots, T^k(J_o), \ldots$ for all $k$. In terms of the infinite horizon model, it seems appropriate to define a function $J_\infty$ by

$$J_\infty(\omega) = \lim_{N\to\infty} T^N(J_o)(\omega) \quad \text{for every} \quad \omega \in \Omega.$$

The rest of this chapter will be devoted to the discussion of whether $J_\infty$ exists and whether $J_\infty = J^*$. In other words, it is a concern that whether the dynamic programming algorithm converges under the two assumptions and if it does, will the limit be the same as the optimal total return.

Fortunately, if Assumption I and I.1 hold, all the above questions can be responded affirmatively (Theorem 5.14). However, under Assumption D, the equality $J_\infty = J^*$ may fail to hold even in very simple situations. A preliminary result shows that in order to have $J_\infty = J^*$, it is necessary and sufficient to have $J_\infty = T(J_\infty)$ (Theorem 5.16). The convergence of the dynamic programming algorithm under Assumption D is essential to arrive at the optimal total return.

To show tha fact that $J_\infty = T(J_\infty)$ requires some elaborate technical details concerning the algebraic structure of the various functions. These are presented in detail in Chapter V*. It can be generally stated that if the incentive constraint set is compact, then $J_\infty = T(J_\infty) = T^*$ under Assumptions D, D.1 and D.2 (Theorems 5.16 through 5.18).

Once convergence is established, it is shown (Theorem 5.19) that the limit of the dynamic programming algorithm is an optimal contract and it is stationary.

## 5.6: Remarks

In Chapters III through V, the focus of attention is on a basic multiperiod agency model in which the period payoffs are observable to both the principal and the agent and the sequence of payoffs are stochastically well-defined with a known probability distribution. Analysis have been conducted on a finite and infinite horizon settings. The circumstances or conditions under both settings that will guarantee the principal's ability to negotiate an optimal contract with the agent are extremely mild.

The two conditions imposed on the finite horizon model are both on the manner in which the total net return to the principal in the $k+1$st period behaves relative to the total net return to the $k$th period. One of the conditions says that if the total return, or total accumulated wealth in the entity, in period $k$ changes by a small amount, then given all other factors equal, the expected change in the total wealth of period $k+1$ is also small. The second condition goes on further to say that the change is somewhat linear. These two together implies that given some changes in the wealth of a company, the effect of such changes in subsequent periods is proportionate to the change. Both are conditions on the economic environment in which the entity operates. I believe that most if not all business enterprises are operating under these circumstances. If future wealth were so unpredictable with respect to current changes in company wealth, companies will be very hesitent to declare dividends, acquire ventures which requires capital outlays.

The contraction and monotonicity assumptions under the infinite horizon model are also conditions on the economic environment of the enterprise. The contraction assumption relies on the same assumptions that any discounted cash flow or present value models take on. These assumptions which include bounded payoffs and discount factor less than one are widely accepted in the literature. The monotonicity assumption is based on the belief that no company will remain in business if its expected period payoff is negative. If any one of these two assumptions is met and consideration of the model is not restricted to a finite number of periods, one would be able to construct an optimal stationary contract. This is possible because by extending the planning horizon to an infinite period of time, one allows time to become a monitoring device. By repeating the process over long enough periods, the behavior of the agent becomes extremely predictable. On the other hand, if the agent intends to remain in the company "forever", it will be difficult for him to cheat "forever" and remain undiscovered. It will then be in his best interest to act cooperatively with the principal.

Certainly, if the number of periods in the finite horizon model is allowd to extend long enough, all the infinite horizon model effects will be carried through. The limiting effects of the infinite horizon model provide an additional nicety to the optimal contract: that it is stationary.

# CHAPTER V*

# INFINITE HORIZON MODEL

## 5.1* The Contraction Assumption

The following assumption is motivated by the contraction property of the mapping associated with discounted stochastic optimal total return with bounded return per period.

Assumption C: Let $B$ be the Banach space of all bounded real-valued functions on $\Omega$ with the supremum norm. There exists a closed subset $B_o \subset B$ such that $J_o \in B_o$, and for all $J \in B_o$, $I \in M$, the functions $T(J)$ and $T_I(J)$ belong to $B_o$. Furthermore, for every $\pi = (I_o, I_1, \ldots) \in \Pi$, the limit

$$\lim_{N\to\infty} (T_{I_o} T_{I_1} \cdots T_{I_{N-1}})(J_o)(\omega)$$

exists and is a real number for each $\omega \in \Omega$. In addition, there exists a positive integer $m$ and scalars $\varpi, \alpha$, with $0 < \varpi < 1$, $\alpha > 0$, such that

$$\|T_I(J) - T_I(J')\| \leq \alpha \|J - J'\| \quad \forall\ I \in M,\ J, J' \in B$$

$$\|(T_{I_o} T_{I_1} \cdots T_{I_{m-1}})(J) - (T_{I_o} T_{I_1} \cdots T_{I_{m-1}})(J')\|$$

$$\leq \varpi \|J - J'\| \quad \forall\ I_o, \ldots, I_{m-1} \in M,\ J, J' \in B_o.$$

The assumption is stated in a very general setting. It is often convenient to take $\bar{B} = B$ and assume $\alpha < 1$ with $g$ uniformly bounded above and below. However, in some special cases, the contraction property can be verified only on a strict subset $\bar{B}$ of $B$.

To start the analysis, it is first shown that the function $H$ meets the requirements of the assumption.

<u>Theorem 5.1</u>: Let $H(\omega,I,J) = E^*\{g(\omega,I,y) + \alpha J[f(\omega,I,y)] | \omega,I\}$ be a mapping. Let $J_o(\omega) = 0$ for every $\omega \in \Omega$. Assume that $\alpha < 1$ and for some $b \in R$, there hold

$$0 \leqq g(\omega,I,y) \leqq b \quad \forall\ \omega \in \Omega,\ I \in U(\omega),\ y \in \mathcal{Y}.$$

Then Assumption C is satisfied with $B_o = \{J : J \in B,\ J \geqq 0\}$ and $\alpha = 2\alpha$, $\varphi = \alpha$ and $m = 1$.

Proof: Clearly $J_o \in B_o$ and $T(J)$, $T_I(J) \in B_o$ for all $J \in B_o$ and $I \in M$. Then for any $\pi = (I_o, I_1, \ldots) \in \Pi$,

$$J_o \leqq T_{I_o}(J_o) \leqq \cdots \leqq (T_{I_o} \cdots T_{I_k})(J_o)$$

$$\leqq (T_{I_o} \cdots T_{I_{k+1}})(J_o) \leqq \cdots$$

$$\therefore \lim_{N\to\infty} (T_{I_o} \cdots T_{I_{N-1}})(J_o)(\omega) \quad \text{exists for all} \quad \omega \in \Omega.$$

Since

$$H = (\omega, I, J) = \int^* \{g(\omega, I, y) + \alpha J[f(\omega, I, y)]\} p(dy|\omega, I)$$
$$\leqq \int^* g(\omega, I, y) p(dy|\omega, I) + \alpha \int^* J[f(\omega, I, Y)] p(dy|\omega, I)$$
$$\leqq b + \alpha \int^* J[f(\omega, I, y)] p(dy|\omega, I).$$

Thus

$$(T_{I_o} \cdots T_{I_{N-1}})(J_o)(\omega) \leqq \sum_{k=0}^{N-1} \alpha^k b$$
$$\leqq \frac{b}{1-\alpha} \quad \forall\ \omega \in \Omega,\ N = 1, 2, \ldots$$

$$\therefore \lim_{N \to \infty} (T_{I_o} \cdots T_{I_{N-1}})(J_o)(\omega) \in R \quad \forall\ \omega \in \Omega.$$

$$g[\omega, I(\omega), y] + \alpha J[f(\omega, I(\omega), y)]$$
$$\leqq g[\omega, I(\omega), y] + \alpha J'[f(\omega, I(\omega), y)] + \alpha \|J - J'\|$$
$$\forall\ \omega \in \Omega,\ J, J' \in B,\ I \in M \text{ and } y \in \mathcal{Y}.$$

By Lemma 4.1,

$$\int^* \{g[\omega, I(\omega), y] + \alpha J[f(\omega, I(\omega), y)]\} p(dy|\omega, I)$$
$$\leqq \int^* \{g[\omega, I(\omega), y] + \alpha J'[f(\omega, I(\omega), y)]\} p(dy|\omega, I)$$
$$+ 2\alpha \|J - J'\|.$$

Hence

$$T_I(J)(\omega) - T_I(J')(\omega) \leqq 2\alpha \|J - J'\|.$$

Similarly, we obtain

$$T_I(J')(\omega) - T_I(J)(\omega) \leqq 2\alpha \|J - J'\|$$

$$\therefore |T_I(J)(\omega) - T_I(J')(\omega)| \leqq 2\alpha \|J - J'\|.$$

Taking supremum on the left-hand side over $\omega \in \Omega$

$$\|T_I(J)(\omega) - T_1(J')(\omega)\| \leqq 2\alpha \|J - J'\| \quad \forall\ I \in M,\ J, J' \in B.$$

If $J, J' \in B_o$, again we obtain

$$g[\omega, I(\omega), y] + \alpha J[f(\omega, I(\omega), y] \leqq g[\omega, I(\omega), y]$$
$$+ \alpha J'[f(\omega, I(\omega), y)] + \alpha \|J - J'\|.$$

Then

$$\int^* \{g[\omega, I(\omega), y] + \alpha J[f(\omega, I(\omega), y]\} p(dy|\omega, I)$$
$$\leq \int^* \{g[\omega, I(\omega), y] + \alpha J'[f(\omega, I(\omega), y)] + \alpha\|J - J'\|\} p(dy|\omega, I)$$
$$\leq \int^* \{g[\omega, I(\omega), y] + \alpha J'[f(\omega, I(\omega), y)]\} p(dy|\omega, I) + \alpha\|J - J'\|.$$

Proceeding as before, we obtain

$$\|T_I(J) - T_I(J')\| \leq \alpha\|J - J'\| \quad \forall\ I \in M,\ J, J' \in B_o. \qquad \text{QED}$$

Theorem 5.2: Let Assumption C hold, then

(a) For every $J \in B_o$ and $\pi \in \Pi$,

$$J_\pi = \lim_{N\to\infty} (T_{I_o} \cdots T_{I_{N-1}})(J_o) = \lim_{N\to\infty} (T_{I_o} \cdots T_{I_{N-1}})(J).$$

(b) For each positive integer $N$ and each $J \in B_o$,

$$\sup_{\pi\in\Pi} (T_{I_o} \cdots T_{I_{N-1}})(J) = T^N(J) \quad \text{and}$$

$$J_N^* = \sup_{\pi\in\Pi} (T_{I_o} \cdots T_{I_{N-1}})(J_o) = T^N(J_o).$$

(c) The mappings $T^M$ and $T_I^M$, $I \in M$ are contraction mappings in $B_o$ with modulus $\varphi$.

Proof: (a) Let $k \geq 0$ be any integer and $k = nm + q$ where $q, n \geq 0$ and $0 \leq q < m$. By C, for any $J, J' \in B_o$

$$\|(T_{I_o} \cdots T_{I_{k-1}})(J) - (T_{I_o} \cdots T_{I_{k-1}})(J')\| \leq \varphi^n \alpha^q \|J - J'\|.$$

Since $J_o \in B_o$

$$\|(T_{I_o} \cdots T_{I_{k-1}})(J) - (T_{I_o} \cdots T_{I_{k-1}})(J_o)\|$$
$$\leq \varphi^n \alpha^q \|J - J_o\|.$$

Taking limits as $k \to \infty$ and $n \to \infty$

$$\lim_{k\to\infty} (T_{I_o} \cdots T_{I_{k-1}})(J) = \lim_{k\to\infty} (T_{I_o} \cdots T_{I_{k-1}})(J_o).$$

(b) By assumption $T^k(J) \in B_o$ for all $k$. Thus $T^k(J)(\omega) < \infty \ \forall\ \omega \in \Omega$ and $k$. For any $\varepsilon > 0$, let $\bar{I}_k \in M$, $k = 0,1,\ldots,N-1$ be such that

$$T_{\bar{I}_{N-1}}(J) \geqq T(J) - \varepsilon$$

$$(T_{\bar{I}_{N-2}} T)(J) \geqq T^2(J) - \varepsilon$$

$$(T_{\bar{I}_o} T^{N-1})(J) \geqq T^N(J) - \varepsilon.$$

By Assumption C

$$T^N(J) \leqq (T_{\bar{I}_o} T^{N-1})(J) + \varepsilon$$

$$\leqq T_{\bar{I}_o}[(T_{\bar{I}_1} T^{N-2})(J) + \varepsilon] + \varepsilon$$

$$\leqq (T_{\bar{I}_o} T_{\bar{I}_1}, T^{N-2})(J) + \alpha\varepsilon + \varepsilon$$

$$\leqq (T_{\bar{I}_o} T_{\bar{I}_1} \cdots T_{\bar{I}_{N-1}})(J) + (\sum_{k=0}^{N-1} \alpha^k \varepsilon)$$

$$\leqq \sup_{\pi\in\Pi} (T_{I_o} \cdots T_{I_{N-1}})(J) + (\sum_{k=0}^{N-1} \alpha^k \varepsilon)$$

$$\therefore\ T^N(J) \leqq \sup_{\pi\in\Pi} (T_{I_o} \cdots T_{I_{N-1}})(J).$$

But $T^N(J) \geqq \sup_{\pi\in\Pi} (T_{I_o} \cdots T_{I_{N-1}})(J)$ by definition

$$\therefore\ J_N^* = \sup_{\pi\in\Pi} (T_{I_o} \cdots T_{I_{N-1}})(J) = T^N(J).$$

(c) By Assumption C, $T_I^m$ is a contraction mapping. Also for all $I_k \in M$, $k = 0,\ldots,m-1$, and $J, J' \in B_o$

$$(T_{I_o} \ldots T_{I_{m-1}})(J) \leqq (T_{I_o} \ldots T_{I_{m-1}})(J') + \varphi\|J - J'\|.$$

Taking supremum of both sides over $I_k \in M$ and from part (b)

$$T^m(J) \leqq T^m(J') + \varphi\|J - J'\|.$$

Similarly $T^m(J') \leqq T^m(J) + \varphi\|J - J'\|$

$$\therefore \|T^m(J) - T^m(J')\| \leqq \varphi\|J - J'\|. \quad \text{QED}$$

Theorem 5.2 provides some very preliminary results. It establishes the validity of the dynamic programming algorithm. The next theorem is the well-known Fixed Point Theorem in Banach Space which is quoted below without proof.

<u>Theorem 5.3</u> (Fixed Point Theorem): If $B_o$ is a closed subset of a Banach space with an appropriate norm and $L : B_o \to B_o$ is a mapping such that for some positive integer $m$ and scalar $\varphi \in (0,1)$,

$$\|L^m(Z) - L^m(Z')\| \leqq \varphi\|Z - Z'\| \text{ for all } Z, Z' \in B_o.$$

Then $L$ has a unique fixed point in $B_o$. Furthermore, for every $Z \in \bar{B}$,

$$\lim_{N\to\infty} \|L^N(Z) - Z^*\| = 0$$

where $Z^* \in B_o$ such that $L(Z^*) = Z^*$

With the aid of the Fixed Point Theorem, Theorem 5.4 characterizes the optimal total return function $J^*$ and Theorem 5.5 the total return function $J_I$ corresponding to any stationary contract $(I, I, \ldots) \in \Pi$. It also shows that these functions can be obtained in the limit via successive application of $T$ and $T_I$ on any $J \in B$.

Theorem 5.4: Let Assumption C hold. Then

(a) The optimal profit function $J^* \in B_o$ and is the unique fixed point of $T$ within $B_o$. Furthermore, if $J' \in B_o$ is such that $T(J') \geqq J'$, then $J^* \geqq J'$

(b) For every $I \in M$, the function $J_I \in B_o$ and is the unique fixed point of $T_I$ within $B_o$.

(c) $\lim_{N\to\infty} \|T^N(J) - J^*\| = 0 \quad \forall\, J \in B_o$

$\lim_{N\to\infty} \|T_I^N(J) - J_I\| = 0 \quad \forall\, J \in B_o,\ I \in M.$

Proof: By Theorems 5.2(c) and 5.3, $T$ and $T_I$ have unique fixed points in $B_o$. Clearly $T_I^N(J) = J_I$. Thus part (b) is proved.

Let $\tilde{J}^*$ be the fixed point

$$\tilde{J}^* = T(\tilde{J}^*).$$

For any $\bar{\epsilon} > 0$, let $\bar{I} \in M$ be such that

$$T_{\bar{I}}(\tilde{J}^*) \geqq \tilde{J}^* - \bar{\epsilon}.$$

By Assumption C that $\|T_I(J) - T_I(J')\| \leqq \alpha\|J - J'\|$

$$T_{\bar{I}}^2(\tilde{J}^*) \geqq T_{\bar{I}}(\tilde{J}^*) - \alpha\bar{\epsilon} \geqq \tilde{J}^* - (1+\alpha)\bar{\epsilon}.$$

Continuing, $T_{\bar{I}}^m(\tilde{J}^*) \geqq \tilde{J}^* - (1+\alpha+\ldots+\alpha^{m-1})\bar{\epsilon}$. By the assumption that $\|(T_{I_o}T_{I_1}\ldots T_{I_{m-1}})(J) - (T_{I_o}T_{I_1}\ldots T_{I_{m-1}})(J')\|$ $\leqq \varphi\|J - J'\|$

$$T_{\bar{I}}^{2m}(\tilde{J}^*) \geqq T_{\bar{I}}^m(\tilde{J}^*) - \varphi(1+\alpha+\ldots+\alpha^{m-1})\bar{\epsilon}$$
$$\geqq \tilde{J}^* - (1+\varphi)(1+\alpha+\ldots+\alpha^{m-1})\bar{\epsilon}.$$

Thus for all $k \geqq 1$

$$T_{\bar{I}}^{km}(\tilde{J}^*) \geqq \tilde{J}^* - (1+\varphi+\ldots+\varphi^{k-1})(1+\alpha+\ldots+\alpha^{m-1})\bar{\epsilon}.$$

Since $J_{\bar{I}} = \lim_{k\to\infty} T_{\bar{I}}^{km}(\tilde{J}^*)$. Taking limits as $k \to \infty$

$$J_{\bar{I}} \geqq \tilde{J}^* - \frac{1}{1-\varphi}(1+\alpha+\ldots+\alpha^{m-1})\bar{\epsilon}.$$

Let $\bar{\epsilon} = (1-\varphi)(1+\alpha+\ldots+\alpha^{m-1})^{-1}\epsilon$

$$J_{\bar{I}} \geqq \tilde{J}^* - \epsilon.$$

But $J^* \geqq J_{\bar{I}}$ and $\epsilon > 0$ is arbitrary

$$\therefore J^* \geqq \tilde{J}^*.$$

On the other hand,

$$J^* = \sup_{I \in \Pi} \lim_{N \to \infty} (T_{I_o} \ldots T_{I_{N-1}})(\tilde{J}^*)$$

$$\leq \lim_{N \to \infty} T^N(\tilde{J}^*) = \tilde{J}^*$$

$$\therefore J^* = \tilde{J}^*.$$

By Theorem 5.3, part (c) follows immediately. Since $T$ is a monotonic mapping by assumption by part (c) it follows that

if $J' \in B_o$ such that $T(J') \geq J'$

then $J^* \geq J'$. QED

<u>Theorem 5.5</u>: Let Assumption C hold. Then

(a) A stationary contract $\pi^* = (I^*, I^*, \ldots) \in \Pi$ is optimal if and only if $T_{I^*}(J^*) = T(J^*)$. Equivalently, $\pi^*$ is optimal if and only if

$$T_{I^*}(J_{I^*}) = T(J_{I^*})$$

(b) If for each $\omega \in \Omega$ there exists a contract which is optimal at $\omega$, then there exists a stationary optimal contract.

Proof: (a) If $\pi^*$ is optimal, then $J_{I^*} = J^*$. By Theorem 5.4(a) and (b)

$$T_{I^*}(J^*) = T(J^*).$$

If $T_{I^*}(J^*) = T(J^*)$, then $T_{I^*}(J^*) = J^*$. By Theorem 5.4(a), $J_{I^*} = J^*$. If $T_{I^*}(J_{I^*}) = T(J_{I^*})$. Again, by Theorem 5.4(a)

$$J_{I^*} = T(J_{I^*}) = J^*.$$

(b) Let $\Pi^*_\omega = (I^*_{o,\omega}, I^*_{1,\omega}, \ldots)$ be an optimal contract at $\omega \in \Omega$.

By Theorems 5.2(a) and 5.4(a)

$$\begin{aligned} J^*(\omega) &= J_{I^*_\omega}(\omega) \\ &= \lim_{k\to\infty} (T_{I^*_{o,\omega}} \cdots T_{I^*_{k,\omega}})(J_o)(\omega) \\ &= \lim_{k\to\infty} (T_{I^*_{o,\omega}} \cdots T_{I^*_{k,\omega}})(J^*)(\omega) \\ &\leqq \lim_{k\to\infty} (T_{I^*_{o,\omega}} T^k)(J^*)(\omega) \\ &= T_{I^*_{o,\omega}}(J^*)(\omega) \\ &\leqq T(J^*)(\omega) \\ &= J^*(\omega) \end{aligned}$$

$\therefore$ $T_{I^*_{o,\omega}}(J^*)(\omega) = T(J^*)(\omega)$ for each $\omega$.

Define $I^*(\omega) = I^*_{o,\omega}(\omega)$ for $I^* \in M$. Then $T_{I^*}(J^*) = T(J^*)$. By part (a) the stationary contract $(I^*, I^*, \ldots)$ is optimal. QED

Theorem 5.5 also establishes the existence and characterization of stationary optimal contracts. Part (a) of Theorem 5.5 shows that there exists a stationalry optimal contract if and only if the supremum is attained for every $\omega \in \Omega$ in the optimality equation, $J^* = T(J^*)$. Theorem 5.6 strengthens this result by showing that the supremum is attained if $U_k(\omega,\lambda)$, the incentive constraint set is compact. It also shows that stationary optimal contracts may be obtained in the limit form finite horizon optimal contracts by successively computing $T(J), T^2(J), \cdots$. At the same time, it also proves the convergence of the algorithm.

<u>Theorem 5.6</u>: Let Assumption C hold. Let the incentive space $C$ be a Hausdorff space. Suppose that for some $J \in B_o$ and some integers $k_o > 0$, the sets

$$U_k(\omega,\lambda) = \{I \in U(\omega) \,|\, H[\omega, I, T^k(J)] \geqq \lambda\}$$

are compact for all $\omega \in \Omega$, $\lambda \in \mathbb{R}$ and $k \geqq k_o$. Then

(a) there exists a contract $\pi^* = (I_o^*, I_1^*, \ldots) \in \Pi$ attaining the supremum for all $\omega \in \Omega$ and $k \geqq k_o$ with initial function $J$, i.e.,

$$(T_{I_k^*} T^k)(J) = T^{k+1}(J) \;\; \forall \; k \geqq k.$$

(b) For every contract $\pi^*$ satisfying (a), the sequence $\{I_k^*(\omega)\}$ has at least one limit point for each $\omega \in \Omega$.

(c) If $I^*: \Omega \to C$ is such that $I^*(\omega)$ is a limit point of $\{I_k^*(\omega)\}$ for each $\omega \in \Omega$, then the stationary policy $(I^*, I^*, \ldots)$ is optimal.

Proof: (a) Since $T^{k+1}(J)(\omega) = \sup\limits_{I \in U(\omega)} H[\omega, I, T^k(J)]$ and $U_k(\omega, \lambda)$ are compact for $k \geq k_o$. By Lemma 4.5, $T^{k+1}(J)(\omega)$ attains a maximum

$$\therefore \; (T_{I_k^*} T^k)(J) = T^{k+1}(J).$$

(b) Let $I^* = (I_o^*, I_1^*, \ldots)$ satisfy part (a).

Define

$$\epsilon_k = \sup\{\|J^* - T^i(J)\| \mid i \geq k\} \quad k = 0, 1, \cdots.$$

Since $T(J^*) = J^*$. By Assumption C and part (a)

$$\begin{aligned} \|J^* - (T_{I_n^*} T^n)(J)\| &= \|T(J^*) - T^{n+1}(J)\| \\ &\leq \alpha \|J^* - T^n(J)\| \quad \forall\, n \geq k_o \end{aligned}$$

and

$$\begin{aligned} \|(T_{I_n^*} T^n)(J) - (T_{I_k^*} T^k)(J)\| &\leq \alpha \|T^n(J) - T^k(J)\| \\ &\leq \alpha \|J^* - T^n(J)\| + \alpha \|J^* - T^k(J)\| \\ &\qquad \forall\, n \geq k_o, \quad k = 0, 1, \cdots. \end{aligned}$$

Thus

$$\begin{aligned} H[\omega, I_n^*(\omega), T^k(J)] &\geq H[\omega, I_n^*(\omega), T^n(J)] - 2\alpha\epsilon_k \\ &\geq J^*(\omega) - 3\alpha\epsilon_k \quad \forall\, n \geq k, \quad k \geq k_o \end{aligned}$$

$$\therefore \; I_n^*(\omega) \in U_k[\omega, J^*(\omega) - 3\alpha\epsilon_k] \quad \text{for all} \quad n \geq k \quad \text{and} \quad k \geq k_o.$$

Since $U_k[\omega, J^*(\omega) - 3\alpha\epsilon_k]$ is compact

$\therefore$ $\{I_n^*(\omega)\}$ has a limit point in $U_k[\omega, J^*(\omega) - 3\alpha\epsilon_k]$.

(c) Let $I^*(\omega)$ be a limit point of $\{I_n^*(\omega)\}$.

By part (b), $I^*(\omega) \in U_k[\omega, J^*(\omega) - 3\alpha\epsilon_k]$ $\forall$ $k \geqq k_o$. Thus

$$(T_{I^*}T^k)(J)(\omega) \geqq J^*(\omega) - 3\alpha\epsilon_k \;\forall\; \omega \in \Omega,\; k \geqq k_o.$$

By Assumption C, for all $k$

$$\|T_{I^*}(J^*) - (T_{I^*}T^k)(J)\| \leqq \alpha\|J^* - T^k(J)\| \leqq \alpha\epsilon_k$$

$$\therefore\; T_{I^*}(J^*)(\omega) \geqq J^*(\omega) - 4\alpha\epsilon_k.$$

By Theorem 5.4(c), $\epsilon_k \to 0$

$$\therefore\; T_{I^*}(J^*) \geqq J^*(\omega).$$

But $J^* = T(J^*) \geqq T_{I^*}(J^*)$

$$\therefore\; T_{I^*}(J^*) = J^*.$$

By Theorem 5.5, the stationary contract $(I^*, I^*, \ldots)$ is optimal. QED

## 5.2* The Monotonicity Assumptions

For the rest of this chapter, the two parallel sets of monotonicity assumptions are considered.

Assumption I: $J_o(\omega) \leqq H(\omega, I, J_o)$ $\forall$ $\omega \in \Omega$, $I \in U(\omega)$.

Assumption I.1: Let $\{J_k\} \subset F$ be a sequence such that $J_o \leqq J_k \leqq J_{k+1}$ for all $k$, then

$$\lim_{k\to\infty} H(\omega, I, J_k) = H(\omega, I, \lim_{k\to\infty} J_k) \;\forall\; \omega \in \Omega,\; I \in U(\omega).$$

Assumption I.2: There exists a scalar $\alpha > 0$ such that for all scalars $r > 0$ and functions $J \in F$ with $J_o \leqq J$,

$$H(\omega,I,J) \leqq H(\omega,I,J+r) \leqq H(\omega,I,J) + \alpha r$$
$$\forall\ \omega \in \Omega \text{ and } I \in U(\omega).$$

Assumption D: $J_o(\omega) \geqq H(\omega,I,J_o)\ \forall\ \omega \in \Omega,\ I \in U(\omega)$

Assumption D.1: Let $\{J_k\} \subset F$ be a sequence such that $J_{k+1} \leqq J_k \leqq J_o$ for all $k$. Then

$$\lim_{k\to\infty} H(\omega,I,J_k) = H(\omega,I,\lim_{k\to\infty} J_k)\ \forall\ \omega \in \Omega \text{ and } I \in U(\omega).$$

Assumption D.2: There exists a scalar $\alpha > 0$ such that for all scalars $r > 0$ and functions $J \in F$ with $J \leqq J_o$

$$H(\omega,I,J) - \alpha r \leqq H(\omega,I,J-r) \leqq H(\omega,I,J)$$
$$\forall\ \omega \in \Omega \text{ and } J \in U(\omega)$$

Clearly, under either set of assumptions, $J_\pi$ is guaranteed to be well-defined by the monotonicity of $J$ for all $\pi \in \Pi$. It is also easy to see that under each of these sets of assumptions the limit, $\lim_{N\to\infty} (T_{I_o} T_{I_1} \cdots T_{I_{N-1}})(J_o)(\omega)$ is well-defined as a real number or $\pm\infty$. Indeed, in the case of Assumption I,

$$J_o \leqq T_{I_o}(J_o) \leqq (T_{I_o}T_{I_1})(J_o) \leqq \cdots$$

$$\leqq (T_{I_o}T_{I_1}\cdots T_{I_{N-1}})(J_o) \leqq \cdots, \quad \text{and}$$

$$J_o \geqq T_{I_o}(J_o) \geqq (T_{I_o}T_{I_1})(J_o) \geqq \cdots$$

$$\geqq (T_{I_o}T_{I_1}\cdots T_{I_{N-1}})(J_o) \geqq \cdots$$

in the case of Assumption D. In both cases, the limit clearly exists in the extended real numbers for each $\omega \in \Omega$.

Once again, as a first step, the function $H$ is shown to satisfy these assumptions.

<u>Theorem 5.7</u>: Consider the mapping

$$H(\omega,I,J) = E^*\{g(\omega,I,y) + \alpha J[f(\omega,I,y)] \mid \omega,I\}.$$

Let $J_o(\omega) = 0 \;\; \forall \;\; \omega \in \Omega$. If

$$g(\omega,I,y) \geqq 0 \quad \forall \quad \omega \in \Omega,\; I \in U(\omega),\; y \in \mathcal{Y}$$

then Assumptions I, I.1 and I.2 are satisfied with the scalar in I.2 equal to $\alpha$. If

$$g(\omega,I,y) \leqq 0 \quad \forall \quad \omega \in \Omega,\; I \in U(\omega),\; y \in \mathcal{Y}$$

then Assumptions D, D.1 and D.2 are satisfied with the scalar in D.2 equal to $\alpha$.

Proof: Since $J_o(\omega) = 0 \;\; \forall \; \omega \in \Omega$ and $g(\omega,I,y) \geqq 0$ or $g(\omega,I,y) \leqq 0 \;\; \forall \;\; \omega \in \Omega,\; I \in U(\omega),\; y \in \mathcal{Y}$

Assumptions I and D are trivially satisfied. By the monotone convergence theorem of integration, Assumptions I.1 and D.1 are satisfied since $g(\omega,I,y) \geq 0$.

For all $r > 0$ and $J \in F$ with $J_0 \leq J$

$$\begin{aligned} H(\omega,I,J+r) &= E^*\{g(\omega,I,y) + \alpha J[f(\omega,I,y)] + \alpha r | \omega,I\} \\ &= E^*\{g(\omega,I,y) + \alpha J[f(\omega,I,y)] | \omega,I\} + \alpha r \\ &= H(\omega,I,J) + \alpha r. \end{aligned}$$

Hence I.2 is satisfied. Similarly, using $g(\omega,I,y) \leq 0$ for all $r > 0$, $J \in F$, $J \leq J_0$

$$H(\omega,I,J-r) = H(\omega,I,J) - \alpha r$$

and D.2 is satisfied. QED

In proving the next theorem, the following, admittedly confusing notation is adopted.

Notation: The contract $\pi_k[\omega] = (I_0^k[\omega], I_1^k[\omega], \ldots)$ is associated with $\omega$. $I_i^k[\omega]$ denotes, for each $\omega \in \Omega$ and $k$ a function in $M$ while $I_i^k[\omega][Z]$ denotes the value of $I_i^k[\omega]$ at an element $Z \in \Omega$.

<u>Theorem 5.8</u>: Let Assumptions D, D.1 and D.2 hold. Let $J^* < \infty$ and $\epsilon > 0$ be given, there exist an $\epsilon$-optimal contract. Furthermore, if, in D.2, the scalar $\alpha < 1$, the contract $\pi_\epsilon$ is stationary.

Proof: Let $\{\epsilon_k\}$ be a sequence such that $\epsilon_k > 0$ for all $k$ and

$$\sum_{k=0}^{\infty} \alpha^k \epsilon_k = \epsilon.$$

For each $\omega \in \Omega$, let $\{\pi_k[\omega]\} \subset \Pi$ be a sequence of contracts of the form $\pi_k[\omega] = (I_0^k[\omega], I_1^k[\omega], \ldots)$ such that for $k = 0,1,\ldots$

$$J_{\pi_k[\omega]}(\omega) \geqq J^*(\omega) - \epsilon_k \quad \forall \quad \omega \in \Omega.$$

Since $J^* < \infty$, such a sequence exists. Let $\bar{I}_k \in M$ be defined by $\bar{I}_k(\omega) = I_0^k[\omega](\omega) \quad \forall \quad \omega \in \Omega$ and $\bar{J}_k$ defined by

$$\bar{J}_k(\omega) = H[\omega, \bar{I}_k(\omega), \lim_{i\to\infty}(T_{I_1^k[\omega]} \cdots T_{I_i^k[\omega]})(J_0)] \quad \forall \quad \omega \in \Omega,$$

$$k = 0,1,\cdots.$$

By D and D.1

$$\bar{J}_k(\omega) = \lim_{i\to\infty}(T_{I_0^k[\omega]} \cdots T_{I_i^k[\omega]})(J_0)(\omega) = J_{\pi_k[\omega]}(\omega)$$

$$\geqq J^*(\omega) - \epsilon_k \quad \forall \quad \omega \in \Omega, \quad k = 0,1,\cdots.$$

By D.2, for all $k = 1,2,\ldots$ and $\omega \in \Omega$

$$\begin{aligned} T_{\bar{I}_{k-1}}(\bar{J}_k)(\omega) &= H[\omega, \bar{I}_{k-1}(\omega), \bar{J}_k] \\ &\geqq H[\omega, \bar{I}_{k-1}(\omega), (J^* - \epsilon_k)] \\ &\geqq H[\omega, \bar{I}_{k-1}(\omega), J^*] - \alpha\epsilon_k \\ &\geqq H[\omega, \bar{I}_{k-1}(\omega), \lim(T_{I_1^{k-1}[\omega]} \cdots T_{I_i^{k-1}[\omega]})(J_0)] \\ &\quad - \alpha\epsilon_k = \bar{J}_{k-1}(\omega) - \alpha\epsilon_k. \end{aligned}$$

Then

$$T_{\bar{I}_{k-2}}[T_{\bar{I}_{k-1}}(\bar{J}_k)] \geqq T_{\bar{I}_{k-2}}(\bar{J}_{k-1} - \alpha\varepsilon_k)$$

$$\geqq T_{\bar{I}_{k-2}}(\bar{J}_{k-1} - \alpha^2\varepsilon_k)$$

$$\geqq \bar{J}_{k-2} - (\alpha\varepsilon_{k-1} + \alpha^2\varepsilon_k)$$

$$\therefore\ (T_{\bar{I}_o} \cdots T_{\bar{I}_{k-1}})(\bar{J}_k) \geqq \bar{J}_o - (\alpha\varepsilon_1 + \ldots + \alpha^k\varepsilon_k)$$

$$\geqq J^* - \sum_{i=0}^{k} \alpha^i\varepsilon_i .$$

But $J_o \geqq \bar{J}_k$

$$\therefore\ (T_{\bar{I}_o} \cdots T_{\bar{I}_k})(J_o) \geqq J^* - \sum_{i=0}^{k} \alpha^i\varepsilon_i .$$

Let $\pi_\varepsilon = (\bar{I}_o, \bar{I}_1, \ldots)$ and taking limits

$$J_{\pi_\varepsilon} \geqq J^* - \varepsilon .$$

If $\alpha < 1$, take $\varepsilon_k = \varepsilon(1-\alpha)$ for all $k$. Let $\pi_\varepsilon = (\bar{I}, \bar{I}, \ldots)$, where $\bar{I}(\omega) = I_o[\omega](\omega)$ for all $\omega \in \Omega$.

The $\pi_\varepsilon$ is an $\varepsilon$-optimal stationary policy. QED

## 5.3* <u>The Optimality Equation</u>

The next two theorems with their corollaries prove the optimality equation, $J^* = T(J^*)$, hereby establishing the validity of the dynamic programming algorithm. The corollaries attempt to set up the algorithm for stationary contracts. For Assumption D, the optimality equation requires the compliance of all D, D.1 and D.2, whereas under Assumption I, the same results hold only under I and one of the additional conditions, I.1 or I.2.

<u>Theorem 5.9</u>: Let D, D.1 and D.2 hold. Then $J^* = T(J^*)$. Furthermore, if $J' \in F$ is such that $J' \leqq J_o$ and $J' \leqq T(J')$, then $J' \leqq J^*$.

Proof: For every $\pi = (I_o, I_1, \ldots) \in \Pi$ and $\omega \in \Omega$. By D.1

$$J_\pi(\omega) = \lim_{k\to\infty} (T_{I_o} T_{I_1} \ldots T_{I_k})(J_o)(\omega)$$

$$= T_{I_o}[\lim_{k\to\infty} (T_{I_1} \ldots T_{I_k})(J_o)](\omega)$$

$$\leqq T_{I_o}(J^*)(\omega) \leqq T(J^*)(\omega).$$

Taking supremum of the left-hand side over $\pi \in \Pi$

$$J^* \leqq T(J^*).$$

Let $\epsilon_1$ and $\epsilon_2 > 0$. By Theorem 5.2, $\exists$ a contract $\bar{\pi} = (\bar{I}_o, \bar{I}_1, \ldots)$ such that

$$T_{\bar{I}_o}(J^*) \geqq T(J^*) - \epsilon_1$$

and

$$J_{\bar{\pi}_1} > J^* - \epsilon_2$$

where

$$\bar{\pi}_1 = (\bar{I}_1, \bar{I}_2, \ldots)$$

$$J_{\bar{\pi}} = \lim_{k\to\infty} (T_{\bar{I}_o} T_{\bar{I}_1} \ldots T_{\bar{I}_k})(J_o)$$

$$= T_{\bar{I}_o}[\lim_{k\to\infty} (T_{\bar{I}_1} \ldots T_{\bar{I}_k})(J_o)]$$

$$= T_{\bar{I}_o}(J_{\bar{\pi}_1}) \geqq T_{\bar{I}_o}(J^*) - \alpha\epsilon_2$$

$$\geqq T(J^*) - (\epsilon_1 + \alpha\epsilon_2).$$

But $J^* \geqq J_{\bar{\pi}}$ and $\epsilon_1$ and $\epsilon_2$ are arbitrary. Then

$$J^* \geqq T(J^*)$$

$$\therefore\ J^* = T(J^*).$$

Let $J' \in F$ be such that $J' \leqq J_o$ and $J' \leqq T(J')$.
Let $\{\epsilon_k\}$ and a sequence with $\epsilon_k > 0$ and a contract $\bar{\pi} = (\bar{I}_o, \bar{I}_1, \ldots) \in \Pi$ be such that

$$T_{\bar{I}_k}(J') \geqq T(J') - \epsilon_k \qquad k = 0,1,\cdots.$$

From D.2

$$J^* = \sup_{\pi\in\Pi} \lim_{k\to\infty} (T_{I_o} \ldots T_{I_k})(J_o)$$

$$\geqq \sup_{\pi\in\Pi} \limsup_{k\to\infty} (T_{I_o} \ldots T_{I_k})(J')$$

$$\geqq \limsup_{k\to\infty} (T_{I_o} \ldots T_{I_k})(J')$$

$$\geqq \limsup_{k\to\infty} (T_{I_o} \ldots T_{I_{k-1}})[T(J') - \epsilon_k]$$

$$\geqq \limsup_{k\to\infty} (T_{I_o} \ldots T_{I_{k-1}})(J' - \epsilon_k)$$

$$\geqq \limsup_{k\to\infty} (T_{I_o} \ldots T_{I_{k-1}})(J') - \alpha^k\epsilon_k$$

$$\geqq \lim_{k\to\infty} [T(J') - (\sum_{i=0}^{k} \alpha^i\epsilon_i)] \geqq J' - \sum_{i=0}^{k} \alpha^i\epsilon_i$$

since $\epsilon_i$ are arbitrary

$\therefore\ J^* \geqq J'$. QED

Corollary 5.9.1: Let D, D.1 and D.2 hold. Then for every stationary contract

$$\pi = (I, I, \ldots), \quad J_I = T_I(J_1).$$

Furthermore, if $J' \in F$ is such that $J' \leqq J_o$ and $J' \leqq T_I(J')$ then $J' \leqq J_I$.

Proof: Use $U(\omega) = \{I(\omega)\}$ instead of $U(\omega)\ \forall\ \omega \in \Omega$. Use Theorem 5.9. QED

Lemma 5.1: Let I hold. Then $J^* = \lim_{N\to\infty} J_N^*$.

Proof: Clearly $J^* \geqq J_N^*$ for all N. Hence $J^* \geqq \lim_{N\to\infty} J_N^*$. Also for all $\pi = (I_o, I_1, \ldots) \in \Pi$,

$$(T_{I_o} \ldots T_{I_{N-1}})(J_o) \leqq J_N^*.$$

Taking limits on both sides

$$J_\pi \leqq \lim_{N\to\infty} J_N^*.$$

Taking supremum of the left hand side

$$J^* \leqq \lim_{N\to\infty} J_N^*$$

$\therefore\ J^* = \lim_{N\to\infty} J_N^*$. QED

Theorem 5.10: Let I and I.1 hold. Then $J^* = T(J^*)$. Furthermore, if $J' \geqq J_o$ and $J' \geqq T(J')$, then $J' \geqq J^*$.

Proof: Using the arguments in Lemma 5.1 for all $\omega \in \Omega$

$$\lim_{N\to\infty} \sup_{\pi \in U(\omega)} H(\omega, I, J_N^*) = \sup_{\pi \in U(\omega)} \lim_{N\to\infty} H(\omega, I, J_N^*).$$

By I.1, then

$$\lim_{N\to\infty} T(J_N^*) = T(\lim_{N\to\infty} J_N^*).$$

Since I and I.1 are equivalent to Assumption F.1′, by Corollary 4.2.1,

$$J_N^* = T^N(J_o).$$

Thus $T(J_N^*) = T^{N+1}(J_o) = J_{N+1}^*$. By Lemma 5.1 and combining the results we have

$$J^* = T(J^*).$$

Let $J' \in F$ be such that $J' \geqq J_o$ and $J' \geqq T(J')$. Then

$$J = \sup_{\pi\in\Pi} \lim_{N\to\infty} (T_{I_o} \cdots T_{I_{N-1}})(J_o)$$

$$\leqq \lim_{N\to\infty} \sup_{\pi\in\Pi} (T_{I_o} \cdots T_{I_{N-1}})(J_o)$$

$$\leqq \lim_{N\to\infty} \sup_{\pi\in\Pi} (T_{I_o} \cdots T_{I_{N-1}})(J')$$

$$\leqq \lim_{N\to\infty} T^N(J') \leqq J'. \quad \text{QED}$$

<u>Corollary 5.10.1</u>: Let I. and I.2 hold. Let $\Omega$ be a finite and $J^*(\omega) < \infty$ for all $\omega \in \Omega$. Then $J^* = T(J^*)$. Furthermore, if $J' \in F$ is such that $J' \geqq J_o$ and $J' \geqq T(J')$, then $J' \geqq J^*$.

Proof: Using a nearly verbation repetition of the proof of Theorem 4.2 (b), we have $J_N^* = T^N(J_o)$ for all $N = 1,2,\cdots$. We will now show that

$$\lim_{N\to\infty} H(\omega,I,J_N^*) = H(\omega,I,\lim_{N\to\infty} J_N^*) \quad \forall\ \omega \in \Omega,\ I \in U(\omega).$$

Suppose for some $\tilde{\omega} \in \Omega$, $\tilde{I} \in U(\tilde{\omega})$ and $\epsilon > 0$

$$H(\tilde{\omega},\tilde{I},J_k^*) + \epsilon < H(\tilde{\omega},\tilde{I},\lim_{N\to\infty} J_N^*) \quad k = 1,2,\cdots.$$

Since $\Omega$ is finite and $J^*(\omega) = \lim_{N\to\infty} J_N^*(\omega) < \infty$ for all $\omega$, $\exists$ an integer $k_o > 0$ such that

$$J_k^* + (\epsilon/\alpha) \geqq \lim_{N\to\infty} J_N^* \quad \forall\ k \geqq k_o$$

By I.2, for all $k \geqq k_o$

$$H(\tilde{\omega},\tilde{I},J_k^*) + \epsilon \geqq H(\tilde{\omega},\tilde{I},J_k^* + (\epsilon/\alpha)) \geqq H(\tilde{\omega},\tilde{I},\lim_{N\to\infty} J_N^*)$$

which contradicts the earlier inequality

$$\therefore \lim_{N\to\infty} H(\omega,I,J_N^*) = H(\omega,I,\lim_{N\to\infty} J_N^*)$$

and the results follow by Theorem 5.10. QED

<u>Corollary 5.10.2</u>: Let I and I.1 hold. Then for every stationary contract $\pi = (I,I,\ldots)$, $J_I = T_I(J_I)$. Furthermore, if $J' \in F$ is such that $J' \geqq J_o$ and $J' \geqq T_I(J')$, then $J' \geqq J_I$.

Next, necessary and sufficient conditions for the optimality of a stationary contract under the two assumptions are studied.

<u>Theorem 5.11</u>: Let D, D.1 and D.2. Then a stationary contract $\pi^* = (I^*,I^*,\ldots)$ is optimal if and only if

$$T_{I^*}(J^*) = T(J^*).$$

Furthermore, if for each $\omega \in \Omega$, there exists a contract which is optimal at $\omega$, then there exists a stationary optimal contract.

Proof: If $\pi^*$ is optimal, then $J_{I^*} = J^*$. By Theorem 5.9 and Corollary 5.9.1, the result follows. Conversely, if $T_{I^*}(J^*) = T(J^*)$, By Theorem 5.9, $J^* = T(J^*)$ then it follows that $T_{I^*}(J^*) = J^*$. By Corollary 5.9.1, $J_{I^*} \geqq J^*$.

$\therefore$ $\pi^*$ is optimal.

Let $\pi^*_\omega = (I^*_{o,\omega}, I^*_{1,\omega}, \ldots)$ be optimal at $\omega \in \Omega$. By D.1,

$$\begin{aligned} J^*(\omega) &= J_{I^*,\omega}(\omega) \\ &= \lim_{k\to\infty} (T_{I^*_{o,\omega}} \cdots T_{I^*_{k,\omega}})(J_o)(\omega) \\ &= T_{I^*_{o,\omega}} [\lim_{k\to\infty} (T_{I^*_{1,\omega}} \cdots T_{I^*_{k,\omega}})(J_o)](\omega) \\ &\leqq T_{I^*_{o,\omega}} (J^*)(\omega) \leqq T(J^*)(\omega) = J^*(\omega) \end{aligned}$$

$\therefore$ $T_{I^*_{o,\omega}}(J^*)(\omega) = T(J^*)(\omega)$ for all $\omega \in \Omega$.

Define $I^* \in M$ by $I^*(\omega) = I^*_{o,\omega}(\omega)$. Then $T_{I^*}(J^*) = T(J^*)$ and by result just proved $(I^*, I^*, \ldots)$ is optimal. QED

<u>Theorem 5.12</u>: Let I and I.1 hold. Then a stationary contract $\pi^* = (I^*, I^*, \ldots)$ is optimal if and only if

$$T_{I^*}(J_{I^*}).$$

Proof: If $\pi^*$ is optimal, then $J_{I^*} = J^*$. By Theorem 5.10 and Corollary 5.10.2, the result follows.

Conversely, if $T_{I*}(J_{I*}) = T(J_{I*})$. By Corollary 5.10.2, $J_{I*} = T(J_{I*})$. By Theorem 5.10, $J_{I*} \geqq J*$

$\therefore$ $\pi*$ is optimal. QED

Theorem 5.11 states that under Assumption D, if the supremum is attain in the optimality equation

$$J*(\omega) = \sup_{I \in U(\omega)} H(\omega, I, J*)$$

for every $\omega \in \Omega$, then there exists a stationary contract. However, if the supremum cannot be attained for some $\omega \in \Omega$, the optimality equation can still be used to construct a nearly optimal contract, which is stationary whenever the scalar $\alpha$ in D.2 is strictly less than one.

<u>Theorem 5.13</u>: Let D, D.1 and D.2 hold. Then

(a) Let $\epsilon > 0$, and $\{\epsilon_i\}$ be such that $\sum_{k=0}^{\infty} \alpha^k \epsilon_k = \epsilon$. $\epsilon_i > 0$, $i = 0, 1, \cdots$. Let $\pi* = (I_0^*, I_1^*, \ldots) \in \Pi$ be such that

$$T_{I_k^*}(J*) \geqq T(J*) - \epsilon_k \quad k = 0, 1, \ldots$$

then $J* \geqq J_{I*} \geqq J* - \epsilon$.

(b) Let $\epsilon > 0$ and the scalar in D.2, $\alpha < 1$. Suppose $I* \in M$ is such that $T_{I*}(J*) \geqq T(J*) - \epsilon(1-\alpha)$. Then $J* \geqq J_{I*} \geqq J* - \epsilon$.

Proof: (a) Since $T(J^*) = J^*$

$$T_{I_k^*} \geqq J^* - \epsilon_k$$

Apply $T_{I_{k-1}^*}$ to both sides

$$(T_{I_{k-1}^*} T_{I_k^*})(J^*) \geqq T_{I_{k-1}^*}(J^*) - \alpha\epsilon_k$$

$$\geqq J^* - (\epsilon_k + \alpha\epsilon_k).$$

Repeat the process, for every $k = 1,2,...$

$$(T_{I_o^*} \cdots T_{I_k^*})(J^*) \geqq J^* - (\sum_{i=0}^{k} \alpha^i \epsilon_i).$$

Since $J_o \geqq J^*$, it follows that

$$(T_{I_o^*} \cdots T_{I_k^*})(J_o) \geqq J^* - \sum_{i=0}^{k} \alpha^i \epsilon_i \qquad k = 1,2,\cdots.$$

Taking limit as $k \to \infty$

$$J_{I^*} \geqq J^* - \epsilon.$$

(b) Let $\epsilon_k = \epsilon(1-\alpha)$ and $I_k^* = I^*$ for all $k$. The result follows by part (a). QED

A weak counterpart of part (a) of Theorem 5.13 under Assumption I is given in Theorem 5.15. However, I am unable to give a counterpart of part (b) under Assumption I or conditions for existence of a stationary optimal contract.

## 5.4* Convergence to Optimality and Existence of Optimal Contracts

Define a function $J_\infty \in F$ by

$$J_\infty(\omega) = \lim_{N\to\infty} T^N(J_0)(\omega) \quad \text{for every} \quad \omega \in \Omega.$$

This section is devoted to investigate whether $J_\infty = J^*$. Fortunately, the relationship hold for Assumption I under very mild conditions.

Theorem 5.14: Let I hold and assume that either I.1 holds or $J_N^* = T^N(J_0)$ for all N. Then $J_\infty = J^*$ where

$$J_\infty(\omega) = \lim_{N\to\infty} T^N(J_0)(\omega) \quad \forall \quad \omega \in \Omega.$$

Proof: By Lemma 6, $J^* = \lim_{N\to\infty} J_N^*$. By Corollary 4.2.1 $J_N^* = T^N(J_0)$

$$\therefore\ J^* = \lim_{N\to\infty} J_N^* = J_\infty. \quad \text{QED}$$

The following is a counterpart of Theorem 5.8 and part (a) of Theorem 5.13 under Assumption I for the existence of nearly optimal contracts.

Theorem 5.15: Let I and I.2 hold. Let $\Omega$ be a finite set and $J^*(\omega) < \infty$ for all $\omega \in \Omega$. Then for any $\varepsilon > 0$, there exists an $\varepsilon$-optimal policy.

Proof: For each $N$, let $\epsilon_N = \epsilon/2(1+\alpha+\ldots+\alpha^{N-1})$ and $\pi_N = \{I_o^N, I_1^n, \ldots, I_{N-1}^N, I, I, \ldots\}$ be such that $I \in M$ and for $k = 0,1,\ldots,N-1$, $I_k^N \in M$ and

$$(T_{I_k^N} T^{N-k-1})(J_o) \geqq T^{N-k}(J_o) - \epsilon_N.$$

Thus $T_{I_{N-1}^N}(J_o) \geqq T(J_o) - \epsilon_N$. Apply $T_{I_{N-2}^N}$ to both sides

$$(T_{I_{N-2}^N} T_{I_{N-1}^N})(J_o) \geqq (T_{I_{N-2}^N} T)(J_o) - \alpha\epsilon_N$$

$$\geqq T^2(J_o) - (1+\alpha)\epsilon_N.$$

Continuing $(T_{I_o^N} \ldots T_{I_{N-1}^N})(J_o) \geqq T^N(J_o) - (1+\alpha+\ldots+\alpha^{N-1})\epsilon_N$.

$\therefore$ For $N = 0,1,\ldots$

$$J_{\pi_N} \geqq T^N(J_o) - \epsilon/2.$$

As in the proof of Corollary 5.10.1, the assumptions imply

$$J_N^* = T^N(J_o) \quad \text{for all } N.$$

By Theorem 5.14, $\lim_{N\to\infty} T^N(J_o) = J_N^*$. Since $\Omega$ is finite and $J^*(\omega) < \infty$ for all $\omega \in \Omega$ $\exists$ $N_o$ such that $T^{N_o}(J_o) \geqq J^* - \epsilon/2$. Then

$$J_{\pi_{N_o}} > J^* - \epsilon$$

and $\pi_{N_o}$ is the desired contract. QED

Under Assumption D, D.1 and D.2, the equality $J_\infty = J^*$ may fail to hold even in very simple situations. The following preliminary result shows that in order to have $J_\infty = J^*$, it is necessary and sufficient to have $J_\infty = T(J_\infty)$, a condition implying the convergence of the dynamic programming algorithm.

Theorem 5.16: Let D, D.1 and D.2 hold. Then

$$J_\infty \geqq T(J_\infty) \geqq T(J^*) = J^*.$$

Furthermore, $J_\infty = T(J_\infty) = T(J^*) = J^*$ if and only if $J_\infty = T(J_\infty)$.

Proof: Clearly $J_\infty \geqq J_\pi$ for all $\pi \in \Pi$

$\therefore\ J_\infty \geqq J^*$.

By Theorem 5.9 $T(J^*) = J^*$. For all $k \geqq 0$

$$T(J_\infty) = \sup_{I \in U(\omega)} H(\omega, I, J_\infty)$$

$$\leqq \sup_{I \in U(\omega)} H(\omega, I, T^k(J_o)] = T^{k+1}(J_o).$$

Taking limit on right side, $T(J_\infty) \leqq J_\infty$

$\therefore\ J_\infty \geqq T(J_\infty) \geqq T(J^*) = J^*$.

Let $J_\infty = T(J_\infty) = T(J^*) = J^*$

$$J_\infty = T(J_\infty) \text{ by hypothesis.}$$

Let $J_\infty = T(J_\infty)$. Since $J_o \geqq J_\infty$, by Theorem 5.9, $J_\infty \leqq J^*$. Then $J_\infty \geqq T(J_\infty) \geqq T(J^*) = J^*$ implies that $J_\infty = T(J_\infty) = T(J^*) = J^*$. QED

To prove the fact that $J_\infty = T(J_\infty)$, the following definitions and notations are needed.

Notations:

(1) For $J \in F$, let $E(J)$ denote the epigraph of $J$, i.e., the subset of $\Omega \times \mathbb{R}$ given by

$$E(J) = \{(\omega,\lambda) | J(\omega) \geqq \lambda\}.$$

Under D, since $T^k(J_o) \geqq T^{k+1}(J_o)$ for all $k$ and $J_\infty = \lim_{k\to\infty} T^k(J_o)$ thus

$$E(J_\infty) = \bigcap_{k=1}^{\infty} E[T^k(J_o)].$$

(2) For each $k \geqq 1$, the subset $C_k$ of $\Omega \times C \times \mathbb{R}$ is given by

$$C_k = \{(\omega,I,\lambda) | H[\omega,I,T^{k-1}(J_o)] \geqq \lambda,\ \omega \in \Omega,\ I \in U(\omega)\}.$$

(3) Let $P(C_k)$ denote the projection of $C_k$ on $\Omega \times \mathbb{R}$, i.e.,

$$P(C_k) = \{(\omega,\lambda) | \exists\ I \in U(\omega) \text{ such that } (\omega,I,\lambda) \in C_k\}.$$

(4) Let the set $\overline{P(C_k)}$ be defined as

$$\overline{P(C_k)} = \{(\omega,\lambda) | \exists\ \{\lambda_n\} \text{ such that } \lambda_n \to \lambda,\ (\omega,\lambda_n) \in P(C_k)\quad n = 0,1,\ldots\}.$$

<u>Lemma 5.2</u>: Let D hold. Then for all $k \geqq 1$

$$P(C_k) \subset \overline{P(C_k)} = E[T^k(J_o)].$$

Furthermore, $P(C_k) = \overline{P(C_k)} = E[T^k(J_o)]$ if and only if the supremum is attained for each $\omega \in \Omega$ in

$$T^k(J_o)(\omega) = \sup_{I \in U(\omega)} H[\omega, I, T^{k-1}(J_o)].$$

Proof: If $(\omega, \lambda) \in E[T^k(J_o)]$, then

$$T^k(J_o)(\omega) = \sup_{I \in U(\omega)} H[\omega, I, T^{k-1}(J_o)] \geqq \lambda.$$

Let $\{\epsilon_n\}$ be a sequence such that $\epsilon_n > 0$, $\epsilon_n \to 0$ and let $\{I_n\} \subset U(\omega)$ be such that

$$H[\omega, I_n, T^{k-1}(J_o)] \geqq T^k(J_o)(\omega) - \epsilon_n \geqq \lambda - \epsilon_n.$$

Then $(\omega, I_n, \lambda - \epsilon_n) \in C_k$ and $(\omega, \lambda - \epsilon_n) \in P(C_k)$ for all $n$. Since $\lambda - \epsilon_n \to \lambda$, $(\omega, \lambda) \in \overline{P(C_k)}$

$$\therefore E[T^k(J_o) \subset \overline{P(C_k)}.$$

Let $(\omega, \lambda) \in \overline{P(C_k)}$ $\exists$ a sequence $\{\lambda_n\}$ such that $\lambda_n \to \lambda$ and a corresponding sequence $\{I_n\} \subset U(\omega)$ such that

$$T^k(J_o)(\omega) \geqq H[\omega, I_n, T^{k-1}(J_o)] \geqq \lambda_n.$$

Let $n \to \infty$, $T^k(J_o)(\omega) \geqq \lambda$. Thus

$$(\omega, \lambda) \in E[T^k(J_o)]$$

$$\overline{P(C_k)} \subset E[T^k(J_o)]$$

$$\therefore P(C_k) \subset \overline{P(C_k)} = E[T^k(J_o)].$$

Assume that the supremum is attained by $I^*_{k-1}(\omega)$ for each $\omega \in \Omega$. Then for each $(\omega,\lambda) \in E[T^k(J_o)]$

$$H[\omega, I^*_{k-1}(\omega), T^{k-1}(J_o)] \geqq \lambda.$$

This implies $(\omega,\lambda) \in P(C_k)$. Hence $E[T^k(J_o)] \subset P(C_k)$.

By the first part of this theorem

$$P(C_k) = \overline{P(C_k)} = E[T^k(J_o)].$$

Now let $P(C_k) = \overline{P(C_k)} = E[T^k(J_o)]$. For every $\omega$ for which $T^k(J_o)(\omega) > -\infty$

$$[\omega, T^k(J_o)(\omega)] \in P(C_k).$$

This implies that $\exists$ a $I^*_{k-1}(\omega) \in U(\omega)$ such that

$$H[\omega, I^*_{k-1}(\omega), T^{k-1}(J_o)] \geqq T^k(J_o)(\omega)$$

$$= \sup_{I \in U(\omega)} H[\omega, I, T^{k-1}(J_o)]$$

$\therefore$ The supremum is attained for all $\omega$ for which $T^k(J_o)(\omega) > -\infty$.

It also is trivially attained by all $I \in U(\omega)$ whenever $T^k(J_o)(\omega) = -\infty$. QED

Definition:

$$P(\bigcap_{k=1}^{\infty} C_k) = \{(\omega,\lambda) \mid \exists\ I \in U(\omega) \text{ such that } (\omega, I, \lambda) \in \bigcap_{k=1}^{\infty} C_k\}$$

$$\overline{P(\bigcap_{k=1}^{\infty} C_k)} = \{(\omega,\lambda) \mid \exists\ \{\lambda_n\} \text{ such that } \lambda_n \to \lambda,\ (\omega,\lambda_n) \in P(\bigcap_{k=1}^{\infty} C_k)\}.$$

Lemma 5.3:

$$P(\bigcap_{k=1}^{\infty} C_k) \subset \bigcap_{k=1}^{\infty} P(C_k) \subset \bigcap_{k=1}^{\infty} \overline{P(C_k)}$$

$$= \bigcap_{k=1}^{\infty} E[T^k(J_o) = E(J_\infty), \quad \text{and}$$

$$\overline{P(\bigcap_{k=1}^{\infty} C_k)} \subset \bigcap_{k=1}^{\infty} P(C_k) = \bigcap_{k=1}^{\infty} E[T^k(J_o)] = E(J_\infty).$$

Theorem 5.17: Let D, D.1 and D.2 hold. Then

(a) $J_\infty = T(J_\infty)$ (equivalently $J_\infty = J^*$) if and only if

$$\overline{P(\bigcap_{k=1}^{\infty} C_k)} = \bigcap_{k=1}^{\infty} \overline{P(C_k)}.$$

(b) $J_\infty = T(J_\infty)$ and the supremum in

$$J_\infty(\omega) = \sup_{I \in U(\omega)} H(\omega, I, J_o)$$

is attained for each $\omega \in \Omega$ if and only if

$$P(\bigcap_{k=1}^{\infty} C_k) = \bigcap_{k=1}^{\infty} \overline{P(C_k)}.$$

Proof: (a) Let $J_\infty = T(J_\infty)$ and $(\omega, \lambda) \in E(J_\infty)$. Thus $J_\infty(\omega) = \sup_{I \in U(\omega)} H(\omega, I, J_\infty) \geqq \lambda$. Let $\{\epsilon_n\}$ be a sequence such that $\epsilon_n > 0$, $\epsilon_n \to 0$ $\exists$ a sequence $\{I_n\}$ such that

$$H(\omega, I_n, J_\infty) \geqq \lambda - \epsilon_n \qquad n = 1, 2, \ldots$$

so $H[\omega, I_n, T^{k-1}(J_o)] \geqq \lambda - \epsilon_n \qquad k, n = 1, 2, \ldots$

$\therefore$ $(\omega, I_n, \lambda - \epsilon_n) \in C_k$ for all $k, n$

and $(\omega, I_n, \lambda - \epsilon_n) \in \bigcap_{k=1}^{\infty} C_k$ for all $n$. Thus

$(\omega, \lambda - \epsilon_n) \in P(\bigcap_{k=1}^{\infty} C_k)$ for all $n$. But $\lambda - \epsilon_n \to \lambda$

implies

$$(\omega, \lambda) \in \overline{P(\bigcap_{k=1}^{\infty} C_k)}$$

$$\therefore E(J_\infty) \subset \overline{P(\bigcap_{k=1}^{\infty} C_k)}.$$

By Lemma 5.3

$$\therefore \bigcap_{k=1}^{\infty} \overline{P(C_k)} = E(J_\infty) = \overline{P(\bigcap_{k=1}^{\infty} C_k)}.$$

Let $\overline{P(\bigcap_{k=1}^{\infty} C_k)} = \bigcap_{k=1}^{\infty} \overline{P(C_k)}$. By Lemma 5.2,

$$\overline{P(\bigcap_{k=1}^{\infty} C_k)} = E(J_\infty).$$

Let $\omega \in \Omega$ be such that $J_\infty(\omega) > -\infty$. Then

$[\omega, J_\infty(\omega)] \in \overline{P(\bigcap_{k=1}^{\infty} C_k)}$ $\exists$ a sequence $\{\lambda_n\}$ with

$\lambda_n \to J_\infty(\omega)$ and a sequence $\{I_n\} \subset U(\omega)$ such that

$$H[\omega, I_n, T^{k-1}(J_o)] \geqq \lambda_n \quad k,n = 1,2,\cdots.$$

By D.1, taking limit with respect to $k$

$$H[\omega, I_n, J_\infty] \geqq \lambda_n \quad n = 1,2,\cdots.$$

Thus

$$T(J_\infty)(\omega) \geqq H[\omega, I_n, J_\infty] \geqq \lambda_n.$$

Let $n \to \infty$ implies $T(J_\infty)(\omega) \geqq J_\infty(\omega)$ for all $\omega \in \Omega$ and $J_\infty(\omega) > -\infty$. The inequality also holds if $J_\infty(\omega) = -\infty$

$$\therefore\ T(J_\infty) \geqq J_\infty$$

By Theorem 5.16, $J_\infty \geqq T(J_\infty)$

$$\therefore\ J_\infty = T(J_\infty).$$

(b) Let $J_\infty = T(J_\infty)$ and the supremum is attained for each $\omega \in \Omega$ in $J_\infty(\omega) = \sup_{I \in U(\omega)} H(\omega, I, J_\infty)$ $\exists$ $I^* \in M$ such that for each $(\omega,\lambda) \in E(J_\infty)$

$$H[\omega, I^*(\omega), J_\infty] \geqq \lambda.$$

Thus $H[\omega, I^*(\omega), T^{k-1}(J_o)] \geqq \lambda$ for $k = 1,2,\ldots$ and

$$[\omega, I^*(\omega), \lambda] \in \bigcap_{k=1}^{\infty} C_k$$

$$\therefore\ (\omega,\lambda) \in P(\bigcap_{k=1}^{\infty} C_k)$$

$$\therefore\ E(J_\infty) \subset P(\bigcap_{k=1}^{\infty} C_k).$$

By Lemma 5.3, $P(\bigcap_{k=1}^{\infty} C_k) = E(J_\infty) = \bigcap_{k=1}^{\infty} \overline{P(C_k)}$. Conversely, let $P(\bigcap_{k=1}^{\infty} C_k) = E(J_\infty)$. For all $\omega \in \Omega$ with $J_\infty(\omega) > -\infty$

$$[\omega, J_\infty(\omega)] \in E(J_\infty) = P(\bigcap_{k=1}^{\infty} C_k)$$

$\exists$ a $I^*(\omega) \in U(\omega)$ such that

$$[\omega, I^*(\omega), J_\infty(\omega)] \in \bigcap_{k=1}^{\omega} C_k.$$

Thus

$$H[\omega, I^*(\omega), T^{k-1}(J_o)] \geqq J_\infty(\omega) \qquad k = 0,1,\cdots.$$

By D.1 and taking limits

$$T(J_\infty)(\omega) \geqq H[\omega, I^*(\omega), J_\infty] \geqq J_\infty(\omega).$$

By Theorem 5.16, $T(J_\infty) = J_\infty$. If $J_\infty(\omega) = -\infty$, every $I \in U(\omega)$ attains the supremum and the proof is complete. QED

<u>Theorem 5.18</u>: Let D, D.1 and D.2 hold. Let the incentive space C be a Hausdorff space. Suppose there exists an integer $k_o \geqq 0$ such that for each $\omega \in \Omega$, $\lambda \in R$ and $k \geqq k_o$, the set

$$U_k(\omega,\lambda) = \{I \in U(\omega) | H[\omega, I, T^k(J_o)] \geqq \lambda$$

is compact. Then

$$P(\bigcap_{k=1}^{\infty} C_k) = \bigcap_{k=1}^{\infty} \overline{P(C_k)}.$$

Proof: Let $(\omega,\lambda)$ be in $\bigcap_{k=1}^{\infty} P(C_k)$ $\exists$ a sequence $\{I_n\} \subset U(\omega)$ such that

$$H[\omega, I_n, T^k(J_o)] \geqq H[\omega, I_n, T^n(J_o)] \geqq \lambda \quad \forall \quad n \geqq k.$$

Thus $I_n \in U_k(\omega,\lambda)$ $\forall$ $n \geqq k$, $k = 0,1,\cdots$. $U_k(\omega,\lambda)$ is compact for $k \geqq k_o$. This implies that $\{I_n\}$ has a limit point $\bar{I} \in U_k(\omega,\lambda)$ $\forall$ $k \geqq k_o$. But $U_o(\omega,\lambda) \supset U_1(\omega,\lambda) \supset \ldots$ so $\bar{I} \in U_k(\omega,\lambda)$ for $k = 0,1,\ldots$

$$H[\omega,\bar{I},T^k(J_o)] \geqq \lambda \quad k = 0,1,\ldots$$

$$(\omega,\bar{I},\lambda) \in \bigcap_{k=1}^{\infty} C_k.$$

This implies $(\omega,\lambda) \in P(\bigcap_{k=1}^{\infty} C_k)$

$$\therefore P(\bigcap_{k=1}^{\infty} C_k) \supset \bigcap_{k=1}^{\infty} P(C_k).$$

Since $U_k(\omega,\lambda)$ is compact, by Lemma 4.5, the supremum in $T^k(J_o)(\omega) = \sup_{I \in U(\omega)} H[\omega,I,T^{k-1}(J_o)]$ is attained for every $\omega \in \Omega$ and $k > k_o$. By Lemma 5.2, $P(C_k) = \overline{P(C_k)}$ for $k > k_o$. But $P(C_1) \supset P(C_2) \supset \ldots$ and $\overline{P(C_1)} \supset \overline{P(C_2)} \supset \ldots$

$$\therefore \bigcap_{k=1}^{\infty} P(C_k) = \bigcap_{k=1}^{\infty} \overline{P(C_k)}.$$

By Lemma 5.2,

$$P(\bigcap_{k=1}^{\infty} C_k) = \bigcap_{k=1}^{\infty} P(C_k). \quad \text{QED}$$

After proving the fact that $J_\infty = T(J_\infty)$ and hence establishing the convergence of the dynamic programming algorithm under Assumption D, the following provides the conditions for the existence and computation of optimal stationary contracts under the decreasing assumption .

Theorem 5.19: Let the assumptions of Theorem 5.18 hold. Then

(a) there exists a contract $\pi^* = (I_o^*, I_1^*, \ldots) \in \Pi$ attaining the supremum for all $k \geq k_o$, i.e.,

$$(T_{I_k^*} T^k)(J_o) = T^{k+1}(J_o) \quad \forall \quad k \geq k_o.$$

(b) For every contract $\pi^*$ satisfying (a), the sequence $\{I_k^*(\omega)\}$ has at least one limit point for each $\omega \in \Omega$ with $J^*(\omega) > -\infty$.

(c) Let $I^* : \Omega \to C$ be such that $I^*(\omega)$ is a limit point of $\{I_k^*(\omega)\}$ for all $\omega \in \Omega$ with $J^*(\omega) > -\infty$ and $I^*(\omega) \in U(\omega)$ for all $\omega \in \Omega$ with $J^*(\omega) = -\infty$. Then the stationary contract $(I^*, I^*, \ldots)$ is optimal.

Proof: (a) This follows from Lemma 4.5.

(b) Let $\pi^* = (I_o^*, I_1^*, \ldots)$ satisfy

$$(T_{I_k^*} T^k)(J_o) = T^{k+1}(J_o) \quad \forall\ k \geq k_o,\ \omega \in \Omega$$

and $J^*(\omega) > -\infty$

$$H[\omega, I_n^*(\omega), T^k(J_o)] \geq H[\omega, I_n^*(\omega), T^n(J_o)]$$

$$\geq J^*(\omega) \quad \forall\ k \geq k_o, \quad n \geq k$$

$\therefore\ I_n^*(\omega) \in U_k[\omega, J^*(\omega)] \quad \forall\ k \geq k_o, \quad n \geq k.$

$U_k[\omega, J^*(\omega)]$ is compact. $\{I_n^*(\omega)\}$ has at least one limit point.

(c) Each limit point $I^*(\omega) \subset U(\omega)$ and

$$H[\omega, I^*(\omega), T^k(J_o)] \geqq J^*(\omega) \quad \forall\ k \geqq k_o.$$

Using D.1 and taking limits

$$H[\omega, I^*(\omega), J_\infty] = H[\omega, I^*(\omega), J^*] \geqq J^* \quad \text{for all} \quad \omega \in \Omega.$$

This relation holds trivially for all $\omega \in \Omega$ with $J^*(\omega) = -\infty$.

$$\therefore\ T_{I^*}(J^*) \geqq J^* = T(J^*).$$

This implies $T_{I^*}(J^*) = T(J^*)$. By Theorem 5.11 $(I^*, I^*, \ldots)$ is optimal. QED.

# CHAPTER VI

# BOREL MODELS

## 6.1: Introduction

In the previous chapters, a basic multiperiod agency model is developed. It was shown that under appropriate conditions, optimal or nearly optimal contracts exist and the dynamic programming algorithm can be implemented to construct such contracts. All these results rely on the assumption that the disturbance term, $y_k$, behaves in a reasonable manner, that is, it is countable and there is a well-defined distribution on its behavior over time. Put into the context of the model, the assumption implies that once an initial payoff is specified, the sequence of subsequent payoffs for the entire planning horizon will be defined stochastically. This means that at time 0 all payoffs are defined with a known probability distribution conditioned on the initial payoff. The optimization process will then be reduced to finding an optimal contract for the corresponding payoffs.

If the assumptions in Chapters IV or V are met, an optimal or nearly optimal contract can be guaranteed.

However, the disturbance can be arbitrary. Such arbitrariness may be due to random externalities which the company has no control of or to the internal operating procedures. The imperfect state information model which is to be discussed in the next chapter is perhaps the most common situation that gives rise to an arbitrary disturbance. The actual payoff is not observable by the principal who receives a report or signal from the agent concerning the outcome. The agent can freely choose his reporting function. This will make it impossible for the principal to define his expected payoffs at time 0 to search for an optimal or nearly optimal contract.

In fact, if the disturbance is allowed to be arbitrary, various complications arise in the optimization process. This chapter will discuss the problems involved in the different phrases of the dynamic algorithm. The main intent of going through the technical details is to set the stage for the imperfect state information model such that the dynamic programming algorithm can be utilized as a solution procedure. As both the problems and their corresponding remedies are highly technical in nature, discussions in this chapter can only be carried on at a very general level. A significant amount of detail is omitted. The technically oriented reader is referred to the starred chapter for a complete development.

It was mentioned earlier in this work that there are three operations performed repetitively. First, there is the evaluation of a conditional expectation. Second, an extended real-valued function in two variables (state and incentive) is maximized over one of these variables (incentive). Finally, if an optimal or nearly optimal contract is to be constructed, a "selector" which maps each state or payoff to a contract which achieves or nearly achieves the optimum for the second step must be chosen. The following sections will take each of these operations in turn and discuss the problems associated at each stage if the disturbance is not countable.

## 6.2: Existence of Probability Measures

Elementary statistics say that probability is the measurement of the likelihood of the occurrence of a certain event from a collection (set) of events. It is a measure of the likelihood of occurrence. If the set of events or the set of all possible combinations of events is countable, a probability distribution on the elements of the set always exists. If the set is arbitrary, then very little can be said about the probability distribution of its elements. In the context of the model in the previous chapters, the conditional expectation involves not only the probability distribution on one set, but on the product of two sets, the payoff and the disturbance. It becomes essential to investigate the interplay of the distributions on these two sets.

Since probability distribution is a measure of the likelihood of occurrence of the elements of a set, its existence is closely related to the measurability of the set. When arbitrary sets are encountered, measurability is always a crucial issue. One can envision measurability of a set as the ability to count the elements in the set (counting measure) or the ability to induce a distance between the elements in the set in a one-dimensional setting or the area in a two-dimensional case (a metric or norm). A probability measure can be viewed as a function which maps the elements in the set to the real line. The space of all probability measures that can be defined on the given set $S$ is called the space of probability measures on $X$. It is denoted by $P(X)$. In Appendix I, it is shown that P(X) inherits all the properties of the original set X. Hence, P(X) is measurable whenever X is measurable. Or conversely, if X is measurable, then an unconditional probability measure always exists with respect to the specific measure of X. Nevertheless, an arbitrary measurable space is an extremely large space for any meaningful analysis to be conducted on. In order to draw any useful implications one has to restrict the research on a smaller subset which is typical enough to encompass most if not all the characteristics of the original set. The Borel set is the most common candidate for such a purpose. To define Borel sets, the idea of a $\sigma$-algebra is needed.

A collection of subsets of a set X is said to be a $\sigma$-algebra in X if it has the following properties: (1) it contains X, (2) it contains any subset A of X and the complement of A relative to X, and (3) it contains all possible unions of subsets of X. Then the Borel sets of X is the smallest $\sigma$-algebra in X such that it contains every open set in X. Since P(X) inherits all the properties of the original space, it is also a Borel space.

As mentioned earlier, the dynamic programming algorithm requires the evaluation of a conditional expectation which involves probability measures on a product of the payoff and incentive spaces. It can be shown (Theorem 6.2 and its corollaries) that a probability measure on a product of Borel spaces can be decomposed into a marginal and a conditional probability. Such decomposition is possible even when the parameters or arguments of the distribution function are dependent. In addition, such a process can be reversed (Theorem 6.3). Given a probability measure and one or more conditional probabilities, a unique probability measure on the product space can be constructed. All these distributions can be shown to be measurable if the original sets are Borel sets.

With the establishment of probability distributions on both the payoff and incentive spaces and the interplay of these probabilities between these two spaces, the conditional expectation operations in Borel spaces are well-defined.

## 6.3: Analytic Sets

The second stage of the dynamic programming algorithm involves the maximization of an extended real-valued function in two variables over one of these variables. When the disturbance is countable, the whole array of payoffs is defined stochastically given an initial payoff. The incentive function is defined on the payoff space. Under such circumstances, the resulting problem is a standard multiperiod maximization problem which has been treated somewhat in detail in Chapters IV and V. When the disturbance term is an arbitrary element from a Borel space, then $\omega_k$ cannot be deduced from the knowledge of $\omega_0$ at time period 0. Also the exact form of the optimal or nearly optimal contract cannot be specified at time 0 even if the existence of such contracts are guaranteed. The best one can do is to be able to construct a "selector" which maps each payoff to a contract which achieves or nearly achieves the maximum. Essentially, the algorithm searches for the maximum along the projection of the incentive function on the payoff space in this second stage of the process.

In searching along the projection of sets from Borel space, a very serious problem is encountered. So far in Chapters III through V, the multiperiod agency problem is formulated to involve dealing with "nice" sets. These "nice" sets have been either measurable sets or Borel sets. But, at this stage, when projections of these "nice" sets are used to search for a solution, it would be desirable that the projections are "nice" also. It is at this point that the use of measurable sets and Borel sets breaks down, because one cannot be sure that the projections required will be of the same type. The projections do not carry over the behavior of the original sets. In fact, they may not be measurable with respect to the spaces of the original sets.

Fortunately, there is another class of sets available, the so-called "analytic sets" which has the desirable properties that are required in the current model. There are many approaches to analytic sets, but maybe the best for the current purposes is that the analytic sets consist of the images of the Borel sets under continuous functions. The image of an analytic set under a continuous function is itself an analytic set. The Borel sets thus form a subclass of the analytic sets: each Borel set is an analytic set, but there are analytic sets which are not Borel sets. Also, a projection is a continuous function. Now, letting the analytic sets be the "nice" set, one obtains some control of the results of projections, that is, a guarantee of the measurability of the projections. This will enable the investigation to carry forward. By enlarging the Borel sets to include the analytic sets, the model is ready for the implementation of the dynamic programming algorithm. Technically, through the analytic sets, the projection becomes measurable with respect to the original sets.

## 6.4: Construction of the "Selector"

The last stage of the optimization process is the construction of a "selector" which maps each payoff to an optimal or nearly optimal contract. In the last section, analytic sets are utilized to enhance the measurability of the projection. If analytic sets are to be employed, it becomes inherent that the various functions should be defined such that they are measurable with respect to the analytic sets. The main result in this section is that one can construct a selector which is measurable (Theorem 6.14). Because of various technical measurability problems, a much more general and larger space is used. It is in this larger space, the universally measurable functional space in which all functions and composition of functions are measurable with respect to the relative analytic sets that the "selector" is defined and constructed. Throughout the process of constructing such a selector, a very elaborate abstract algebraic structure is imposed on the payoff and incentive sets and the various functions. The actual implementation of the dynamic programming algorithm to numerically evaluate the optimal contract and meet all these measurability requirements will not be easy.

However, the projections may not be that badly behaved. Under certain conditions, it can be shown that when extended real-valued functions involved are semicontinuous, the selectors can be chosen to be measurable with respect to the original Borel sets. Such a selector is produced in Theorems 6.15 through 6.17.

The main concern in this chapter is to take care of the technical difficulties in executing the dynamic programming algorithm when the disturbance is un-accountable. Admittedly, all these details have no direct bearing on the original economic model. However, if one were to adopt the algorithm to solve the im-perfect state information model which is discussed in the next chapter, one would have to guarantee the feasibility of obtaining a solution through the algorithm.

# CHAPTER VI*

# BOREL MODELS

## 6.1* Introduction

If the state, incentive and disturbance spaces are all arbitrary measure spaces, very little can be done. Hence, for the general model, only sparse works are done in the literature. One attempt in this direction is the work of Striebel [1975] involving p-essential suprema. The following objective function is adopted.

$$J_{k+1}(\omega) = p_k\text{-essential} \sup_{I} E\{g[\omega, I(\omega), y] + J_k[f(\omega, I(\omega), y, J_{k-1})]\} \qquad k = 0, \ldots, N-1,$$

where the p-essential supremum is taken over all measurable $I$ from the payoff space $\Omega$ to incentive space $C$ satisfying any constraints which may have been imposed. The functions $J_k$ are measurable and if the probability measures $p_o, \ldots, p_{N-1}$ are chosen properly and the so-called countable $\varepsilon$-lattice property (refer to the monograph for a precise definition) holds, the above modified dynamic programming algorithm generates the optimal net return function and can be used to obtain contracts which are optimal or nearly optimal for $p_{N-1}$ for almost all initial states. However, the selection of the proper probability measures is as difficult as executing the dynamic programming algorithm and the verification of the countable $\epsilon$-lattice property is nontrivial even in very simple situations.

A second approach is to investigate models in which the payoff (state) and incentive spaces are Borel spaces or even $R^n$ and the expected net return function

$$h(\omega,I) = g(\omega,I,y)p(dy|\omega,I)$$

is assumed to be semicontinuous and/or convex. Semi-continuous models of this type are mainly focused on various combinations of semicontinuity and compactness assumptions such that the functions $J_k$ are semicontinuous. Most of the researches that were done in this model (Freedman [1974], Furukawa [1972], Himmelberg, et. al [1976], Maitra [1968] and Schal [1972]) are carried out in a finite-dimension Euclidean state space with assumptions of convexity, semicontinuity or both made on the net return function. Results are not readily generalizable beyond Euclidean spaces (Rockafellar [1976]).

Another approach, the Borel space framework was introduced by Blackwell [1965]. The payoff (state) $\Omega$ and incentive $C$ spaces were assumed to be Borel spaces, and the functions defining the model were assumed to be Borel-measurable. However, even over a finite horizon the optimal total return function to the principal need not be Borel-measurable and there need not exist an everywhere $\epsilon$-optimal policy (Blackwell (1965), Example 2). The problem arises from the inability to choose a Borel-measurable function $\mu_k : \Omega \to C$ which nearly achieves the supremum uniformly in $\omega$. The nonexistence of such a function interferes with the construction of optimal contracts via the dynamic programming algorithm, since one must first determine at each stage the measure $p$ with respect to which it is satisfactory to nearly achieve the supremum for $p$ almost every $\omega$. This is essentially the same difficulty encountered with the Striebel approach. The difficulties in constructing nearly optimal contracts over an infinite horizon are more acute. Furthermore, from an applications point of view, a $p-\epsilon$-optimal contract, even if it can be constructed, is a much less appealing object than an everywhere $\epsilon$-optimal contract, since in many situations the distribution $p$ is unknown or may change when the system is operated repetitively, in which case a new $p - \epsilon$-optimal contract must be computed.

In the formulation that follows, the class of admissible contracts in the Borel model is enlarged to include all universally measurable contracts. It will be shown that this class is sufficiently rich to ensure that there exist everywhere $\epsilon$-optimal contracts, and, if the supremum in the dynamic programming algorithm is attained for every $\omega$ and $k$, then an everywhere optimal contract exists. Thus the notion of p-optimality can be dispensed with. Another advantage of working with the class of universally measurable functions is that this class is closed under certain basic operations such as integration with respect to a universally measurable stochastic kernel and composition.

In a dynamic programming algorithm, there are three operations performed repetitively. First, there is the evaluation of a conditional expectation. Second, an extended real-valued function in two variables (state and incentive) is supremized over one of these variables (incentive). Finally, if an optimal or nearly optimal contract is to be constructed, a "selector" which maps each state to a contract which achieves or nearly achieves the supremum in the second step must be chosen. The following sections will discuss the problems arising in each of these operations and suggest solutions whenever feasible.

## 6.2* Probability Measures on Borel Spaces

The construction of a rigorous multiperiod agency model via the dynamic programming algorithm is impossible when the payoff space and the incentive space are arbitrary sets or even when they are arbitrary measurable spaces. For this reason, the concept of a Borel space is adopted and the properties of Borel spaces are used to develop the construction.

In evaluating the conditional expectation of the total net return function, several properties of the probability measures need to be developed, the first and the obvious one being the unparameterized probability measure. Since conditional expectation involves probability measures on a product of Borel spaces, it becomes essential to investigate the interplay of the measures. It can be shown that a probability measure on a product of Borel spaces can be decomposed into a marginal and a Borel-measurable stochastic kernel. This decomposition is possible even when a measurable dependence on a parameter is admitted. Such a result is essential to the filtering algorithm for the imperfect state information model which will be developed in the next chapter. In addition, such a process can be reversed, that is, given a probability measure and one or more Borel-measurable stochastic kernels on Borel spaces, a unique probability measure on the product space can be constructed.

If $X$ is a topological space, $\mathcal{F}_X$ is the collection of closed subsets of $X$ and $\mathcal{B}_X$ the Borel $\sigma$-algebra on $X$. The space of probability measures on $(X, \mathcal{B}_X)$ is denoted by $P(X)$. $C(X)$ is the Banach space of bounded, real-valued continuous functions on $X$ with the supremum norm for any metric $d$ on $X$ consistent with its topology. A probability measure $p \in P(X)$ determines a linear functional $l_p : C(X) \to R$ defined by $l_p(f) = \int fdp$. On the other hand, a function $f \in C(X)$ determines a real-valued $\theta_f : P(X) \to R$ defined by $\theta_f(p) = \int fdp$.

The properties of the probability measure space $P(X)$ have been given much attention in statistics literature (Ash [1972], Feller [1971] are just a couple of classics), they are summarized in Appendix A. In general, one can say that P(X) inherits all characteristics of the space X. For example, if X is a separable metrizable space, then P(X) is separable and metrizable (Theorem A.4).

Definition: Let X and Y be separable metrizable spaces. A stochastic kernel $q(dy|x)$ on Y given X is a collection of probability measures in P(Y) parameterized by $x \in X$. If $\mathcal{J}$ is an $\sigma$-algebra on X and $\gamma^{-1}[\mathcal{B}_{P(Y)}] \subset \mathcal{J}$, where $\gamma : X \to P(Y)$ is defined by $\gamma(x) = q(dy|x)$, then $q(dy|x)$ is said to be $\mathcal{J}$-measurable. If $\gamma$ is continuous, $q(dy|x)$ is said to be continuous.

Before proving the decomposition theorem for stochastic kernels, the following theorem states their general behavior when the state spaces are Borel spaces.

Theorem 6.1: Let X and Y be Borel spaces, $\mathcal{E}$ a collection of subsets of Y which generates $\mathcal{B}_Y$ and is closed under finite intersections, and $q(dy|x)$ a stochastic kernel on Y given X. Then $q(dy|x)$ is Borel-measurable if and only if the mapping $\lambda_E : X \to [0,1]$ defined by $\lambda_E(x) = q(E|x)$ is Borel measurable for every $E \in \mathcal{E}$.

Proof: Let $\gamma : X \to P(Y)$ be defined by $\gamma(x) = q(dy|x)$. For $E \in \mathcal{E}$, $\lambda_E = \theta_E \circ \gamma$. If $q(dy|x)$ is Borel-measurable, then Theorem A.11 implies $\lambda_E$ is Borel-measurable for every $E \in \mathcal{E}$. Conversely, if $\lambda_E$ is Borel-measurable for every $E \in \mathcal{E}$, then

$$\sigma[\bigcup_{E\in\mathcal{E}} \lambda_E^{-1}(\mathcal{B}_R)] \subset \mathcal{B}_X.$$

By Theorem A.11, it implies

$$\gamma^{-1}[\mathcal{B}_{P(Y)}] = \gamma^{-1}[\sigma(\bigcup_{E\in\mathcal{E}} \theta_E^{-1}(\mathcal{B}_R))]$$

$$= \sigma[\bigcup_{E\in\mathcal{E}} \gamma^{-1}(\theta_E^{-1}(\mathcal{B}_R))]$$

$$= \sigma[\bigcup_{E\in\mathcal{E}} \lambda_E^{-1}(\mathcal{B}_R))] \subset \mathcal{B}_X$$

$\therefore$ $q(dy|x)$ is measurable. QED

<u>Corollary 6.1.1</u>: Let $X$ and $Y$ be Borel spaces and $q(dy|x)$ a Borel-measurable stochastic kernel on $Y$ given $X$. If $B \in \mathcal{B}_{XY}$, then the mapping $\Lambda_B : X \to [0,1]$ defined by $\Lambda_B(x) = q(B_X|x)$ where $B_X = \{y \in Y : (x,y) \in B\}$ is Borel-measurable.

Proof: If $B \in \mathcal{B}_{XY}$ and $x \in X$, then $B_X \subset Y$ is homeomorphic to $B \cap [\{x\}Y] \in \mathcal{B}_{XY}$. Thus $B_X \in \mathcal{B}_Y$ so $q(B_X|x)$ is defined. Let $\mathcal{D} = \{B \in \mathcal{B}_{XY} : \Lambda_B$ is Borel-measurable$\}$. $\mathcal{D}$ is a Dynkin system. By Theorem 6.1, $\mathcal{D}$ contains the measurable rectangles

$\therefore$ $\mathcal{D} = \mathcal{B}_{XY}$. QED

The next two results are the decomposition and integration theorems for stochastic kernels. The first one says that any probability measure on a product of Borel spaces can be decomposed into a marginal and a Borel-measurable stochastic kernel. The second theorem is the reversed statement: given a probability measure and one or more Borel-measurable stochastic kernels, a unique probability measure on the product space can be constructed. Together with their corollaries, the two theorems provide relationships between two or more probability spaces which are useful in the later development of the models.

Theorem 6.2: Let $(X,\mathcal{J})$ be a measurable space, let $Y$ and $Z$ be Borel spaces, and let $q(d(y,z)|x)$ be a stochastic kernel on $YZ$ given $X$. Assume that $q(B|x)$ is $\mathcal{J}$-measurable in $x$ for every $B \in \mathcal{B}_{XY}$. Then there exists a stochastic kernel $r(dz|x,y)$ on $Z$ given $XY$ and a stochastic kernel $s(dy|x)$ on $Y$ given $X$ such that $r(\underline{Z}|x,y)$ is $\mathcal{J}\mathcal{B}_Y$-measurable in $(x,y)$ for every $\underline{Z} \in \mathcal{B}_Z$, $s(\underline{Y}|x)$ is $\mathcal{J}$-measurable in $X$ for every $\underline{Y} \in \mathcal{B}_Y$, and

$$q(\underline{YZ}|x) = \int_Y r(\underline{Z}|x,y)\,s(dy|x) \quad \forall\ \underline{Y} \in \mathcal{B}_Y,\ \underline{Z} \in \mathcal{B}_Z.$$

Proof: Assume without loss of generality that $Y = Z = (0,1]$. Let $s(dy|x)$ be the marginal of $q(d(y,z)|x)$ on $Y$ i.e., $s(\underline{Y}|x) = q(\underline{Y}Z|x)$ for every $\underline{Y} \in \mathcal{B}_Y$. For each positive integer $n$, define subsets of $Y$

$$M(j,n) = ((j-1)/2^n, j/2^n)] \quad j = 1,\ldots,2^n.$$

Thus each $M(j,n+1)$ is a subset of some $M(k,n)$ and the collection $\{M(j,n) : n = 1,2,\ldots; j = 1,\ldots,2^n\}$ generates $\mathcal{B}_Y$. For $Z \in Q \cap Z$, where $Q$ is the set of rational numbers, define $q(dy(0,Z]|x)$ to be the measure on $Y$ whose value at $\underline{Y} \in \mathcal{B}_Y$ is $q(\underline{Y}(0,Z]|x)$. Then $q(dy(0,Z]|x)$ is absolutely continuous with respect to $s(dy|x)$ for every $z \in Q \cap Z$ and $x \in X$. Define for $z \in Q \cap Z$

$$G_n(z|x,y) = \begin{cases} q[M(j,n)(0,z]|x]/s[M(j,n)|x] \\ \quad \text{if } y \in M(j,n) \text{ and } s[M(j,n)|x] > 0 \\ 0 \quad \text{if } y \in M(j,n) \text{ and } s[M(j,n)|x] = 0. \end{cases}$$

For each $z$, the set

$$\begin{aligned} B(z) &= \{(x,y) \in XY : \lim_{n\to\infty} G_n(z|x,y) \text{ exists in } R\} \\ &= \{(x,y) \in XY : \{G_n(z|x,y)\} \text{ is Cauchy}\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{m,n \geqq N} \{(x,y) \in XY : |G_n(z|x,y) - \\ &\qquad - G_m(z|x,y)| < \tfrac{1}{k}\} \end{aligned}$$

is $\mathcal{X}\mathcal{B}_Y$-measurable.

For fixed $x$ and $y$ and for $m \geqq n$

$$q[M(j,n)(0,z]|x] = \int_{M(j,n)} G_m(z|x,y)s(dy|x).$$

But $\{M(j,n) : j = 1,\ldots,2^n\}$ is the $\sigma$-algebra generated by $G_n(z|x,y)$. This implies $G_n(z|x,y)$ is a martingale on $Y$ under the measure $s(dy|x)$. Each $G_n(z|x,y)$ is bounded above by 1, so by the Martingale convergence theorem, $G_n(z|x,y)$ converges for $s(dy|x)$ almost every $y$

$$\therefore\ s[B(z)_x|x] = 1.$$

Let

$$G(z|x,y) = \begin{cases} \lim_{n\to\infty} G_n(z|x,y) & \text{if } (x,y) \in B(z) \\ z & \text{otherwise} \end{cases}$$

Let $m \to \infty$, then

$$q[\underline{Y}(0,z]|x] = \int_{\underline{Y}} G(z|x,y)s(dy|x) \quad \forall\ x \in X,\ z \in Q \cap Z$$

$$\text{and} \quad \underline{Y} = M(j,n).$$

But $\underline{Y}$ is a Dynkin system and by Theorem A.10, then

$$q[\underline{Y}(0,z]|x] = \int_{\underline{Y}} G(z|x,y)s(dy|x) \quad \forall\ x \in X,\ z \in Q \cap Z,\ \underline{Y} \in \mathcal{B}_Y.$$

For each $z_o \in Q \cap Z$, define

$$C(z_o) = \{(x,y) \in XY : \exists\ z \in Q \cap Z \text{ with } z \leqq z_o \text{ and } G(z|x,y) > G(z_o|x,y)\}$$

$$= \bigcup_{\substack{z \in Q \cap Z \\ z \leqq z_o}} \{(x,y) \in XY : G(z|x,y) > G(z_o|x,y)\}$$

$$C = \bigcup_{z_o \in Q \cap Z} C(z_o)$$

$$D(z_o) = \{(x,y) \in XY : G(\cdot|x,y) \text{ is not right-continuous at } z_o\}$$

$$= \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcup_{\substack{z \in Q \cap Z \\ z_o \leqq z \leqq z_o + \frac{1}{k}}} \{(x,y) \in XY : |G(z|x,y) - G(z_o|x,y)| \geqq \frac{1}{n}\}$$

$$D = \bigcup_{z_o \in Q \cap Z} D(z_o)$$

$$E = \{(x,y) \in XY : G(z|x,y) \text{ does not converge to zero as } z \downarrow 0\}$$

$$= \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcup_{\substack{z \in Q \cap Z \\ z < 1/k}} \{(x,y) \in XY : |G(z|(x,y) \geqq \frac{1}{n}\}$$

$$F = \{(x,y) \in XY : G(1|x,y) \neq 1\}.$$

For fixed $x \in X$ and $z_o \in Q \cap Z$, and for all $z \in Q \cap Z$, $z \leqq z_o$

$$\int_{\underline{Y}} G(z|x,y)s(dy|x) \leqq \int_{\underline{Y}} G(z_o|x,y)s(dy|x)$$

$\therefore$ $G(z|x,y) \leqq G(z_o|x,y)$ for $s(dy|x)$ almost all $y$

$\therefore$ $s[C(z_o)_x|x] = 0$ and $s(C_x|x) = 0$.

This implies that if $z \downarrow z_o$, $z \in Q \cap Z$, then

$$\int_{\underline{Y}} G(z|x,y)s(dy|x) \downarrow \int_{\underline{Y}} G(z_o|x,y)s(dy|x)$$

and $G(z|x,y) \downarrow G(z_o|x,y)$ for $s(dy|x)$ almost all $y$

$$\therefore\ s[D(z_o)_x|x] = 0 \text{ and } s(D_x|x) = 0.$$

Furthermore, as $z \downarrow 0$, $z \in Q \cap Z$

$$\int_{\underline{Y}} G(z|x,y)\, s(dy|x) \downarrow 0 \quad \forall\ \underline{Y} \in \mathcal{B}_Y.$$

Since $G(z|x,y)$ is non-decreasing in $Z$ for $s(dy|x)$ almost all $y$, then $G(z|x,y) \downarrow 0$ for $s(dy|x)$ almost all $y$

$$\therefore\ s(E_X|x) = 0.$$

Let $z = 1$ in $q[\underline{Y}(0,z]|x]$, we obtain

$$\int_{\underline{Y}} G(1|x,y)\, s(dy|x) = s(\underline{Y}|x) \quad \forall \quad Y \in \mathcal{B}_Y.$$

Thus $G(1|x,y) = 1$ for $s(dy|x)$ almost all $y$

$$\therefore\ s(F_X|x) = 0.$$

For $z \in Z$, let $\{z_n\}$ be a sequence in $Q \cap Z$ such that $z_n \downarrow z$. For every $x \in X$, $y \in Y$, define

$$F(z|x,y) = \begin{cases} \lim\limits_{n\to\infty} G(z_n|x,y) & \text{if } (x,y) \in XY - (C \cup D \cup E \cup F) \\ z & \text{otherwise} \end{cases}$$

Clearly $F(z|x,y)$ is well-defined, nondecreasing and right-continuous.

It also satisfies for every $(x,y) \in XY$

$$0 \leq F(z|x,y) \leq 1 \quad \forall \quad z \in Z$$

$$F(1|x,y) = 1$$

and

$$\lim_{z \downarrow 0} F(z|x,y) = 0$$

$\therefore$ For each $(x,y)$ $\exists$ a probability measure $r(dz|x,y)$ on $Z$ such that $r((0,Z]|x,y) = F(Z|x,y)$ $\forall$ $z \in (0,1]$.

The collection of subsets $\underline{Z} \in \mathcal{B}_Z$ for which $r(\underline{Z}|x,y)$ is $\mathcal{J}\mathcal{B}_Y$-measurable in $(x,y)$ forms a Dynkin system which contains $\{(0,z]|z \in Z\}$

$\therefore$ $r(\underline{Z}|x,y)$ is $\mathcal{J}\mathcal{B}_Y$-measurable for every $\underline{Z} \in \mathcal{B}_Z$.

By the m notone convergence theorem, then $\forall$ $x \in X$, $z \in Z$, $\underline{Y} \in \mathcal{B}_Y$

$$\begin{aligned} q[\underline{Y}(0,z]|x] &= \int_{\underline{Y}} F(z|x,y)\,s(dy|x) \\ &= \int_{\underline{Y}} r((0,z]|x,y)\,s(dy|x). \end{aligned}$$

Again, the collection of subsets $\underline{Z} \in \mathcal{B}_Z$ for which $q(\underline{Y}Z|x] = \int_{\underline{Y}} r(z|x,y)\,s(dy|x)$ holds forms a Dynkin system which contains $\{(0,z]|z \in Z\}$

$\therefore$ $q(\underline{Y}\underline{Z}|x] = \int_{\underline{Y}} r(\underline{Z}|x,y)\,s(dy|x)$ $\forall$ $\underline{Y} \in \mathcal{B}_Y$, $\underline{Z} \in \mathcal{B}_Z$.

QED

<u>Corollary 6.2.1</u>: Let $X,Y$ and $Z$ be Borel spaces and let $q(d(y,z)|x)$ be a Borel-measurable stochastic kernel on $YZ$ given $X$. Then there exist

Borel-measurable stochastic kernels $r(dz|x,y)$ and $s(dy|x)$ on $Z$ given $XY$ and on $Y$ given $X$ respectively such that

$$q(\underline{YZ}|x) = \int_{\underline{Y}} r(\underline{Z}|x,y)s(dy|x) \quad \forall\ \underline{Y} \in \mathcal{B}_Y,\ \underline{Z} \in \mathcal{B}_Z.$$

Corollary 6.2.2: Let $Y$ and $Z$ be Borel spaces and $q \in P(YZ)$. Then there exists a Borel-measurable stochastic kernel $r(dz|y)$ on $Z$ given $Y$ such that

$$q(\underline{YZ}) = \int_{\underline{Y}} r(\underline{Z}|y)s(dy) \quad \forall\ \underline{Y} \in \mathcal{B}_Y,\ \underline{Z} \in \mathcal{B}_Z$$

where $s$ is the marginal of $q$ on $Y$.

Theorem 6.3: Let $X_1, X_2, \ldots$ be a sequence of Borel spaces, $Y_n = X_1X_2\ldots X_n$ and $Y = X_1X_2\cdots$. Let $p \in P(X_1)$ be given, and, for $n = 1,2,\ldots$ let $q_n(dx_{n+1}|y_n)$ be a Borel-measurable, stochastic kernel on $X_{n+1}$ given $Y_n$. Then, for $n = 2,3,\ldots$ there exist unique probability measures $r_n \in P(Y_n)$ such that

$$(6.1)\quad r_n(\underline{X}_1\underline{X}_2\ldots\underline{X}_n)$$
$$= \int_{\underline{X}_1}\int_{\underline{X}_2}\cdots\int_{\underline{X}_{n-1}} q_{n-1}(\underline{X}_n|x_1,x_2,\ldots,x_{n-1})$$
$$q_{n-2}(dx_{n-1}|x_1,\ldots,x_{n-2})\ldots q_1(dx_2|x_1)p(dx_1)$$
$$\forall\ \underline{X}_1 \in \mathcal{B}_{X_1},\ldots,\underline{X}_n \in \mathcal{B}_{X_n}.$$

If $f : Y_n \to R^*$ is Borel-measurable and either $\int f^+ dr_n < \infty$ or $\int f^- dr_n < \infty$, then

$$(6.2) \quad \int_{Y_n} f dr_n = \int_{X_1} \int_{X_2} \cdots \int_{X_n} f(x_1, x_2, \ldots, x_n) \, q_{n-1}(dx_n | x_1, x_2, \ldots, x_{n-1}) \cdots q_1(dx_2 | x_1) p(dx_1).$$

Furthermore, there exists a unique probability measure $r$ on $Y = X_1 X_2 \cdots$ such that for each $n$, the marginal of $r$ on $Y_n$ is $r_n$.

Proof: The spaces $Y_n$, $n = 2,3,\ldots$ and $Y$ are Borel. Let $n = 2$, for $B \in \mathcal{B}_{Y_2}$, by Corollary 6.2.1, define

$$r_2(B) = \int_{X_1} q_1(B_{x_1} | x_1) p(dx_1)$$

it is easy to see that $r_2 \in P(Y_2)$ and satisfies (6.1). Let $f : Y_2 \to R^*$ be Borel-measurable and $\int f^- dr_2 < \infty$. Consider $f^+ : Y_2 \to [0,\infty]$, $\exists$ an increasing sequence of simple functions such that $f_n \uparrow f^+$. By the monotone convergence theorem

$$\lim_{n\to\infty} \int_{X_2} f_n(x_1, x_2) q_1(dx_2 | x_1)$$

$$= \int_{X_2} f^+(x_1, x_2) q_1(dx_2 | x_1) \quad \forall \quad x_1 \in X_1$$

$$\therefore \int_{X_2} f^+(x_1, x_2) q_1(dx_2 | x_1) \quad \text{is Borel-measurable}$$

and

$$\lim_{n\to\infty} \int_{Y_2} f_n dr_2 = \lim_{n\to\infty} \int_{X_1}\int_{X_2} f_n(x_1,x_2) q_1(dx_2|x_1) p(dx_1)$$

$$= \int_{X_1}\int_{X_2} f^+(x_1,x_2) q_1(dx_2|x_1) p(dx_1).$$

But $\int_{Y_2} f_n dr_2 \uparrow \int_{Y_2} f^+ dr_2$

$\therefore$ (6.2) holds for $f^+$.

Similar arguments show (6.2) holds for $f^-$

$$\therefore \int_{Y_2} f dr_2 = \int_{Y_2} f^+ dr_2 - \int_{Y_2} f^- dr_2$$

$$= \int_{X_1} [\int_{X_2} f^+(x_1,x_2) q_1(dx_2|x_1)$$

$$- \int_{X_2} f^-(x_1,x_2) q_1(dx_2|x_1)] p(dx_1)$$

$$= \int_{X_1}\int_{X_2} f(x_1,x_2) q_1(dx_2|x_1) p(dx_1).$$

Now assume $r_k \in P(Y_k)$ exists for which (6.1) and (6.2) hold when $n = k$.

For $B \in Y_{k+1}$, let

$$r_{k+1}(B) = \int_{Y_k} q_k(B_{y_k}|y_k) r_k(dy_k).$$

Then $r_{k+1} \in P(Y_{k+1})$. If $B = \underline{X}_1\underline{X}_2 \ldots \underline{X}_k\underline{X}_{k+1}$, where $\underline{X}_j \in \mathcal{B}_{X_j}$, then

$$r_{k+1}(b) = \int \chi_{X_1 X_2 \ldots X_k}(y_k) q_k(\underline{x}_{k+1}|y_k) r_k(dy_k)$$

$$= \int_{\underline{X}_1}\int_{\underline{X}_2} \ldots \int_{\underline{X}_k} q_k(\underline{X}_{k+1}|x_1,x_2\ldots x_k) q_{k-1}(dx_k|x_{k-1})$$

$$\ldots q_1(dx_2|x_1) p(dx_1)$$

by (6.2) when $n = k$.

$\therefore$ (6.1) holds for $n = k+1$.

Then use the previous result to show (6.2) for $n = k+1$ proceeding as when $n = 2$ case.

By the induction hypothesis (6.1) and (6.2) holds. Suppose $r_n' \in P(Y_n)$ satisfies (6.1). Then the collection $\mathcal{D} = \{B \in \mathcal{B}_{Y_n} | r_n(B) = r_n'(B)\}$ is a Dynkin system containing the measurable rectangles

$\therefore\ \mathcal{D} = \mathcal{B}_{Y_n}$ and $r_n = r_n'$.

Each of the measures $r_n$ is consistent, i.e., if $m \geqq n$ then the marginal of $r_m$ on $Y_n$ is $r_n$. If each $X_k$ is complete, by the Kolmogorov theorem, $\exists$ a unique $r \in P(Y)$ whose marginal on each $Y_n$ is $r_n$. If $X_k$ is not complete, consider the completion $\tilde{X}_k$ and $Y_n$ on its completion $\tilde{Y}_n$. Again, $\exists$ $\tilde{r} \in P(\tilde{Y})$ whose marginal on each $\tilde{Y}_n$ is $\tilde{r}_n$. The uniqueness of $\tilde{r}$ implies the uniqueness of its correspondence $r \in P(Y)$. QED

In the course of proving Theorem 6.3, the following result has also been proved.

<u>Theorem 6.4</u>: Let $X$ and $Y$ be Borel spaces and $q(dy|x)$ a Borel-measurable stochastic kernel on $Y$ given $X$. If $f : XY \to R^*$ is Borel-measurable, then the function $\lambda : X \to R^*$ defined by $\lambda(x) = \int f(x,y)q(dy|x)$ is Borel-measurable.

<u>Corollary 6.4.1</u>: Let $X$ be a Borel space and let $f : X \to R^*$ be Borel-measurable. Then the function $\theta_f : P(X) \to R^*$ defined by $\theta_f(p) = \int f dp$ is Borel-measurable.

Proof: Define a Borel-measurable stochastic kernel on $X$ given $P(X)$ by $q(dx|p) = p(dx)$. Let $\tilde{f} : p(x)x \to R^*$ be defined by $\tilde{f}(p,x) = f(x)$. Then

$$\theta_f(p) = \int f(x)p(dx) = \int \tilde{f}(p,x)q(dx|p).$$

$\therefore$ By Theorem 6.4, $\theta_f$ is Borel-measurable. QED

If $f \in C(XY)$ and $q(dy|x)$ is continuous, then the mapping $\gamma$ as defined in Theorem 6.4 is also continuous. This is proved with the aid of two lemmas.

<u>Lemma 6.1</u>: Let $Y$ be a metrizable space, $d$ a metric on $Y$ consistent with its topology, and $X \subset Y$. If $g \in U_d(X)$, then $g$ has a continuous extension to $Y$.

Proof: Since $g$ is uniformly continuous on $X$, given $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ such that if $x_1, x_2 \in X$ and $d(x_1,x_2) < \delta(\varepsilon)$, then $|g(x_1) - g(x_2)| < \varepsilon$. Let $\bar{X}$ be the closure of $X$. Suppose $y \in \bar{X}$. Then there exists a sequence $\{x_n\} \subset X$ for which $x_n \to y$. Let $\varepsilon > 0$ be given, $\exists$ $N(\varepsilon)$ such that $d(x_n,x_m) < \delta(\varepsilon)$ for all $n,m > N(\varepsilon)$.

$\therefore$ $\{g(x_n)\}$ is Cauchy in $R$.

Define $\hat{g}(y) = \lim_{n\to\infty} g(x_n)$. Thus $|g(x_n) - \hat{g}(y)| < \epsilon$ whenever $n > N(\epsilon)$. Suppose now $x \in X$ and $d(x,y) < \delta(\epsilon)/2$. Choose $n > N(\epsilon)$ such that $d(x_n,y) < \delta(\epsilon)/2$. Then $d(x,x_n) < \delta(\epsilon)$ and

$$|g(x) - \hat{g}(y)| \leqq |g(x) - g(x_n)| + |g(x_n) - \hat{g}(y)|$$
$$< 2\epsilon$$

$\therefore$ For any sequence $\{x_n'\} \subset X$ with $x_n' \to y$ $\hat{g}(y) = \lim_{n\to\infty} g(x_n')$.

$\hat{g}(y)$ is independent of the sequence $\{x_n\}$ chosen. If $y \in X$, take $x_n = y$, $n = 1,2,\ldots$ and obtain $g(y) = \hat{g}(y)$

$\therefore$ $\hat{g}$ is an extension of $g$.

If $\{y_m\}$ is a sequence in $\bar{X}$ which converges to $y \in \bar{X}$, then there exist sequences $\{x_{mn}\}$ in $X$ with $y_m = \lim_{n\to\infty} x_{mn}$. Choose $n_1 < n_2 < \ldots$ such that $\lim_{m\to\infty} x_{mn_m} = y$ and $d(x_{mn_m}, y_m) < \delta(\frac{1}{m})/2$. Then $\hat{g}(y) = \lim_{m\to\infty} g(x_{mn_m})$ and $|g(x_{mn_m}) - \hat{g}(y_m)| < 2/m$. Letting $m \to \infty$, then

$$\hat{g}(y) = \lim_{m\to\infty} \hat{g}(y_m)$$

and $\hat{g}$ is continuous on $\bar{X}$. Clearly

$$\sup_{x\in X} |g(x)| = \sup_{y\in\bar{X}} |g(y)|.$$

If $\bar{X} = Y$, $\hat{g}$ is clearly unique. If $\bar{X}$ is a proper subset of $Y$, by Tietze extension theorem, $\hat{g}$ can be extended to all of $Y$ such that

$$\|g\| = \sup_{y \in Y} |\hat{g}(y)|. \quad \text{QED}$$

Lemma 6.2: Let $X$ and $Y$ be separable metrizable spaces. Then the mapping $\sigma: P(X)P(Y) \to P(XY)$ defined by $\sigma(p,q) = pq$ where $pq$ is the product of the measures $p$ and $q$ is continuous.

Proof: By Theorem A.7, $X$ and $Y$ can be homeomorphically embedded in the Hilbert cube $\mathcal{H}$. Let $X, Y \subset \mathcal{H}$ and $d$ be a metric on $\mathcal{H}\mathcal{H}$ consistent with its topology. Let $g \in U_d(X,Y)$, by Lemma 6.1, $g$ can be extended to a function $\hat{g} \in C(\mathcal{H}\mathcal{H})$. Consider the set of finite linear combinations of the form $\sum_{i=1}^{k} \hat{f}_i(x)\hat{h}_i(y)$ where $\hat{f}_i$ and $\hat{h}_i$ range over $C(\mathcal{H})$ and $k$ any integer. Let $\varepsilon > 0$ be given, by the Stone-Weierstrass Theorem such a linear combination exists and satisfies

$$\left\| \sum_{i=1}^{k} \hat{f}_i \hat{h}_i - \hat{g} \right\| < \varepsilon/3.$$

Let $\{p_n\}$ be a sequence in $P(X)$ and $p_n \to P$, $p \in P(X)$ and $\{q_n\}$ a sequence in $P(Y)$ with $q_n \to q$, $q \in P(Y)$ Consider $f_i, h_i$ the restrictions of $\hat{f}_i$ and $\hat{h}_i$ to

X and Y

$$\limsup_{n\to\infty} \left|\int_{XY} gd(p_nq_n) - \int_{XY} gd(pq)\right|$$

$$\leqq \limsup_{n\to\infty} \left|\int_{XY} (g - \sum_{i=1}^{k} f_ih_i)d(p_nq_n)\right|$$

$$+ \sum_{i=1}^{k} \lim_{n\to\infty} \left|\int_X f_idp_n \int_Y h_idq_n - \int_X f_idp \int_Y h_idq\right|$$

$$+ \lim_{n\to\infty} \left|\int_{XY} (\sum_{i=1}^{k} f_ih_i - g)d(pq)\right| \leqq \varepsilon$$

By the equivalence of Theorem A.5 (a) and (c)

$$p_n \to p \Rightarrow \int fdp_n \to \int fdp \quad \text{and} \quad q_n \to p \Rightarrow \int fdq_n \to \int fdq$$

$\therefore$ $\sigma$ is continuous. QED

<u>Theorem 6.5</u>: Let $X$ and $Y$ be separable metrizable spaces and let $q(dy|x)$ be a continuous stochastic kernel on $Y$ given $X$. If $f \in C(XY)$, then the function $\lambda : X \to R$ defined by $\lambda(x) = \int f(x,y)q(dy|x)$ is continuous.

Proof: The mapping $v : X \to P(XY)$ defined by $v(x) = p_xq(dy|x)$ is continuous by Corollary A.5.1 and Lemma 6.2. Thus $\lambda(x) = (\theta_f \circ v)(x)$ where $\theta_f : P(XY) \to R$ is defined by $\theta_f(r) = \int fdr$. By Theorem A.5, $\theta_f$ is continuous

$\therefore$ $\lambda$ is continuous. QED

With the above results, it can be seen that the conditional expectation operators in Borel spaces are well-defined.

## 6.3* Analytic Sets and Universally Measurability

The dynamic programming algorithm is centered around maximization of functions, and this is intimately connected with projections of sets. More specifically, if $f : XY \to R^*$ is given and $f^* : X \to R^*$ is defined by

$$f^*(x) = \sup_{y \in Y} f(x,y),$$

then for each $c \in R$

$$\{x \in X | f^*(x) > c\} = \text{proj}_X(\{(x,y) \in XY \; f(x,y) > c\}).$$

If $f$ is a Borel-measurable function, then $\{(x,y) | f(x,y) > c\}$ is a Borel-measurable set. Unfortunately, the projection of a Borel-measurable set need not be Borel-measurable.

As mentioned earlier in the introduction, the second stage of the dynamic programming algorithm involves in the supremization of an extended real-valued function in two variables over one of these variables. Essentially, the algorithm searches for the supremum along the projection of the set. If one were unable to guarantee the Borel-measurability of the projection, it would become impossible to implement the algorithm. However, in Borel spaces, the projection of a Borel set is an analytic set. By enlarging the Borel space to include all universally measurable functions, the model is amendable for the implementation of the dynamic programming algorithm.

Analytic sets have a very standard place in the mathematical literature. The properties that are required to develop the multiperiod agency model are summarized in Appendix B. This section will start off with the measurability properties of analytic sets.

Definition: Let X be a set. A paving $\mathcal{P}$ of X is a nonempty collection of subsets of X. The pair $(X,\mathcal{P})$ is called a paved space.

Let $\sigma(\mathcal{P})$ be the $\sigma$-algebra generated by $\mathcal{P}$. $\mathcal{P}_\delta$ denotes the collection of all intersections of countably many members of $\mathcal{P}$ and $\mathcal{P}_\sigma$ the collection of all unions of countably many members of $\mathcal{P}$. N denotes the set of positive integers. $\mathfrak{N}$ and $\Sigma$ are the set of all infinite and finite sequences of positive integers respectively.

Definition: Let $(X,\mathcal{P})$ be a paved space. A Suslin scheme for $\mathcal{P}$ is a mapping from $\Sigma$ into $\mathcal{P}$. The nucleus of a Suslin scheme $S:\Sigma \rightarrow \mathcal{P}$ is

$$N(S) = \bigcup_{(\sigma_1,\sigma_2,\ldots)\in\mathfrak{N}} \bigcap_{n=1}^{\infty} S(\sigma_1,\ldots,\sigma_n)$$

The set of all nuclei of Suslin schemes for a paving $\mathcal{P}$ is denoted by $\mathcal{S}(\mathcal{P})$.

Definition: Let $X$ be a Borel space. Denote by $\mathcal{F}_X$ the collection of closed subsets of $X$. The analytic subsets of $X$ are the members of $\mathcal{S}(\mathcal{F}_X)$.

Actually, there are a number of ways to define the class of analytic sets in a Borel space. Theorem A.13 provides seven equivalent definitions. At the beginning of this section, it was indicated that extended real-valued functions on a Borel space $X$ whose upper level sets are analytic arise naturally via partial supremization. Because the collection of analytic subsets of an uncountable Borel space is strictly larger than the Borel $\sigma$-algebra, such functions need not be Borel-measurable. Nonetheless, they can be integrated with respect to any probability measure on $(X,\mathcal{B}_X)$. The following will discuss the measurability properties of analytic sets.

Definition: Let $X$ be a Borel space. The universal $\sigma$-algebra $\mathcal{U}_X$ is defined by $\mathcal{U}_X = \bigcap_{p \in P(X)} \mathcal{B}_X(p)$. If $E \in \mathcal{U}_X$, then $E$ is universally measurable.

<u>Theorem 6.6</u> (Lusin's Theorem): Let $X$ be a Borel space and $S$ a Suslin scheme for $\mathcal{U}_X$. Then $N(S)$ is universally measurable, i.e., $\mathcal{S}(\mathcal{U}_X) = \mathcal{U}_X$.

**Proof:** Let $A = N(S)$, where $S$ is a Suslin scheme for $\mathcal{U}_X$. For $(\sigma_1,\ldots,\sigma_k) \in \Sigma$, define

$$N(\sigma_1,\ldots,\sigma_k) = \{(\zeta_1,\zeta_2,\ldots) \in \mathfrak{N} : \zeta_1 = \sigma_1,\ldots,\zeta_k = \sigma_k\}$$

and

$$M(\sigma_1,\ldots,\sigma_k) = \{(\zeta_1,\zeta_2,\ldots) \in \mathfrak{N} : \zeta_1 \leqq \sigma_1,\ldots,\zeta_k \leqq \sigma_k\}$$
$$= \bigcup_{\tau_1 \leqq \sigma_1,\ldots,\tau_k \leqq \sigma_k} N(\tau_1,\ldots,\tau_k)$$

Let

$$R(\sigma_1,\ldots,\sigma_k) = \bigcup_{z \in M(\sigma_1,\ldots,\sigma_k)} \bigcap_{s<z} S(s).$$

Then $R(\sigma_1,\ldots,\sigma_k) \subset K(\sigma_1,\ldots,\sigma_k)$ where

$$K(\sigma_1,\ldots,\sigma_k) = \bigcup_{\tau_1 \leqq \sigma_1,\ldots,\tau_k \leqq \sigma_k} \bigcap_{j=1}^{k} S(\tau_1,\ldots,\tau_j)$$

Thus $M(\sigma_1) \uparrow \mathfrak{N}$ and $R(\sigma_1) \uparrow A$ as $\sigma_1 \uparrow \infty$ and $M(\sigma_1,\ldots,\sigma_{k-1},\sigma_k) \uparrow M(\sigma_1,\ldots,\sigma_{k-1})$ and $R(\sigma_1,\ldots,\sigma_{k-1},\sigma_k) \uparrow R(\sigma_1,\ldots,\sigma_{k-1})$ as $\sigma_k \uparrow \infty$. Let $p \in P(X)$ and $\varepsilon > 0$ be given. Choose $\bar{\zeta}_1,\bar{\zeta}_2,\ldots$ such that

$$p^*(A) \leqq p^*[R(\bar{\zeta}_1)] + \varepsilon/2$$

$$p^*[R(\bar{\zeta}_1,\ldots,\bar{\zeta}_{k-1})] \leqq p^*[R(\bar{\zeta}_1,\ldots,\bar{\zeta}_{k-1},\bar{\zeta}_k)] + \varepsilon/2^k \quad k = 2,3,\cdots.$$

Then $p^*(A) \leqq p^*[R(\bar{\zeta}_1,\ldots,\bar{\zeta}_k)] + \varepsilon$. The set $K(\bar{\zeta}_1,\ldots,\bar{\zeta}_k)$ is universally measurable

$$\therefore\ 1 = p[K(\bar\zeta_1,\ldots,\bar\zeta_k)] + p[X - K(\bar\zeta_1,\ldots,\bar\zeta_k)]$$
$$\geqq p^*[R(\bar\zeta_1,\ldots,\bar\zeta_k)] + p[X - K(\bar\zeta_1,\ldots,\bar\zeta_k)]$$
$$\geqq p^*(A) - \varepsilon + p[X - K(\bar\zeta_1,\ldots,\bar\zeta_k)].$$

Let

$$x \in \bigcap_{k=1}^{\infty} K(\zeta_1,\ldots,\zeta_k)$$
$$= \bigcap_{k=1}^{\infty} \bigcup_{\tau_1\leqq\zeta_1,\ldots,\tau_k\leqq\zeta_k} \bigcap_{j=1}^{k} S(\tau_1,\ldots,\tau_j).$$

Suppose for every $\tau_1 \leqq \zeta_1$, $\exists$ a positive interer $k(\tau_1)$ such that

$$x \notin S(\tau_1) \cap [\bigcap_{k=2}^{k(\tau_1)} \bigcup_{\tau_2\leqq\zeta_2,\ldots,\tau_k\leqq\zeta_k} \bigcap_{j=2}^{k} S(\tau_1,\tau_2,\ldots,\tau_j)]$$

Let $k_o = \max_{\tau_1\leqq\zeta_1} k(\tau_1)$. Then

$$x \notin \bigcup_{\tau_1\leqq\zeta_1} \{S(\tau_1) \cap [\bigcap_{k=2}^{k_o} \bigcup_{\tau_2\leqq\zeta_2,\ldots,\tau_k\leqq\zeta_k} \bigcap_{j=2}^{k} S(\tau_1,\tau_2,\ldots,\tau_j)]\}$$
$$\supset \bigcup_{\tau_1\leqq\zeta_1,\ldots,\tau_{k_o}\leqq\zeta_{k_o}} \bigcap_{j=1}^{k_o} S(\tau_1,\ldots,\tau_j)$$
$$= K(\zeta_1,\ldots,\zeta_{k_o})$$

$\therefore$ Contradicting that $x \in \bigcap_{k=1}^{\infty} K(\zeta_1,\ldots,\zeta_k)$

$\therefore$ For some $\bar\tau_1 \leqq \zeta_1$

$$x \in S(\bar\tau_1) \cap [\bigcap_{k=2}^{\infty} \bigcup_{\tau_2\leqq\zeta_2,\ldots,\tau_k\leqq\zeta_k} \bigcap_{j=1}^{k} S(\bar\tau_1,\tau_2,\ldots,\tau_j)].$$

Similarly, $\exists\ \bar\tau_2 \leqq \zeta_2$ such that

$$x \in S(\bar\tau_1) \cap S(\bar\tau_1,\bar\tau_2)$$
$$\cap [\bigcap_{k=3}^{\infty} \bigcup_{\tau_3\leqq\zeta_3,\ldots,\tau_k} \bigcap_{j=3}^{k} S(\bar\tau_1,\bar\tau_2,\tau_3,\ldots,\tau_j)].$$

Continuing, we obtain a sequence $\bar{\tau}_1 \leqq \zeta_1, \bar{\tau}_2 \leqq \zeta_2, \ldots$ such that

$$x \in \bigcap_{k=1}^{\infty} S(\bar{\tau}_1, \ldots, \bar{\tau}_k) \subset N(S) = A$$

$$\therefore \bigcap_{k=1}^{\infty} K(\zeta_1, \ldots, \zeta_k) \subset A \quad \forall \ (\zeta_1, \ldots) \in \mathfrak{N}.$$

As $k \to \infty$, $K(\zeta_1, \ldots, \zeta_k)$ decreases to a set contained in $A$ and $X - K(\zeta_1, \ldots, \zeta_k)$ increases to a set containing $X - A$.

$\therefore$ Letting $k \to \infty$

$$1 \geqq p^*(A) - \varepsilon + p^*(X - A)$$

This implies $1 \geqq p^*(A) + p^*(X - A)$. But for any $E \subset X$, $p^*(E) + p^*(X - E) \geqq 1$

$\therefore p^*(A) + p^*(X - A) = 1$

$\therefore A$ is measurable with respect to $p$. QED

Corollary 6.6.1: Let $X$ be a Borel space, every analytic subset of $X$ is universally measurable.

Proof: The closed subsets of $X$ are universally measurable, so $\mathcal{S}(\mathcal{T}_X) \subset \mathcal{U}_X$. QED

As remarked earlier, the class of analytic subsets of an uncountable Borel space is not a $\sigma$-algebra, so there are universally measurable sets which are not analytic.

In Theorem A.11, when $X$ is a Borel space, the function $\theta_A : P(X) \to [0,1]$ defined by $\theta_A(p) = p(A)$ is Borel-measurable for every Borel-measurable $A \subset X$. The following theorem and its corollary will investigate the property of this function when $A$ is analytic. The main result is that the set $\{p \in P(X) | p(A) \geq c\}$ is analytic for each real $c$. Thus, there exists universally measurable probability measure for the analytic set $A$.

<u>Theorem 6.7</u>: Let $X$ be a Borel space and $A$ an analytic subset of $X$. For each $c \in R$, the set $\{p \in P(X) : p(A) \geq c\}$ is analytic.

Proof: Let $S$ be a Suslin scheme for $\mathcal{F}_X$ such that $A = N(S)$. For $s \in \Sigma$, let $N(s)$, $M(s)$, $R(s)$ and $K(s)$ be as defined as in Theorem 6.6, and

$$\bigcap_{k=1}^{\infty} K(\zeta_1,\ldots,\zeta_k) \subset A \quad \forall\ (\zeta_1,\ldots,) \in \mathfrak{N}.$$

Each $K(s)$ is closed. Let $\bar{p}(A) \geq c$, for any $n \geq 1$, $\exists\ (\bar{\zeta}_1,\bar{\zeta}_2,\ldots) \in \mathfrak{N}$ such that

$$\bar{p}(A) \leq \bar{p}[R(\bar{\zeta}_1,\ldots,\bar{\zeta}_k)] + 1/n.$$

Since $R(s) \subset K(s)$

$$\begin{aligned} \therefore\ \bar{p}[K(\bar{\zeta}_1,\ldots,\bar{\zeta}_k)] &\geq \bar{p}[R(\bar{\zeta}_1,\ldots,\bar{\zeta}_k)] \\ &\geq \bar{p}(A) - 1/n \\ &\geq c - 1/n \end{aligned}$$

$$\therefore \ \bar{p} \in \bigcap_{n=1}^{\infty} \bigcup_{z\in\mathfrak{N}} \bigcap_{s<z} \{p \in P(X) : p[K(s)] \geq c - 1/n\}.$$

Now, let $\bar{p} \in \bigcap_{n=1}^{\infty} \bigcup_{z\in\mathfrak{N}} \bigcap_{s<z} \{p \in P(X) : p[K(s)] \geq c - 1/n\}$.

For each $n$, $\exists \ (\zeta_1, \zeta_2, \ldots) \in \mathfrak{N}$ such that

$$\bar{p}[\bigcap_{k=1}^{\infty} K(\zeta_1, \ldots, \zeta_k)] = \lim_{k\to\infty} \bar{p}[K(\zeta_1, \ldots, \zeta_k)]$$
$$\geq c - 1/n.$$

But $\bigcap_{k=1}^{\infty} K(\zeta_1, \ldots, \zeta_k) \subset A$. Thus $\bar{p}(A) \geq c - 1/n$

$n = 1, 2,$ and $\bar{p}(A) \geq c$

$$\therefore \ \{p \in P(A) : p(A) \geq c\}$$

$$= \bigcap_{n=1}^{\infty} \bigcup_{z\in\mathfrak{N}} \bigcap_{s<z} \{p \in P(X) : p[K(s)] \geq c - 1/n\}.$$

By Theorem A.11, for each $n \geq 1$ and $s \in \Sigma$, the set

$$T_n(s) = \{p \in P(X) : p[K(s)] \geq c - 1/n\}$$

is Borel-measurable in $P(X)$ and

$$\{p \in P(X) : p(A) \geq c\} = \bigcap_{n=1}^{\infty} N(T_n).$$

By Theorem A.14 and Corollary A.13.2,

$\{p \in P(X) : p(A) \geq c\}$ is analytic. QED

<u>Corollary 6.7.1</u>: Let $X$ be a Borel space and $A$ an analytic subset of $X$. For each $c \in R$, the set $\{p \in P(X) : p(A) > c\}$ is analytic.

Proof: For each $c \in R$

$$\{p \in P(X) : p(A) > c\} = \bigcup_{n=1}^{\infty} \{p \in P(X) : p(A) \geq c + 1/n\}$$

By Corollary A.13.2 and Theorem 6.7, the set is analytic. QED

## 6.4 Universally and Borel Measurable Selection

The last stage of the optimization process is the construction of a "selector" which maps each payoff to a contract that achieves or nearly achieves the supremum. The following discussion will concentrate on the universally measurable functions and show its existence by actually constructing a selector (Theorem 6.14). If the projection of a particular Borel-measurable set turns out to be Borel-measurable, then a similar selector which is Borel-measurable can also be constructed (Theorem 6.17). Clearly, a Borel-measurable selector is a special case of the general universally measurable one.

Definition: Let $X$ be a Borel space. The analytic σ-algebra $\mathcal{A}_X$ is the smallest σ-algebra containing the analytic subsets of $X$. Notationally, $\mathcal{A}_X = \sigma[\mathcal{S}(\mathcal{F}_X)]$, where $\mathcal{F}_X$ is the collection of closed subsets of $X$. If $E \in \mathcal{A}_X$, $E$ is analytically measurable.

It is noted, by Theorems 6.2 and 6.6, that for any Borel space $X$, $\mathcal{B}_X \subset \mathcal{S}(\mathcal{F}_X) \subset \mathcal{A}_X \subset \mathcal{U}_X$.

If $X$ is countable, each of these collections of sets is equal to the power set of $X$. However, if $X$ is uncountable, each set containment in the above relationship is strict.

Definition: Let $X$ and $Y$ be Borel spaces and $f$ a function mapping $D \subset X$ into $Y$. If $D \in \mathcal{A}_X$ and $f^{-1}(B) \in \mathcal{A}_X$ for every $B \in \mathcal{B}_Y$, $f$ is said to be analytically measurable.

If $D \in \mathcal{U}_X$ and $f^{-1}(B) \in \mathcal{U}_X$ for every $B \in \mathcal{B}_Y$, $f$ is said to be universally measurable.

Because of the set containment relationship, it is clear that every Borel-measurable function is analytically measurable, and every analytically measurable function is universally measurable. The converses of these statements are false due to strict containment property.

In developing the model for universally measurable functions, it is necessary to show that universally measurable stochastic kernels can be used to define probability measures on product spaces in a manner similar to Theorem 6.3. To do that, several preliminary results are required.

Lemma 6.3: Let $X$ be a Borel space and $E \subset X$. Then $E \in \mathcal{U}_X$ if and only if given any $p \in P(X)$, there exists $B \in \mathcal{B}_X$ such that $p(E \Delta B) = 0$.

<u>Theorem 6.8</u>: Let $X, Y$ and $Z$ be Borel spaces. $D \in \mathcal{U}_X$ and $E \in \mathcal{U}_Y$. Suppose $f : D \to Y$ and $g : E \to Z$ are universally measurable and $f(D) \subset E$. Then $g \circ f$ is universally measurable.

Proof: For $p \in P(X)$, define $p' \in P(Y)$ by

$$p'(c) = p[f^{-1}(c)] \ \forall \ c \in \mathcal{B}_Y.$$

Let $V \in \mathcal{B}_Y$ be such that

$$p[f^{-1}(V) \ \Delta \ f^{-1}(U)] = p'(V \Delta U) = 0.$$

Since $f^{-1}(V) \in \mathcal{U}_X$, $\exists \ W \in \mathcal{B}_X$ such that $p[W \Delta f^{-1}(V)] = 0$. Then $p[W \Delta f^{-1}(U)] = 0$. By Lemma 6.3, $f^{-1}(U) \in \mathcal{U}_X$ for every $U \in \mathcal{U}_Y$ since $g^{-1}(B) \in \mathcal{U}_Y$ for $B \in \mathcal{B}_Z$

$\therefore \ f^{-1}[g^{-1}(B)]$ is universally measurable. QED

<u>Corollary 6.8.1</u>: Let $X$ and $Y$ be Borel spaces. $D \in \mathcal{U}_X$ and $f : D \to Y$ a universally measurable function. If $U \in \mathcal{U}_Y$, then $f^{-1}(U) \in \mathcal{U}_X$.

<u>Corollary 6.8.2</u>: Let $X, Y$ and $Z$ be Borel spaces, $D \in \mathcal{U}_X$ and $E \in A_Y$. Suppose $f : D \to Y$ and $g : E \to Z$ are analytically measurable and $f(D) \subset E$. Then $g \circ f$ is universally measurable. If $A \in \mathcal{A}_Y$, then $f^{-1}(A) \in \mathcal{U}_X$.

Corollary 6.8.3: Let $X$ and $Y$ be Borel spaces, let $f : X \to Y$ be a function, and let $q(dy|x)$ be a stochastic kernel on $Y$ given $X$ such that, for each $x$, $q(dy|x)$ assigns probability one to the point $f(x) \in Y$. Then $q(dy|x)$ is universally measurable if and only if $f$ is universally measurable.

Proof: By Corollary A.5.1, the mapping $\delta : Y \to P(Y)$ defined by $\delta(y) = p_y$ is a homeomorphism. Let $\gamma : X \to P(Y)$ be the mapping $\gamma(x) = q(dy|x)$. Thus $\gamma = \delta \circ f$ and $f = \delta^{-1} \circ \gamma$. The result follows from Theorem 6.8. QED

Lemma 6.4: Let $X$ be a Borel space and $f : X \to R^*$. The function $f$ is universally measurable if and only if, for every $p \in P(X)$, there is a Borel-measurable function $f_p : X \to R^*$ such that $f(x) = f_p(x)$ for $p$ almost every $x$.

Proof: Let $f$ be universally measurable and $p \in P(X)$ be given. For $r \in Q^*$, let
$U(r) = \{x : f(x) \leq r\}$

$$f(x) = \inf\{r \in Q^* : x \in U(r)\}.$$

Let $B(r) \in \mathcal{B}_X$ be such that $p[B(r) \Delta U(r)] = 0$. Define

$$f_p(x) = \inf\{r \in Q^* : x \in B(r)\}$$
$$= \inf_{r \in Q^*} \psi_r(x)$$

where

$$\psi_r(x) = \begin{cases} r & \text{if } x \in B(r) \\ \infty & \text{otherwise} \end{cases}$$

$\therefore$ $f_p : X \to R^*$ is Borel-measurable and

$\{x : f(x) \neq f_p(x)\} \subset \bigcup_{r \in Q^*} [B(r) \Delta U(r)]$ has

p-measure 0.

Conversely, given $p \in P(X)$, let $f_p$ be a Borel-measurable function such that $f(x) = f_p(x)$ for p almost every x. Then

$$p(\{x : f(x) \leqq c\} \Delta \{x : f_p(x) \leqq c\}) = 0 \quad \forall c \in R^*$$

$\therefore$ f is universally measurable. QED

Lemma 6.5: Let X and Y be Borel spaces and $q(dy|x)$ be a stochastic kernel on Y given X. The following statements are equivalent:

(a) The stochastic kernel $q(dy|x)$ is universally measurable.

(b) For any $B \in \mathcal{B}_Y$, the mapping $\lambda_B : X \to R$ defined by $\lambda_B(x) = q(B|x)$ is universally measurable.

(c) For any $p \in P(X)$, there exists a Borel-measurable stochastic kernel $q_p(dy|x)$ on $Y$ given $X$ such that $q(dy|x) = q_p(dy|x)$ for $p$ almost every $x$.

Proof: Suppose (a) holds. The function $\gamma : X \to P(Y)$ as defined by $\gamma(x) = q(dy|x)$ is universally measurable. Let $B \in \mathcal{B}_X$ and $\lambda_B(x) = q(B|x)$. Let $\theta_B : P(Y) \to X$ be defined by $\theta_B(p) = p(B)$. Then $\lambda_B = \theta_B \circ \gamma$. By Theorem A.11 and 6.8, $\lambda_B$ is universally measurable. Suppose (b) holds and choose $p \in P(X)$. $Y$ is separable and metrizable, $\exists$ a countable base $\mathcal{B}$ for the topology in $Y$. Let $\mathcal{J}$ be the collection of sets in $\mathcal{B}$ and their finite intersections. For $F \in \mathcal{J}$, let $f_F$ be a Borel-measurable function such that

$$f_F(x) = q(F|x) \quad \forall\, x \in B_F$$

where $B_F \in \mathcal{B}_X$ and $p(B_F) = 1$. Such a $f_F$ and $B_F$ exist by (b) and Lemma 6.4. For $x \in \bigcap_{F \in \mathcal{J}} B_F$, let $q_p(dy|x) = q(dy|x)$. For $x \notin \bigcap_{F \in \mathcal{J}} B_F$, let $q_p(dy|x)$ be some fixed probability measure in $P(Y)$.

$\therefore\ q(dy|x) = q_p(dy|x)$ for $p$ almost every $x$.

The class of sets $\underline{Y}$ in $\mathcal{B}_Y$ for which $q_p(\underline{Y}|x)$ is Borel-measurable in $x$ is a Dynkin system containing $\mathcal{J}$. The class $\mathcal{J}$ is closed under finite intersection and generates $B_Y$.

$\therefore$ By Theorem A.10, statement (c) follows.

Suppose (c) holds and let $p \in P(X)$. Define $\gamma, \gamma_p : X \to P(Y)$ by

$$\gamma(x) = q(dy|x)$$
$$\gamma_p(x) = q_p(dy|x).$$

Let $B \in \mathcal{B}_{P(x)}$, $p[\gamma^{-1}(B) \Delta \gamma_p^{-1}(B)] = 0$. By Lemma 6.3, $\gamma^{-1}(B)$ is universally measurable

$\therefore$ (c) $\Rightarrow$ (a). QED

Lemma 6.6: Let $X, Y$ and $Z$ be Borel spaces and let $f : XY \to Z$ be a universally measurable function. For fixed $x \in X$, define $g_x : Y \to Z$ by $g_x(y) = f(x,y)$. Then $g_x$ is universally measurable for every $x \in X$.

Proof: Let $x_o \in X$ be fixed and let $\varphi : Y \to XY$ be the continuous function defined by $\varphi(y) = (x_o, y)$. For $\underline{Z} \in \mathcal{B}_Z$, $\{y \in Y : g_{x_o} \in \underline{Z}\} = \varphi^{-1}\{(x,y) \in XY : f(x,y) \in \underline{Z}\}$. This set is universally measurable by Corollary 6.8.1. QED

Now, the main result is ready to be stated.

Theorem 6.9: Let $X_1, X_2, \ldots$ be a sequence of Borel spaces, $Y_n = X_1X_2\ldots X_n$ and $Y = X_1X_2\ldots$ . Let $p \in P(X_1)$ be given and, for $n = 1,2,\ldots$ let $q_n(dx_{n+1}|y_n)$ be a universally measurable stochastic kernel on $X_{n+1}$ given $Y_n$. Then for $n = 2,3,\ldots,$ there exist unique probability measures $r_n \in P(Y_n)$ such that

$$(6.3) \quad r_n(\underline{X}_1\ \underline{X}_2 \ldots \underline{X}_n) = \int_{\underline{X}_1} \int_{\underline{X}_2} \cdots \int_{\underline{X}_{n-1}} q_{n-1}(\underline{X}_n|x_1,x_2,\ldots,x_{n-1}) \, q_{n-2}(dx_{n-1}|x_1,x_2,\ldots,x_{n-2}) \ldots q_1(dx_2|x_1)p(dx_1)$$

$$\forall\ \underline{X}_1 \in \mathcal{B}_{X_1},\ldots,\underline{X}_n \in \mathcal{B}_{X_n}.$$

If $f: Y_n \to R^*$ is universally measurable and either $\int f^+ dr_n < \infty$ or $\int f^- dr_n < \infty$, then

$$(6.4) \quad \int_{Y_n} f dr_n = \int_{X_1} \int_{X_2} \cdots \int_{X_n} f(x_1,x_2,\ldots,x_n) \, q_{n-1}(dx_n|x_1,x_2,\ldots,x_{n-1}) \ldots q_1(dx_2|x_1)p(dx_1).$$

Furthermore, there exists a unique probability measure $r \in P(Y)$ such that for each $n$ the marginal of $r$ on $Y_n$ is $r_n$.

Proof: There is a Borel-measurable stochastic kernel $\bar{q}_1(dx_2|x_1)$ which agrees with $q(dx_2|x_1)$ for p almost every $x_1$. By Theorem 6.3, define $r_2 \in P(Y_2)$ by specifying it on measurable rectangles to be

$$r_2(\underline{X}_1\ \underline{X}_2) = \int_{\underline{X}_1} \bar{q}_1(\underline{X}_2|x_1)p(dx_1) \quad \forall\ \underline{X}_1 \in \mathcal{B}_{X_1},\ \underline{X}_2 \in \mathcal{B}_{X_2}.$$

Assume $f : Y_2 \to [0,\infty]$ is universally measurable and $\bar{f} : Y_2 \to [0,\infty]$ be Borel-measurable. Let $f = \bar{f}$ on $Y_2 - N$ where $N \in \mathcal{B}_{Y_2}$ and $r_2(N) = 0$. By Theorem 6.3,

$$0 = r_2(N) = \int_{\underline{X}_1} \int_{\underline{X}_2} \chi_N(x_1,x_2)\bar{q}_1(dx_2|x_1)p(dx_1)$$

$$= \int_{X_1} \bar{q}_1(N_{x_1}|x_1)p(dx_1)$$

$$\therefore\ \bar{q}_1(N_{x_1}|x_1) = 0 \quad \text{for p almost every } x_1.$$

Since $f(x_1,x_2) = \bar{f}(x_1,x_2)$ for $x_2 \notin N_{x_1}$. Thus

$$\left|\int_{X_2} [f(x_1,x_2) - \bar{f}(x_1,x_2)]\bar{q}_1(dx_2|x_1)\right|$$

$$\leqq \int_{N_{x_1}} |f(x_1,x_2) - \bar{f}(x_1,x_2)|\bar{q}_1(dx_2|x_1) = 0$$

for p almost every $x_1$. Then

$$\int_{X_1} \bar{f}(x_1,x_2)\bar{q}_1(dx_2|x_1) = \int_{X_2} f(x_1,x_2)\bar{q}_1(dx_2|x_1)$$

$$= \int_{X_2} f(x_1,x_2)q_1(dx_2|x_1)$$

for p almost every $x_1$. By Theorem 6.4, $\int_{X_1} \bar{f}(x_1,x_2)\bar{q}_1(dx_2|x_1)$ is Borel-measurable.

$\therefore$ By Lemma 6.4, $\int_{X_2} f(x_1,x_2)q_1(dx_2|x_1)$ is universally measurable.

Furthermore,

$$\int_{Y_2} f dr_2 = \int_{X_2} \bar{f} dr_2 = \int_{X_1}\int_{X_2} \bar{f}(x_1,x_2)\bar{q}_1(dx_2|x_1)p(dx_1)$$

$$= \int_{X_1}\int_{X_2} f(x_1,x_2)q_1(dx_2|x_1)p(dx_1)$$

$\therefore$ (6.4) holds for $n = 2$ and $f \geqq 0$. If

If $f: Y_2 \to R^*$ is universally measurable and satisfies $\int f^+ dr_2 < \infty$ or $\int f^- dr_2 < \infty$, then (6.4) holds for $f^+$ and $f^-$, so it holds for $f$.

Let $f = \chi_{\underline{X}_1}\chi_{\underline{X}_2}$, we obtain (6.3). Now assume the theorem holds for $n = k$. Let $\bar{q}_k(dx_{k+1}|y_k)$ be a stochastic kernel which agrees with $q_k(dx_{k+1}|y_k)$ for $r_k$ almost every $x_k$. Define $r_{k+1}$ by specifying it on measurable rectangles to be

$$r_{k+1}(\underline{X}_1\ \underline{X}_2\cdots\underline{X}_{k+1}) = \int_{\underline{X}_1\underline{X}_2\cdots\underline{X}_k} \bar{q}_k(\underline{X}_{k+1}|x_1,x_2,\ldots,x_k)dr_k$$

$$\forall\ \underline{X}_1 \in \mathcal{B}_{X_1},\ldots,\underline{X}_{k+1} \in \mathcal{B}_{X_{k+1}}.$$

Proceed as in the case $n = 2$ for $n = k+1$. The proof for the existence of $r \in P(Y)$ such that the marginal of $r$ on $X_n$ is $r_n$, $n = 2,3,\ldots$ is the same as in Theorem 6.3. QED

In the course of proving Theorem 6.9, the following fact has also been established.

Theorem 6.10: Let $X$ and $Y$ be Borel spaces and let $f : XY \to R^*$ be universally measurable. Let $q(dy|x)$ be a universally measurable stochastic kernel on $Y$ given $X$. Then the mapping $\lambda : X \to R^*$ defined by $\lambda(x) = \int f(x,y)q(dy|x)$ is universally measurable.

Corollary 6.10.1: Let $X$ be a Borel space and let $f : X \to R^*$ be universally measurable. Then the function $\theta_f : P(X) \to R^*$ defined by $\theta_f(p) = \int f dp$ is universally measurable.

The functions obtained by supremizing bivariate, extended real-valued, Borel-measurable functions over one of their variables have analytic upper level sets. These functions are called upper semianalytic functions.

Definition: Let $X$ be a Borel space, $D \subset X$, and $f : D \to R^*$. If $D$ is analytic and the set $\{x \in D | f(x) > c\}$ is analytic for every $c \in R$, then $f$ is said to be upper semianalytic.

It is clear that the idea behind upper semianalytic functions is similar to that for upper semicontinuous functions in the Borel model. The next step is to investigate whether or not semianalyticity is preserved in the optimal function and under the expectation operator.

<u>Lemma 6.7</u>: (1) Let $X$ be a Borel space, $D$ an analytic subset of $X$ and $f: D \to X$. The following statements are equivalent. (a) The function $f$ is upper semianalytic, i.e., the set

$\{x \in D: f(x) > c\}$ is analytic for every $c \in R$.

(b) The set in (a) is analytic for every $c \in R^*$.
(c) The set $\{x \in D: f(x) \geqq c\}$ is analytic for every $c \in R^*$. (d) The set in (c) is analytic for every $c \in R^*$.

(2) Let $X$ be a Borel space, $D$ an analytic subset of $X$, and $f_n: D \to R^*$, $n = 1,2,...$ a sequence of upper semianalytic functions. Then the functions $\inf_n f_n$, $\sup_n f_n$, $\liminf_{n\to\infty} f_n$ and $\limsup_{n\to\infty} f_n$ are upper semianalytic. In particular, if $f_n \to f$, then $f$ is upper semianalytic.

(3) Let $X$ and $Y$ be Borel spaces, $g: X \to Y$, and $f: g(X) \to R^*$. If $g$ is Borel-measurable and $f$ is upper semianalytic. Then $f \circ g$ is upper semianalytic.

(4) Let $X$ be a Borel space, $D$ an analytic subset of $X$, and $f,g: D \to R^*$. If $f$ and $g$ are upper semianalytic, then $f + g$ is upper semianalytic.

If, in addition, $g$ is Borel-measurable and $g \geqq 0$ or if $f \geqq 0$ and $g \geqq 0$, then $fg$ is upper semi-analytic, where we define $0 \cdot \infty = \infty \cdot 0 = 0(-\infty) = (-\infty)0 = 0$.

Proof: (1) We show (b) ⇒ (a) ⇒ (d) ⇒ (c) ⇒ (b). Clearly (b) ⇒ (a) and (d) ⇒ (c). Suppose (a) holds, then $\{x \in D : f(x) \geqq -\infty\} = D$ which is analytic by definition, while the sets

$$\{x \in D : f(x) \geqq \infty\} = \bigcap_{n=1}^{\infty} \{x \in D : f(x) > n\}$$

$$\{x \in D : f(x) \geqq c\} = \bigcap_{n=1}^{\infty} \{x \in D : f(x) > c - \tfrac{1}{n}\} \quad c \in R$$

are analytic by Corollary A.13.2

$\therefore$ (a) ⇒ (d).

If (c) holds, then the sets

$$\{x \in D : f(x) > \infty\} = \emptyset$$

$$\{x \in D : f(x) > -\infty\} = \bigcup_{n=1}^{\infty} \{x \in D : f(x) \geqq n\}$$

$$\{x \in D : f(x) > c\} = \bigcup_{n=1}^{\infty} \{x \in D : f(x) \geqq c + \tfrac{1}{n}\}$$

are analytic by Corollary A.13.2

$\therefore$ (c) ⇒ (b).

(2) Let $c \in R$,

$$\{x \in D : \inf_n f_n(x) > c\} = \bigcap_{n=1}^{\infty} \{x \in D : f_n(x) > c\}$$

$$\{x \in D : \sup_n f_n(x) \geqq c\} = \bigcup_{n=1}^{\infty} \{x \in D : f_n(x) \geqq c\}$$

$\therefore$ $\inf_n f_n$ and $\sup_n f_n$ are upper semianalytic by Corollary A.13.2 and part (1).

And

$$\liminf_{n\to\infty} f_n = \sup_{n\geq 1} \inf_{k\geq n} f_k$$

$\limsup_{n\to\infty} f_n = \inf_{n\geq 1} \sup_{k\geq n} f_k$ are upper semianalytic.

(3) By Theorem A.18, the domain $g(X)$ of $f$ is analytic. Let $c \in R$,

$$\{x \in X : (f \circ g)(x) > c\} = g^{-1}(\{y \in g(X) : f(y) > c\})$$

is analytic by Theorem A.18.

(4) Let $c \in R$,

$$\{x \in D : f(x) + g(x) > c\} = \bigcup_{r\in Q} [\{x \in D : f(x) > r\} \cap \{x \in D : g(x) > c - r\}].$$

This is true if we adopt $f(x) + g(x) = \infty + \infty = \infty$ for all $x \in D$ and $c \in R^*$. By Corollary A.13.2, $f + g$ is upper sem analytic whenever $f$ and $g$ are.

Suppose $g$ is Borel-measurable and $g \geq 0$. Let $c > 0$,

$$\{x \in D : f(x)g(x) > c\} = \bigcup_{r\in Q, r>0} \{x \in D : f(x) > r, g(x) > c/r\}.$$

If $c \leqq 0$

$$\{x \in D : f(x)g(x) > c\} = \{x \in D : f(x) \geqq 0\}$$

$$\cup \{x \in D : g(x) \geqq 0\}$$

$$\cup \Big[ \bigcup_{r \in Q, r>0} \{x \in D : f(x) > r,\ g(x) < c/r\}\Big]$$

$\therefore$ $\{x \in D : f(x)g(x) > c\}$ is analytic.

Suppose $f$ and $g$ are both semianalytic and nonnegative. For $c \in R$, the set $\{x \in D : f(x)g(x) > c\}$ is analytic as before

$\therefore$ $fg$ is upper semianalytic. QED

<u>Theorem 6.11</u>: Let $X$ and $Y$ be Borel spaces, let $D$ be an analytic subset of $XY$ and let $f : D \to R^*$ be upper semianalytic. Then the function $f^* : \text{proj}_X(D) \to R^*$ defined by $f^*(x) = \sup_{y \in D_X} f(x,y)$ is upper semianalytic. Conversely, if $f^* : X \to R^*$ is a given upper semianalytic function and $Y$ is an uncountable Borel space, then there exists a Borel-measurable function $f : XY \to R^*$ such that $f^*(x) = \sup_{y \in D_X} f(x,y)$ with $D = XY$.

Proof: If $f : D \to R^*$ is upper semianalytic and $c \in R$. The set

$$\{x \in \text{proj}_X(D) : \sup_{y \in D_X} f(x,y) > c\}$$

$$= \text{proj}_X(\{(x,y) \in D : f(x,y) > c\})$$

is analytic by Theorem A.17. Suppose $f^* : X \to R^*$ is upper semianalytic and $Y$ an uncountable Borel space. For $r \in Q$, let $A(r) = \{x \in X : f^*(x) > r\}$. Clearly $A(r)$ is analytic. By Theorem A.17, there exists $B(r) \in \mathcal{B}_{XY}$ such that $A(r) = \text{proj}_X[B(r)]$. Define $G(r) = \bigcup_{s \in Q, s \geq r} B(s)$ and $f : XY \to R^*$ by

$$f(x,y) = \sup \{r \in Q : (x,y) \in G(r)\}$$

$$= \sup_{r \in Q} \psi_r(x,y)$$

where $\psi_r(x,y) = r$ if $(x,y) \in G(r)$ and $\psi_r(x,y) = -\infty$ otherwise. $f$ is Borel-measurable. Let $g$ be defined by $g(x) = \sup_{y \in Y} f(x,y)$. If $f^*(x) > c$ for some $c \in R$ $\exists$ $r \in Q$ such that $f^*(x) > r > c$. Thus $x \in A(r)$. $\exists$ $y \in Y$ such that $(x,y) \in G(r)$ and $f(x,y) \geq r$ and $g(x) \geq r > c$

$$\therefore \quad f^*(x) \geq g(x).$$

If $g(x) > c$ for some $c \in R$ $\exists$ $r \in Q$ and $y \in Y$ such that $g(x) > r > c$ and $(x,y) \in G(r)$. Thus for some $s \in Q$, $s \geq r$, we have $(x,y) \in B(s)$ and $x \in A(s)$. This implies $f^*(x) > s \geq r > c$

$$\therefore \quad g(x) \geq f^*(x)$$

Thus $g(x) = f^*(x)$. QED

Theorem 6.12: Let $X$ and $Y$ be Borel spaces. $f : XY \to R^*$ be upper semianalytic, and $q(dy|x)$ a Borel-measurable stochastic kernel on $Y$ given $X$. Then the function $\lambda : X \to R^*$ defined by $\lambda(x) = \int f(x,y)q(dy|x)$ is upper semianalytic.

Proof: Let $g : XY \to R^*$ be defined as $g(x,y) = -f(x,y)$. Thus $\{(x,y) \in XY : g(x,y) \leqq b\}$ is analytic $\forall\, b \in R$. Suppose $g \geqq 0$. Let $g_n(x,y) = \min\{n, g(x,y)\}$. Each $g_n$ is lower semianalytic and $g_n \uparrow g$. Let

$$\begin{aligned} E_n &= \{(x,y,b) \in XYR : g_n(x,y) \leqq b \leqq n\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{r \in Q} \{(x,y,r) \in XYR : g_n(x,y) < r, \\ &\qquad\qquad r \leqq b + \tfrac{1}{k} \leqq n + \tfrac{1}{k}\}. \end{aligned}$$

By Corollary A.13.2 and Theorem A.16, $E_n$ is analytic. Let $\mu$ be the Lebesgue measure on $R$, $p \in P(XY)$ and $p\mu$ be the product measure on $XYR$. By Fubini's theorem,

$$\begin{aligned} (p\mu)(E_n) &= \int_{XY} \int_{R} \chi_{E_n} \, d\mu dp \\ &= \int_{XY} [n - g_n(x,y)]dp \\ &= n - \int_{XY} g_n(x,y)\,dp. \end{aligned}$$

For $c \in R$, by the monotone convergence theorem,

$$\{p \in P(XY) : \int g(x,y)\,dp \leqq c\}$$

$$= \bigcap_{n=1}^{\infty} \{p \in P(XY) : \int_{XY} g_n(x,y)\,dp \leqq c\}$$

$$= \bigcap_{n=1}^{\infty} \{p \in P(XY) : (p\mu)(E_n) \geqq n - c\}.$$

By Lemma 6.2, the mapping $p \to p\mu$ is continuous and by Theorem 6.7, the function $\theta_j : P(XY) \to R^*$ is defined by

$-\int g(x,y)\,dp$ is upper semianalytic.

Let $\lambda(x) = \theta_f[q(dy|x)p_x]$. Since the mapping $x \to q(dy|x)$ is Borel-measurable and $x \to p_x$ and $[q(dy|x), p_x] \to q(dy|x)p_x$ are continuous from $X$ to $P(X)$ and $P(X)P(Y)$ to $P(XY)$ respectively by Corollary A.5.1 and Lemma 6.2

$\therefore$ By Lemma 6.7 (3), $\lambda$ is upper semianalytic.

Suppose $g \leqq 0$. Let $g_n(x,y) = \max\{-n, g(x,y)\}$. Each $g_n$ is lower semianalytic and $g_n \downarrow g$. Let

$$E_n = \{(x,y,b) \in XYR : g_n(x,y) \leqq b \leqq 0\}.$$

$E_n$ is analytic

$$(p\mu)(E_n) = \int_{XY}\int_{R} \chi_{E_n}\,d\mu dp = -\int_{XY} g_n(x,y)\,dp.$$

For $c \in R$

$$\{p \in P(XY) : \int g(x,y)\,dp < c\} = \bigcup_{n=1}^{\infty} \{p \in P(XY) : \int_{XY} g_n(x,y)\,dp < c\}$$

$$= \bigcup_{n=1}^{\infty} \{p \in P(XY) : (p\mu)(E_n) > -c\}.$$

Use the same arguments as before. In the general case

$$\int f(x,y)q(dy|x) = \int f^+(x,y)q(dy|x) - \int f^-(x,y)q(dy|x).$$

Since $f^+$ and $-f^-$ are upper semianalytic, and by the preceding arguments each of the summands on the right is upper semianalytic. By Lemma 6.7 (4), $\lambda(x)$ is upper semianalytic. QED

Theorems 6.11 and 6.12 show that both the optimal function and the expectation operator are well behaved. The next two theorems will outline the procedures to obtain a measurable selector which assigns to each $x \in X$ a $y \in Y$ which attains or nearly attains the supremum in

$$f^*(x) = \sup_{y \in Y} f(x,y)$$

<u>Theorem 6.13</u> (Jankov-von Neumann Theorem): Let $X$ and $Y$ be Borel spaces and $A$ an analytic subset of $XY$. There exists an analytically measurable function $\varphi : \mathrm{proj}_X(A) \to Y$ such that $\mathrm{gr}(\varphi) \subset A$.

Proof: Let $f : \mathfrak{N} \to XY$ be continuous such that $A = f(\mathfrak{N})$. Let $g = \mathrm{proj}_X \circ f$. Thus $g : \mathfrak{N} \to X$ is continuous from $\mathfrak{N}$ onto $\mathrm{proj}_X(A)$. For $x \in \mathrm{proj}_X(A)$, $g^{-1}(\{x\})$ is a closed nonempty subset of $\mathfrak{N}$.

Let $\zeta_1(x)$ be the smallest integer which is the first component of an element $z_1 \in g^{-1}\{(x)\}$, and $\zeta_2(x)$ be the smallest integer which is the second component of an element $z_2 \in g^{-1}(\{x\})$ whose first component is $\zeta_1(x)$. In general, let $\zeta_k(x)$ be the smallest integer which is the $k^{th}$ component of an element $z_k \in g^{-1}(\{x\})$ whose first $(k-1)$st components are $\zeta_1(x),\ldots,\zeta_{k-1}(x)$. Let

$$\psi(x) = (\zeta_1(x), \zeta_2(x), \ldots)$$

Since $z_k \to \psi(x)$, $\psi(x) \in g^{-1}(\{x\})$. Define $\varphi: \mathrm{proj}_X(A) \to Y$ by $\varphi = \mathrm{proj}_Y \circ f \circ \psi$, so that $\mathrm{gr}(\varphi) \subset A$. For $(\sigma_1,\ldots,\sigma_k) \in \Sigma$, let

$$N(\sigma_1,\ldots,\sigma_k) = \{(\zeta_1,\zeta_2,\ldots) \in \mathfrak{N} : \zeta_1 = \sigma_1,\ldots,\zeta_k = \sigma_k\}$$

$$M(\sigma_1,\ldots,\sigma_k) = \{(\zeta_1,\zeta_2,\ldots) \in \mathfrak{N} : \zeta_1 \leqq \sigma_1,\ldots,\zeta_k \leqq \sigma_k\}.$$

Let $S = (\sigma_1,\sigma_2,\ldots,\sigma_k) \in \Sigma$. Suppose $x \in \psi^{-1}[N(s)]$. Let $\psi(x) = (\zeta_1(x),\zeta_2(x),\ldots)$. Then $\psi(x) \in N(s) \subset M(s)$.

$$\therefore\ \gamma = g[\psi(x)] \in g[M(s)] \text{ and}$$

$$\zeta_1(x) = \sigma_1,\ldots,\zeta_k(x) = \sigma_k.$$

By the construction of $\psi$, $\sigma_1$ is the smallest integer which is the first component of an element of $g^{-1}(\{x\})$ and for $j = 2,\ldots,k$ $\sigma_j$ is the smallest integer which is the $j^{th}$ component of an element of $g^{-1}(\{x\})$ whose first $(j-1)$ components are $\sigma_1,\ldots,\sigma_{j-1}$

$$\therefore g^{-1}(\{x\}) \cap M(\sigma_1,\ldots,\sigma_{j-1},\sigma_j-1) = \emptyset \quad j = 1,\ldots,k$$

and $x \notin \bigcup_{j=1}^{k} g[M(\sigma_1,\ldots,\sigma_{j-1},\sigma_j - 1)]$. This implies

$$\psi^{-1}[N(s)] \subset g[M(s)] - \bigcup_{j=1}^{k} g[M(\sigma_1,\ldots,\sigma_{j-1}\ \sigma_j - 1)].$$

Now suppose $x \in g[M(s)] - \bigcup_{j=1}^{k} g[M(\sigma_1,\ldots,\sigma_{j-1}\ \sigma_j - 1)]$.

Since $x \in g[M(s)]$, $\exists\ y = (\eta_1,\eta_2,\ldots) \in g^{-1}(\{x\})$ such that

$$\eta_1 \leqq \sigma_1,\ldots,\eta_k \leqq \sigma_k.$$

Clearly, $x \in \text{proj}_X(A) = g(\mathfrak{M})$

$\therefore \psi(x)$ is defined.

Let $\psi(x) = (\zeta_1(x),\zeta_2(x),\ldots)$ and $\psi(x) \in g^{-1}(\{x\})$.

Thus $g[\psi(x)] = x$. This implies $\psi(x) \notin M(\sigma_1,\ldots,\sigma_{j-1}\ \sigma_j - 1)$ $j = 1,2,\ldots,k$. Since $\psi(x) \notin M(\sigma_1 - 1)$, then $\zeta_1(x) \geqq \sigma_1$. But $\zeta_1(x)$ is the smallest integer which is the first component of an element of $g^{-1}(\{x\})$.

$$\therefore \zeta_1(x) \leqq \eta_1 \leqq \sigma_1.$$

This implies $\zeta_1(x) = \sigma_1$. Similarly, since $\psi(x) \notin M(\zeta_1(x),\sigma_2 - 1)$, $\zeta_2(x) \geqq \sigma_2$. Again $\zeta_2(x) \leqq \eta_2 \leqq \sigma_2$

$$\therefore \zeta_2(x) = \sigma_2.$$

Continuing $\psi(x) \in N(s)$ and $x \in \psi^{-1}[N(s)]$. Thus $\psi^{-1}[N(s)] \supset g[M(s)] - \bigcup_{j=1}^{k} g[M(\sigma_1,\ldots,\sigma_{j-1},\sigma_j - 1]$.

$$\therefore \psi^{-1}[N(s)] = g[M(s)] - \bigcup_{j=1}^{k} g[M(\sigma_1, \ldots, \sigma_{j-1}, \sigma_j - 1]$$

$M(t)$ is open in $\mathfrak{N}$ for every $t \in \Sigma$. By Theorem A.18, $g[M(t)]$ is analytic

$$\therefore \psi^{-1}[N(s)] \in \mathcal{A}_X \quad \forall\, s \in \Sigma.$$

But $\{N(s) : s \in \Sigma\}$ is a base for the topology on $\mathfrak{N}$

$$\sigma(\{N(s) : s \in \Sigma\}) = \mathcal{B}_{\mathfrak{N}}.$$

Thus

$$\psi^{-1}(\mathcal{B}_{\mathfrak{N}}) = \psi^{-1}[\sigma(\{N(s) : s \in \Sigma\})]$$

$$= \sigma[\psi^{-1}(\{N(s) : s \in \Sigma\})].$$

This implies $\psi^{-1}(B_{\mathfrak{N}}) \subset \mathcal{A}_X$ and $\psi$ is analytically measurable. By the definition of $\varphi$ and the Borel-measurability of $f$ and $proj_Y$

$$\varphi^{-1}(\mathcal{B}_Y) = \psi^{-1}(f^{-1}[proj_Y^{-1}(\mathcal{B}_Y)])$$

$$\subset \psi^{-1}(f^{-1}[\mathcal{B}_{XY}])$$

$$\subset \psi^{-1}(\mathcal{B}_{\mathfrak{N}}) \subset \mathcal{A}_X$$

$\therefore$ $\varphi$ is analytically measurable. QED

<u>*Theorem 6.14*</u>: Let $X$ and $Y$ be Borel spaces, $D \subset XY$ an analytic set and $f : D \rightarrow R^*$ an upper semi-*analytic* function. Define $f^* : proj_X(D) \rightarrow R^*$ by

$$f^*(x) = \sup_{y \in D_x} f(x,y).$$

(a) For every $\epsilon > 0$, there exists an analytically measurable function $\varphi:\mathrm{proj}_X(D) \to Y$ such that $\mathrm{gr}(\varphi) \subset D$ and for all $x \in \mathrm{proj}_X(D)$

$$f[x,\varphi(x)] \geqq \begin{cases} f^*(x) - \epsilon & \text{if } f^*(x) < \infty \\ 1/\epsilon & \text{if } f^*(x) = \infty \end{cases}$$

(b) The set $I = \{x \in \mathrm{proj}_X(D) : \text{for some } y_x \in D_x,\ f(x,y_x) = f^*(x)\}$ is universally measurable and for every $\epsilon > 0$ there exists a universally measurable function $\varphi:\mathrm{proj}_X(D) \to Y$ such that $\mathrm{gr}(\varphi) \subset D$ and for all $x \in \mathrm{proj}_X(D)$

$$f[x,\varphi(x)] = f^*(x) \qquad \text{if } x \in I$$

$$f[x,\varphi(x)] \geqq \begin{cases} f^*(x) - \epsilon & \text{if } x \notin I,\ f^*(x) < \infty \\ 1/\epsilon & \text{if } x \notin I,\ f^*(x) = \infty \end{cases}$$

Proof: (a) The function $f^*$ is upper semi-analytic by Theorem 6.11. For $k = 0,\pm 1,\pm 2,\ldots,$ define

$$A(k) = \{(x,y) \in D : f(x,y) > k\epsilon\}$$

$$B(k) = \{x \in \mathrm{proj}_X(D) : k\epsilon < f^*(x) \leqq (k+1)\epsilon\}$$

$$B(-\infty) = \{x \in \mathrm{proj}_X(D) : f^*(x) = -\infty\}$$

$$B(\infty) = \{x \in \mathrm{proj}_X(D) : f^*(x) = \infty\}.$$

The sets $A(k)$, $k = 0, \pm 1, \pm 2, \ldots$ and $B(\infty)$ are analytic, while the sets $B(k)$, $k = 0, \pm 1, \pm 2, \ldots$ and $B(-\infty)$ are analytically measurable. By the Jankov-von Neumann theorem, there exists, for each $k = 0, \pm 1, \pm 2, \ldots,$ an analytically measurable $\varphi_k : \mathrm{proj}_X[A(k)] \to Y$ with $(x, \varphi_k(x)) \in A(k)$ for all $x \in \mathrm{proj}_X[A(k)]$ and an analytically measurable $\bar{\varphi} : \mathrm{proj}_X(D) \to Y$ such that $(x, \bar{\varphi}(x)) \in D$ for all $x \in \mathrm{proj}_X(D)$.

Let $k^*$ be an integer such that $k^* \geq 1/\epsilon^2$. Define $\varphi : \mathrm{proj}_X(D) \to Y$ by

$$\varphi(x) = \begin{cases} \varphi_k(x) & \text{if } x \in B(k), \quad k = 0, \pm 1, \pm 2, \ldots \\ \bar{\varphi}(x) & \text{if } x \in B(-\infty) \\ \varphi_{k^*}(x) & \text{if } x \in B(\infty) \end{cases}$$

Since $B(k) \subset \mathrm{proj}_X[A(k)]$ and $B(\infty) \subset \mathrm{proj}_X[A(k)]$ for all $k$, this definition is possible. Clearly, $\varphi$ is analytically measurable and $\mathrm{gr}(\varphi) \subset D$. If $x \in B(k)$, then $(x, \varphi_k(x)) \in A(k)$ and

$$\begin{aligned} f[x, \varphi(x)] &= f[x, \varphi_k(x)] \\ &> k\epsilon \geq f^*(x) - \epsilon. \end{aligned}$$

If $x \in B(-\infty)$, then $f(x,y) = -\infty$ for all $y \in D_X$ and

$$f[x, \varphi(x)] = -\infty = f^*(x).$$

If $x \in B(\infty)$

$$f[x, \varphi(x)] = f[x, \varphi_{k^*}(x)] > k^* \epsilon > 1/\epsilon.$$

Hence $\varphi$ has the required properties.

(b) Consider the set $E \subset XYR^*$ defined by

$$E = \{(x,y,b) : (x,y) \in D,\ f(x,y) \geqq b\}$$

$$= \bigcap_{k=1}^{\infty} \bigcup_{r \in Q^*} \{(x,y,b) : (x,y) \in D,\ f(x,y) \geqq r,\ r \geqq b - 1/k\}.$$

By Theorem A.16 and Corollary A.13.2, $E$ is analytic in $XYR^*$

$\therefore$ The set $A = \text{proj}_{XR^*}(E)$ is analytic in $XR^*$.

And the mapping $T : \text{proj}_X(D) \to R^*$ defined by $T(x) = (x, f^*(x))$ is analytically measurable.

$$I = \{x : (x, f^*(x)) \in A\} = T^{-1}(A)$$

$\therefore$ $I$ is universally measurable by Corollary 6.8.2.

Since $E$ is analytic, by the Jankov-von Neumann theorem, $\exists$ an analytically measurable $\varphi : A \to Y$ such that $(x, \varphi(x,b), b) \in E$ for every $(x,b) \in A$. Define $\psi : I \to Y$ by

$$\psi(x) = \varphi(x, f^*(x)) = (\varphi \circ T)(x) \quad \forall\ x \in I.$$

By Corollary 6.8.2, $\psi$ is universally measurable and by construction $f[x, \psi(x)] \geqq f^*(x)$ for $x \in I$

$\therefore$ $f[x, \psi(x)] = f^*(x) \quad \forall\ x \in I.$

By part (a) there exists an analytically measurable $\psi_\varepsilon : \mathrm{proj}_X(D) \to Y$ such that

$$f[x, \psi_\varepsilon(x)] \geqq \begin{cases} f^*(x) - \varepsilon & \text{if } f^*(x) < \infty \\ 1/\varepsilon & \text{if } f^*(x) = \infty . \end{cases}$$

Define $\varphi : \mathrm{proj}_X(D) \to Y$ by

$$\varphi(x) = \begin{cases} \psi(x) & \text{if } x \in I \\ \psi_\varepsilon(x) & \text{if } x \in \mathrm{proj}_X(D) - I \end{cases}$$

Thus $\varphi$ is universally measurable and has the required properties. QED

It is noted that since the composition of analytically measurable functions can fail to be analytically measurable, the selector in the proof of Theorem 6.14 (b) can fail to be analytically measurable. However, the composition of universally measurable functions is universally measurable, and so a selector which is universally measurable is obtained.

Earlier in the chapter, it was mentioned that the projection of a Borel-measurable function need not be Borel-measurable. Nevertheless, under certain conditions, it can be shown that when the extended real-valued functions involved are semicontinuous, then the selectors can be chosen to be Borel-measurable.

<u>Theorem 6.15</u>: Let $X$ and $Y$ be separable metrizable spaces. Let $q(dy|x)$ be a continuous stochastic kernel on $Y$ given $X$, and let $f : XY \to R^*$ be Borel-measurable. Define $\lambda(x) = \int f(x,y)q(dy|x)$.

(a) If $f$ is lower semicontinuous and bounded below, then $\lambda$ is lower semicontinuous and bounded below.

(b) If $f$ is upper semicontinuous and bounded above, then $\lambda$ is upper semicontinuous and bounded above.

Proof: (a) Since $f$ is lower semicontinuous and bounded below, $\exists$ a sequence, $\{f_n\} \subset C(XY)$ such that $f_n \uparrow f$. Define $\lambda_n(x) = \int f_n(x,y)q(dy|x)$. Then $\lambda_n$ is continuous by Theorem 6.5 and by the monotone convergence theorem $\lambda_n \uparrow \lambda$

$\therefore$ $\lambda$ is lower semicontinuous.

(b) Same argument as (a) by complementation. QED

<u>Theorem 6.16</u>: Let $X$ and $Y$ be metrizable spaces and let $f : XY \to R^*$ be given. Define

$$f^*(x) = \sup_{y \in Y} f(x,y).$$

(a) If $f$ is lower semicontinuous, then $f^*$ is lower semicontinuous.

(b) If $f$ is upper semicontinuous and $Y$ is compact, then $f^*$ is upper semicontinuous and for every $x \in Y$ the supremum is attained by some $y \in Y$.

Proof: (a) Let $d_1$ be a metric on $X$ and $d_2$ a metric on $Y$ consistent with their topologies. Let $G \subset XY$ be open and $x_o \in \mathrm{proj}_X(G)$, $\exists$ $y_o \in Y$ such that $(x_o, y_o) \in G$ and some $\varepsilon > 0$ with

$$N_\varepsilon(x_o, y_o) = \{(x,y) \in XY : d_1(x,x_o) < \varepsilon,$$
$$d_2(y,y_o) < \varepsilon\} \subset G.$$

Then

$$x_o \in \mathrm{proj}_X[N_\varepsilon(x_o,y_o)] = \{x \in X : d_1(x,x_o) < \varepsilon\} \subset \mathrm{proj}_X G$$

$\therefore$ $\mathrm{proj}_X(G)$ is open in $X$.

Let $c \in R$

$$\{x \in X : f^*(x) > c\} = \mathrm{proj}_X(\{(x,y) \in XY : f(x,y) > c\}).$$

Since $f$ is lower semicontinuous, this implies

$$\{(x,y) : f(x,y) > c\} \text{ is open}$$

$\therefore$ $\{x \in X : f^*(x) > c\}$ is open and $f^*$ lower semicontinuous.

(b) Let $x \in X$ be fixed and let $\{y_n\} \subset Y$ be such that $f(x,y_n) \uparrow f^*(x)$. This implies $y_n \to y_o$ where $y_o \in Y$ since $\limsup_{n\to\infty} f(x,y_n) \leqq f(x,y_o)$

$\therefore$ $f(x,y_o) = f^*(x)$.

Let $\{x_n\} \subset X$ be such that $x_n \to x_o$. Choose a sequence $\{y_n\} \subset Y$ such that

$$f(x_n, y_n) = f^*(x_n) \qquad n = 1,2,\ldots$$

$\exists$ a subsequence $\{(x_{n_k}, y_{n_k})\}$ such that

$$\limsup_{n\to\infty} f(x_n, y_n) = \lim_{k\to\infty} f(x_{n_k}, y_{n_k})$$

Since $Y$ is compact, $\{y_{n_k}\} \to y_o$, $y_o \in Y$

$$\begin{aligned}\limsup_{n\to\infty} f^*(x_n) &= \limsup_{n\to\infty} f(x_n, y_n) \\ &= \lim_{k\to\infty} f(x_{n_k}, y_{n_k}) \\ &\leq f(x_o, y_o) \leqq f^*(x_o)\end{aligned}$$

$\therefore$ $f^*$ is upper semicontinuous. QED

The next lemma is very similar to Theorem 6.13 in the analytic model and Theorem 6.17 is the corresponding counterpart of Theorem 6.14 in Borel-measurable model.

<u>Lemma 6.8</u>: Let $X$ be a metrizable space, $Y$ a separable metrizable space, and $G$ an open subset of $XY$. Then $\mathrm{proj}_X(G)$ is open and there exists a Borel-measurable function $\varphi : \mathrm{proj}_X(G) \to Y$ such that $gr(\varphi) \subset G$.

Proof: Let $\{y_n : n = 1,2,\ldots\}$ be a countable dense subset of $Y$. For fixed $y \in Y$, the mapping $x \to (x,y)$ is continuous

$\therefore$ $\{x \in X : (x,y) \in G\}$ is open.

Let $G_n = \{x \in X : (x,y_n) \in G\}$. Thus $\text{proj}_X(G) = \bigcup_{n=1}^{\infty} G_n$ and $\text{proj}_X(G)$ is open. Define $\varphi : \text{proj}_X(G) \to Y$ by

$$\varphi(x) = \begin{cases} y_1 & \text{if } x \in G_1 \\ y_n & \text{if } x \in G_n - \bigcup_{k=1}^{n-1} G_n, \quad n = 2,3,\ldots \end{cases}$$

Clearly $gr(\varphi) \subset G$ and $\varphi$ is Borel-measurable. QED

<u>Theorem 6.17</u>: Let $X$ be a metrizable space, $Y$ a compact separable metrizable space, $D$ an open subset of $XY$, and let $f : D \to R^*$ be upper semicontinuous. Let $f^* : \text{proj}_X(D) \to R^*$ be given by $f^*(x) = \sup_{y \in D_X} f(x,y)$.

Then $\text{proj}_X(D)$ is open in $X$, $f^*$ is upper semicontinuous, and for every $\varepsilon > 0$, there exists a Borel-measurable function $\varphi_\varepsilon : \text{proj}_X(D) \to Y$ such that $gr(\varphi_\varepsilon) \subset D$ and for all $x \in \text{proj}_X(D)$

$$f[x,\varphi_\varepsilon(x)] \geqq \begin{cases} f^*(x) - \varepsilon & \text{if } f^*(x) < \infty \\ 1/\varepsilon & \text{if } f^*(x) = \infty. \end{cases}$$

Proof: The set $\text{proj}_X(D)$ is open in $X$ by Lemma 6.8. Let $\hat{f} : XY \to R^*$ be defined by

$$\hat{f}(x,y) = \begin{cases} f(x,y) & \text{if } (x,y) \in D \\ -\infty & \text{otherwise} \end{cases}$$

For $c \in \mathbb{R}$,

$$\{x \in \text{proj}_X(D) : f^*(x) > c\} = \{x \in X : \sup_{y \in Y} \hat{f}(x,y) > c\}.$$

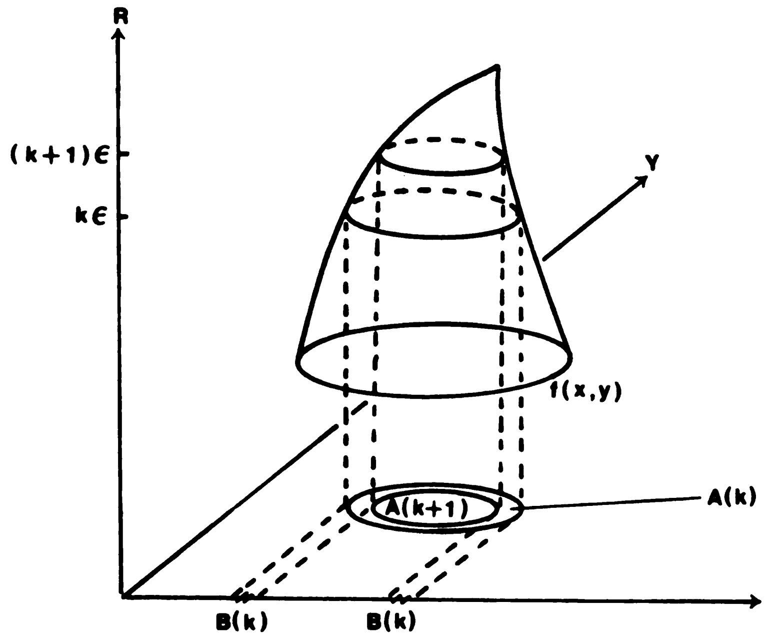
By Theorem 6.16 (b), $f^*$ is upper semicontinuous. Let $\epsilon > 0$ be given. For $k = 0, \pm 1, \pm 2, \ldots,$ define

$$A(k) = \{(x,y) \in D : f(x,y) > k\epsilon\}$$

$$B(k) = \{x \in \text{proj}_X(D) : k\epsilon < f^*(x) \leqq (k+1)\epsilon\}$$

$$B(-\infty) = \{x \in \text{proj}_X(D) : f^*(x) = -\infty\}$$

$$B(\infty) = \{x \in \text{proj}_X(D) : f^*(x) = \infty\}$$




The sets $A(k)$, $k = 0, \pm 1, \pm 2, \ldots,$ are open, while the sets $B(k)$, $B(-\infty)$, $B(\infty)$ are Borel-measurable. By Lemma 6.8, $\exists$ for each $k = 0, \pm 1, \pm 2, \ldots$ a Borel-measurable $\varphi_k : \text{proj}_X(A_k) \to Y$ such that $\text{gr}(\varphi_k) \subset A_k$. $\exists$ a Borel-measurable $\bar{\varphi} : \text{proj}_X(D) \to Y$ such that $\text{gr}(\bar{\varphi}) \subset D$. Let $k^*$ be an integer such that $k^* \geqq 1/\epsilon^2$. Define $\varphi_\epsilon : \text{proj}_X(D) \to Y$ by

$$\varphi_\varepsilon(x) = \begin{cases} \varphi_k(x) & \text{if } x \in B(k) \quad k = 0, \pm 1, \pm 2 \\ \bar{\varphi}(x) & \text{if } x \in B(-\infty) \\ \varphi_{k*}(x) & \text{if } x \in B(\infty) \end{cases}$$

Since $B(k) \subset \text{proj}_X[A(k)]$ and $B(\infty) \subset \text{proj}_X[A(k)]$ for all $k$, this definition is possible. Clearly, $\varphi_\varepsilon$ is Borel-measurable and $gr(\varphi_\varepsilon) \subset D$. If $x \in B(k)$, then, since $(x, \varphi_k(x)) \in A(k)$,

$$f[x, \varphi_\varepsilon(x)] = f[x, \varphi_k(x)] > k\varepsilon \geqq f^*(x) - \varepsilon.$$

If $x \in B(-\infty)$, then $f(x,y) = -\infty$ for all $y \in D_X$

$$f[x, \varphi_\varepsilon(x)] = -\infty = f^*(x).$$

If $x \in B(\infty)$,

$$f[x, \varphi_\varepsilon(x)] = f[x, \varphi_{k*}(x)] > k^*\varepsilon < 1/\varepsilon$$

$\therefore$ $\varphi_\varepsilon$ has the required properties. QED

So far in this chapter, all the rudiments that are necessary to carry out the dynamic programming algorithm in Borel spaces are discussed. The discussion is to a large extent technical in nature. The main ideas are to develop measurability requirements for the various operations of the algorithm. In the next chapter, our attention will be returned to the economic model and an imperfect information model will be built on the results derived in this chapter.

It should be noted that the most logical order of events is to re-examine the finite and infinite models for a basic multiperiod agency using the Borel space ideas presented here. However, I choose not to follow this sequence because the imperfect state information model can be best used to describe the reporting function of an entity about its performance to the owners. Since this application is of the greatest interest to accounting research, I shall proceed directly to the imperfect state information model, leaving the basic multiperiod agency Borel model for further research.

# CHAPTER VII

# IMPERFECT STATE INFORMATION MODEL

## 7.1: Introduction

Chapter VI discusses the problems that arise when the disturbance space is uncountable. Specifically, it becomes impossible to define stochastically the sequence of payoffs given an initial payoff. Later in the chapter, by imposing measurability restrictions on the functions and enlarging the payoff and disturbance spaces to include all analytic sets, the sequential search for optimal or nearly optimal contracts can still be implemented.

The machineries built in Chapter VI, particularly VI*, facilitate the modeling of the following economic phenomenon. A contract is agreed upon by the principal and the agent at the beginning of a typical period, say k. The agent makes his action choice in a manner of a rational decision-maker. A payoff outcome occurs at the end of the period which is only observable by the agent. The agent "reports" the payoff to the principal. There is no reason to believe or to assume that the agent always reports truthfully. In fact, the manner the agent reports the payoff should be one which is in his own best interest. This may induce a discrepancy or disturbance between the actual payoff and the reported payoff. Since the agent "chooses" the manner to report the payoff, the resulting discrepancy can be anything. What is observable to the principal is a sequence of past contracts and reported payoffs. A contract is enforceable only as its arguments are observable by all parties. Under such circumstances, the only candidates for contracting will be the sequence of past contracts and reported payoffs. However, it has been shown in the literature by various writers (Ng and Stoeckenius [1979], Harris and Raviv [1979]) that an incentive contract based on the outcome alone will cause more hazard problems. The question facing the principal will be under what circumstances can he design a long term contract to achieve a Nash equilibrium over the planning horizon.

The discrepancy between the actual payoffs and the reported payoffs under the imperfect state information model does not behave as nicely as the countable disturbance assumed in Chapters III through V. Since the reporting function belongs to some functional space which is of infinite dimension and the principal has no way of knowing what or how the agent is choosing the reporting function, there is no reason to believe that the disturbance is countable and has a well-defined distribution function. It is uncountable because the reporting function is unknown to the principal and it is of infinite dimension, that is, the space is made up of an infinite number of independent vector basis. The principal receives the value of the reporting function, the mapping (or the how) of the actual payoff to the reported payoff is never known to him. For this reason, the principal is not able to induce the actual payoff from the reported payoff. This causes more serious problem in a multiperiod model. Since the principal cannot induce the actual payoff given a reported payoff in any period, it becomes more difficult for him to define the whole sequence of payoffs even stochastically given an initial value. Recall that the payoffs form the state space for the multiperiod agency model proposed in the previous chapters, an alternative state space must be defined before the problem can be formulated. Such a state space must be inducible by the knowledge of the past contracts and reported payoffs. It must have a probability distribution conditioned on the past contracts but independent of the reported payoffs. This condition is necessary to induce truthful reporting.

As indicated earlier, contracts based on reported alone provides incentive for the agent to report untruthfully. It has been suggested that (Myerson [1979]) if the principal desires the agent to tell the truth about a particular event, the contract should be drawn independent of that event. It will be shown later in this chapter that this alternate state space is used as an argument for the incentive contract, it has to be independent of the reported payoff to induce the agent to report truthfully. Lastly, the principal must be able to utilize the state space to design a long run optimal contract.

## 7.2: The Imperfect State Information Model (ISI)

Before proceeding to describe the imperfect state information model, a remark on the solution is in order. Since it is not possible to define the state space stochastically at period 0, the optimal contract, which consists of a sequence of contracts over the time periods, cannot be defined. The idea of a contract has to be modified to a function whose image is on the interval [0,1] and whose function value assign a probability measure on all feasible contracts given the principal's belief of the initial payoff and the past contracts and the reported payoffs.

The model proposed here consists of the following objects:

$\Omega$ Payoff space

C Incentive space

Z Signal space, or the space of possible reporting functions

$\Phi_k$ Information vector. Define for $k = 0,1,\ldots,N-1$

$$\Phi_k = Z_0 C_0 \ldots C_{k-1} Z_k$$

An element of $\Phi_k$ is called a $k^{th}$ information vector, denoted by $\phi_k$.

$U_k$, $k = 0,\ldots,N-1$ Incentive constraints. These constraints exclude all contracts which will give the agent an expected utility less than outside opportunity set.

$\alpha$ Discount factor

g Net return function to the principal

$s_0$ A conditional probability of the initial signal given $\Omega$

s Conditional probability of Z given $\Omega C$

N Time horizon: a positive integer or $\infty$

P Initial payoff probability space

The ISI model can now be described notationally as follows. At period 0 the principal and the agent agree on a contract $I_0$ with the principal's probability belief of getting the initial payoff $\omega_0$ to be p and a reported payoff $z_0$ with probability $s_0$ given $\omega_0$. One can view $s_0$ being the principal's assessment of the truthfulness of the agent's reporting action. The agent chooses $a_o$ optimizing his expected utility with outcome $\omega_o$. He then reports $z_o$ to the principal who then updates his information vector and revises his belief on what he may receive as $z_1$ for the possible contracts that he may entered into with the agent as $I_2$. Then $I_2$ is agreed and the process repeated. Of course, the biggest problem here is that the only information that the principal has is the $\phi_k$. If he were to contract based on reports alone, it has been shown (Ng and Stoeckenius [1979]) that such a contract will induce non-truthful reporting.

At this stage of modeling, not much can be said. A bit of comfort is that it can be shown (Theorem 6.3) that a sequence of probability measures $P_k(\Omega_o Z_o C_o \ldots \Omega_k Z_k C_k | \pi, p)$ can be defined on $\Omega_o Z_o C_o \ldots \Omega_k Z_k C_k$ given a contract $\pi$ and initial distribution p. For notational ease, let $D_k$ denote the set of all sequences of the form $(\omega_o, z_o, I_o, \ldots, \omega_k, z_k, I_k) \in \Omega ZC \ldots \Omega ZC$. This enables one to define the N-stage total discounted expected return to the principal corresponding to a contract $\pi \in \Pi$ as

$$J_{N,\pi}(p) = \int_{D_{k-1}} [\sum_{k=0}^{N-1} \alpha^k g(\omega_k, I_k)] P_{k-1}(\pi, p)\, dH_{k-1}.$$

For the infinite horizon model, then $J_\pi = \lim_{N\to\infty} J_{N,\pi}$.

In order to ensure the integral in $J_{N,\pi}$ to be defined, the following ad hoc condition is imposed on the finite horizon model

$$J_{N,\pi}(p) < \infty \quad \forall\, \pi \in \Pi, \quad p \in P(\Omega).$$

This condition simply implies that the total expected discounted return to the principal is finite.

For the infinite horizon models, to guarantee the limit in the definition of $J_\pi$ to be well-defined, one of the following conditions is needed.

(P) $0 \leqq g(\omega, I)$ for every $(\omega, I) \in \Omega C$

(N) $g(\omega, I) \leqq 0$ for every $(\omega, I) \in \Omega C$

(D) $0 < \alpha < 1$ and for some $b \in R$, $-b \leqq g(\omega, I) \leqq b$ for every $(\omega, I) \in \Omega C$.

Condition (P) implies that the net return per period to the principal $g$ is positive. Condition (N) implies that $g$ is negative all the time. Condition (D) says that $g$ is bounded and the discount factor is less than unity. Of the three conditions, condition (D) certainly describes most common economic situations. It is not unreasonable to assume that at least one of the above conditions hold for the infinite horizon model.

As indicated earlier, based on the information vector alone, there seems little hope to obtain a long run Nash optimal contract. However, by introducing a monitoring device, a sufficient statistic, which is defined below, the ISI model can be transformed into a "perfect state information" (PSI) model. The process of transformation will be described in the next section. The sufficient statistic is defined in such a way that knowledge of its value is sufficient to design an optimal contract and control the system.

First, the term statistic is defined. A statistic for the ISI model is a sequence $(\eta_o,\ldots,\eta_{N-1})$ of functions $\eta_k : P(\Omega)\Phi_k \rightarrow \Xi_k$ where $\Xi_k$ is non-empty with generic elements of $\xi_k$, $k = 0,1,\ldots,N-1$. Then for a statistic $(\eta_o,\ldots,\eta_{N-1})$ to be sufficient for the ISI model, all of the following conditions have to be met.

(a) The statistic must guarantee that the incentive constraint set $U_k(\phi_k)$ can be recovered from $\eta_k(p;\phi_k)$. This enables the PSI model which is defined on $\Xi_k$ to search for an optimum among the same set of feasible contracts.

(b) It must guarantee that the distribution of $\xi_{k+1}$ depends only on the values of $\xi_k$ and $I_k$. Thus, the variables $\xi_k$ can be used to construct the state space of the perfect state information model.

(c) It must guarantee that the net return function to the principal $g$ corresponding to a contract $\pi$ can be computed from the distribution induced on the $(\xi_k,I_k)$ pairs.

What the sufficient statistic does is that the principal can take the information vector and from there construct an independent set of variables which he can use to assess the performance of the agent.

The ISI model describes very closely the reporting function of an entity. Management performs their routine tasks and decision making. A series of payoff outcomes occur and management chooses a reporting function to produce a set of financial statements at the end of each period. In most corporations, the principal or the owner does not take part in both the decision making and the reporting processes. The actual payoff outcomes are then unobservable by them. Then the financial statements are presented to the auditors who perform the various audit tasks, make the necessary recommendations for alteration and attest the financial statements.

If one were to investigate the definition of a sufficient statistic closely, it is not difficult to see that it actually describes the audit function. The auditor takes the reported outcome which is in fact the information vector. After performing the audit tasks and making the necessary changes, he produces the audited financial statements. If an audit is performed in accordance with the general accepted auditing standards, it is believed that the audited financial statements provides fair representation of the financial standing of the company. In other words, from the audited statements one can induce an expected payoff or return to the owner which is exactly Condition (c).

The requirement that the distribution of $\xi_{k+1}$ depends only on the values of $\xi_k$ and $I_k$ and not on the information vector $\phi_k$ parallels the auditing profession's emphasis on independence. Auditors are not to be involved in the reporting function of the company.

Condition (a) apparently is imposed more for control purposes than reporting. However, this condition implies that the principal can use the audited financial statements to search for optimal decision. The possible alternatives induced by these statements are identical with those if the actual payoffs are known. This guarantees that the decisions made are feasible and, as shown in the later section that, they are optimal also with respect to the actual payoff. This fulfills the objective of the financial statements that they should provide relevant information for decision making.

## 7.3: The Perfect State Information Model (PSI)

The manner that the sufficient statistic is defined implies that given a distribution of the initial payoff and the information vector, the principal can transform the reported payoff into an expected payoff with an independent probability distribution. If a sufficient statistic is given, that is assuming it exists, and its values can be computed from the knowledge of $P(\Omega)$ and $\Phi_k$, then the principal can define a perfect state information model in terms of $\Xi_k$. For notational purposes, a $(\hat{\ })$ is used to denote objects in the PSI model.

The perfect state information model consists of the following:

| | |
|---|---|
| $\Xi_k$, $k = 0,1,\ldots,N-1$ | State space |
| $C$ | Incentive space |
| $\hat{U}_k$, $k = 0,1,\ldots,N-1$ | Incentive constraints |
| $\alpha$ | Discount factor |
| $\hat{g}_k$, $k = 0,1,\ldots,N-1$ | One period net return to the principal |
| $\hat{t}_k$, $k = 0,1,\ldots,N-2$ | Probability distribution of $(\xi_{k+1} \mid \xi_k, I_k)$ |
| $N$ | Horizon |

## 7.4: Reduction of the Imperfect State Information Model

This section studies the relationships of the various objects between the ISI and the PSI models. The discussion on the existence of a sufficient statistic will be deferred until the next section. Throughout this part of the analysis, a sufficient statistic is assumed to exist. The Borel model as proposed in Chapter VI guarantees the existence of an optimal solution (Theorems 6.13 and 6.14, Lemma 6.8 and Theorem 6.17). By imposing the measurability restrictions on the $\Xi$ and the incentive spaces, the PSI model consists of a well-defined and measurable state space $\Xi$ with a probability distribution on $\Xi$ which depends on $\Xi_k$ and $I_k$. It becomes a direct application of the Borel model. Hence an optimal contract for the PSI model is guaranteed to exist. The natural question to ask is whether or not the optimal contract obtained from the PSI model is also optimal for the ISI model.

In order to establish correspondence between the two models, the initial probability measure on $\Xi_o$ must be ensured such that it can be induced from knowledge of the distribution of the initial payoff (Theorem 7.1). The inter-relationship between the probability distribution on the set $(\omega_o, I_o, z_o, \ldots, \omega_k, I_k, z_k) \in D_k$ and that on $(\xi_o, I_o, \ldots, I_{k-1}, \xi_k)$ given a contract $\hat{\pi}$ in the PSI model is then derived (Lemma 7.1). The analysis goes on to show that for every contract in the PSI model, a corresponding contract in the ISI model can be constructed (Theorem 7.2).

Next, given a particular contract $\hat{\pi}$ in PSI, the total expected discounted net returns under the two models are related by their respective probability distribution (Theorem 7.3) and so are the optimal total return functions (Theorem 7.4). The final step is to show the correspondence of the optimal contracts between the two models. I can only show a nearly optimal contract for the PSI model under the finiteness assumption of the finite horizon model or the Assumptions (N) and (D) of the infinite horizon model is also nearly optimal for the ISI model. However, correspondence for optimal contracts is shown for all assumptions of both the finite and infinite models (Theorem 7.5).

The ISI model can then be reduced to the PSI model for nearly optimal contracts. In terms of the auditing model, this implies that the nearly optimal contracts derived from the audited financial reports are as good as those as if one were to observe the actual payoff.

## 7.5: Sufficient Statistic

In this section, a sufficient statistic is proposed and shown to meet all the three conditions of a sufficient statistic.

It is derived by a process called filtering. In essence, filtering is very similar to the commonly known process of Bayesian statistics. The system starts with the initial wealth outcome $\omega_o$ which has a priori distribution p. After $z_o$ is observed, the distribution is "up-dated". The up-dated distribution is called a posteriori distribution and is shown to be well-defined and unique (Lemma 7.6). At the $k^{th}$ stage, there will be some a priori distribution $p_k'$ of $\omega_k$ based on $\phi_{k-1}$. Contract $I_{k-1}$ is negotiated, some $z_k$ is observed and an a posteriori distribution of $\omega_k$ conditioned on $(\phi_{k-1}, I_{k-1}, z_k)$ is computed. This distribution is again well-defined and unique. The process of passing from an a priori to an a posteriori distribution in this manner is called filtering. Then the sequence of a posterior distributions of $\omega_k$, $p_k : P(\Omega)\Phi_k \rightarrow P(\Omega)$ is shown to be a sufficient statistic (Theorem 7.6).

The filtering process seems to capture very closely the audit function. The auditor comes into engagement with a client. Based on initial interview with management and evaluation of the company's internal contral system, the auditor form some a priori opinion about the 'correctness' of the reported outcome, or what the initial payoff should be. After performing the necessary substance and compliance tests, the audit will up-date the distribution. Such a distribution will again be up-dated in subsequent periods based on the prior years' working papers. The audit report thus become the outcome of the filtering process.

In the process described above, again I am assuming that there is no incentive nor moral hazard problems exist between the principal and the auditor. The auditor is acting strictly in the best interest of the principal. If the auditor is performing the audit tasks in the best possible manner, the audited financial statements is adequate for the derivation of a long-run optimal contract to be negotiated between the principal and the agent.

CHAPTER VII*

# IMPERFECT STATE INFORMATION MODEL

## 7.1* Introduction

The imperfect state information model consists of two stochastic time series, namely the payoff and the signal on the payoff. Obviously, the later series is parameterized by the first one. In terms of the economic model, the payoff is the wealth outcome of a particular action selected by the agent. He observes the payoff and then decides on the manner that the outcome is to be reported to the principal.

The information which is available to the principal is a vector of the following form

$$\Phi_k = Z_o C_o \ldots C_{k-1} Z_k, \quad k = 0, \ldots, N-1$$

where $C$ is the incentive space and $Z$ is the signal space both of which are assumed to be nonempty Borel spaces. To the principal, since he has no direct control on the form and magnitude of $Z_k$, the report or signal in his mind is nothing but a random event stochastically generated via a signal kernel $s(dz_{k+1} | I_k, \omega_{k+1})$. The $z_{k+1}$ signal is then added to the past signals and incentives $(z_o, I_o, \ldots, z_k, I_k)$ to form the $(k+1)$st information vector $\phi_{k+1} = (z_o, I_o, \ldots, z_k, I_k, z_{k+1})$. The first information vector $\phi_o = (z_o)$ is generated by the initial observation kernel $s_o(dz_o | \omega_o)$, and the initial payoff $\omega_o$ has some given initial distribution $p$. Since $\Phi_k$ is observable by both the agent and the principal, it becomes a basis for incentive contracting. However, it is well-known that any incentive based solely on $\Phi_k$ is likely to induce misrepresentation of the outcome. The following section will develop a sufficient statistic for the payoff, or more specific, a control on the reporting function.

To describe the system in terms of $p$, the initial distribution of $\omega_o$, define

$$\begin{aligned} t(B | \omega, I, J) &= p(\{y : f(\omega, I, J, y) \in B\} | \omega, I, J) \\ &= p(f^{-1}(B)_{(\omega, I, J)} | \omega, I, J) \end{aligned}$$

$$\text{for } B \in \mathcal{B}_\Omega.$$

Thus $t(B | \omega, I, J)$ is the probability that the $(k+1)$st state is in $B$ given that the $k^{th}$ state is $\omega$, the $k^{th}$ contract is $I$ and the $k^{th}$ total net return to the principal is $J$. Alternatively, the system can be viewed as moving from $\omega_k$ to $\omega_{k+1}$ via the state transition kernel $t(d\omega_{k+1}|\omega_k, I_k, J_k)$ and generates a return to the principal of $g(\omega_k, I_k)$.

## 7.2* The Imperfect State Information Model (ISI)

The model will consist of the following objects and their corresponding assumptions

$\Omega$ Payoff space: a nonempty Borel space

$C$ Incentive space: a nonempty Borel space

$Z$ Signal space: a nonempty Borel space

$U_k$ Incentive constraints: For $k = 0,\ldots,N-1$, let $\Phi_k = Z_o C_o \ldots C_{k-1} Z_k$. An element of $\Phi_k$ is called a $k^{th}$ information vector. For each $k$, $U_k$ is a mapping from $\Phi_k$ to the set of nonempty subsets of $C$ such that

$$\Gamma_k = \{(\Phi_k, I_k) : \emptyset_k \in \Phi_k,\ I_k \in U_k(\emptyset_k)\}$$

is analytic

$\alpha$ Discount factor: a positive real number

$g$ Payoff to the owner: an upper semianalytic function from $\Gamma_k$ to $R^*$

$s_o$ Initial signal kernel: a Borel-measurable stochastic kernel on $Z$ given $\Omega$

$s$ Signal kernel: a Borel-measurable stochastic kernel on $Z$ given $C\Omega$

t  State transition kernel: a Borel-measurable stochastic kernel on $\Omega$ given $\Omega C$

N  Horizon: a positive integer or $\infty$

P  Initial payoff probability space: a nonempty Borel space.

Definition: A contract for ISI is a sequence $\pi = (\mu_0,\ldots,\mu_{N-1})$ such that, for each $k$, $k = 0,\ldots,N-1$. $\mu_k(dI_k|p;\phi_k)$ is a universally measurable stochastic kernel or $C$ given $P(\Omega)\Phi_k$ satisfying

$$\mu_k(U_k(\phi_k)|p;\phi_k) = 1 \quad \forall \quad (p,\phi_k) \in P(\Omega)\Phi_k.$$

If for each $p,k$ and $\phi_k$, $\mu_k(dI_k|p;\phi_k)$ assigns mass one to some point in $C$, $\pi$ is said to be non-randomized. Let $\Pi$ denote the set of all contracts $\pi$.

For ease of notation, let $D_k$ denote the set of all sequences of the form $(\omega_0,Z_0,I_0,\ldots,\omega_k,Z_k,I_k) \in \Omega ZC\ldots\Omega ZC$. Thus, given $p \in p(\Omega)$ and $\pi = (\mu_0,\ldots,\mu_{N-1}) \in \Pi$, by Theorem 6.3, there exists a sequence of consistent probability measures $P_k(\pi,p)$ on $D_k$, $k = 0,\ldots,N-1$ defined on measurable rectangles by

$$P_k(\pi,p)(\underline{\Omega}_o\underline{Z}_o\underline{C}_o\cdots\underline{\Omega}_k\underline{Z}_k\underline{C}_k)$$

$$= \int_{\underline{\Omega}_o}\int_{Z_o}\int_{\underline{C}_o}\cdots\int_{\underline{\Omega}_k}\int_{\underline{Z}_k} \mu_k(\underline{C}_k|p;Z_o,I_o,\ldots,I_{k-1}Z_k)$$

$$s(dZ_k|I_{k-1},\omega_k)\,t(d\omega_k|I_{k-1},\omega_{k-1})$$

$$\cdots\mu_o(dI_o|p;Z_o)\,s_o(dZ_o|\omega_o)\,p(d\omega_o)$$

where $\underline{\Omega} \in \mathcal{B}_{\Omega}$, $\underline{Z} \in \mathcal{B}_Z$, $\underline{C} \in \mathcal{B}_C$.

Definition: Given $p \in P(\Omega)$, a contract $\pi = (\mu_o,\ldots,\mu_{N-1}) \in \Pi$ and a positive integer $K \leqq N$, the K-stage payoff corresponding to $\pi$ at $p$ is

$$J_{K,\pi}(p) = \int_{D_{K-1}}[\sum_{k=0}^{K-1} \alpha^k g(\omega_k,I_k)]dP_{K-1}(\pi,p).$$

If $N < \infty$, the total net return to the principal corresponding to $\pi$ is $J_{N,\pi}$. For the finite horizon model, either one of the following conditions on the integral is assumed

$$(F^-)\quad \int_{D_{N-1}}[\sum_{k=0}^{N-1} \alpha^k g^+(\omega_k,I_k)]dP_{N-1}(\pi,p) < \infty$$

$$\forall \quad \pi \in \Pi,\ p \in P(\Omega),$$

$$(F^+)\quad \int_{D_{N-1}}[\sum_{k=0}^{N-1} \alpha^k g^-(\omega_k,I_k)]dP_{N-1}(\pi,p) < \infty$$

$$\forall \quad \pi \in \Pi,\ p \in P(\Omega).$$

Then, $J_{N,\pi}$ can be rewritten as follows

$$J_{N,\pi}(p) = \sum_{k=0}^{N-1} \alpha^k \int_{D_k} g(\omega_k,I_k)\,dP_{k-1}(\pi,p)$$

$$\forall \quad \pi \in \Pi,\ p \in P(\Omega).$$

If $N = \infty$ for infinite horizon models, then $J_\pi = \lim_{N\to\infty} J_{N,\pi}$, the return to the principal corresponding to $\pi$. In order to ensure the limit $J_\pi$ is well-defined in $R^*$, one of the following conditions is assumed on $g$

(P) $0 \leqq g(\omega, I)$ for every $(\omega, I) \in \Omega C$,

(N) $g(\omega, I) \leqq 0$ for every $(\omega, I) \in \Omega C$,

(D) $0 < \alpha < 1$ and for some $b \in R$,

$-b \leqq g(\omega, I) \leqq b$ for every $(\omega, I) \in \Omega C$,

Hence

$$J_\pi(p) = \sum_{k=0}^{\infty} \alpha^k \int_{D_k} g(\omega_k, I_k)\, dP_k(\pi, p) \quad \forall\ \pi \in \Pi,\ p \in P(\Omega).$$

The concepts of optimality at $p$, optimality, $\epsilon$-optimality at $p$ and $\epsilon$-optimality of contract are analogous to those given in Section 3.3*.

To aid in the analysis of the imperfect state model, a sufficient statistic is introduced. It is defined in such a way that knowledge of its value is sufficient to design an optimal contract and control the system.

Definition: A statistic for the model ISI is a sequence $(\eta_0, \ldots, \eta_{N-1})$ of Borel-measurable functions $\eta_k : P(\Omega)\Phi_k \to \Xi$ where $\Xi_k$ is a nonempty Borel space, $k = 0, \ldots, N-1$.

Definition: The statistic $(\eta_o,\ldots,\eta_{N-1})$ is sufficient (SS) if:

(a) For each $k$, there exists an analytic set $\hat{\Gamma}_k \subset \Xi_k C$ such that $\mathrm{proj}_k(\hat{\Gamma}_k) = \Xi_k$ and for every $p \in P(\Omega)$

$$\Gamma_k = \{(\phi_k, I) : [\eta_k(p;\phi_k), I] \in \hat{\Gamma}_k\}.$$

Define $\hat{U}_k(\xi_k) = (\hat{\Gamma}_k)_{\xi_k}$.

(b) There exist Borel-measurable stochastic kernels $\hat{t}_k(d\xi_{k+1}|\xi_k, I_k)$ on $\Xi_{k+1}$ given $\Xi_k C$ such that for every $p \in P(\Omega)$, $\pi \in \Pi$, $\underline{\Xi}_{k+1} \in \mathcal{B}_{\Xi_{k+1}}$, for $k = 0,\ldots,N-2$, we have that

$$\hat{t}_k(\Xi_{k+1}|\bar{\xi}_k, \bar{I}_k) = P_{k+1}(\pi,p)[\eta_{k+1}(p;\phi_{k+1}) \in \Xi_{k+1}| \eta_k(p;\phi_k) = \bar{\xi}_k, I_k = \bar{I}_k]$$

for $P_k(\pi,p)$-almost every $(\bar{\xi}_k, \bar{I}_k)$, that is, the set

$$\{(\omega_o, z_o, I_o, \ldots, \omega_k, z_k, I_k) \in D_k : \hat{t}_k(\Xi_{k+1}|\bar{\xi}_k, \bar{I}_k) = P_{k+1}(\pi,p)[\cdot] \text{ when } \bar{\xi}_k = \eta_k(p;\phi_k), \bar{I}_k = I_k\}$$

has $P_k(\pi,p)$-measure one.

(c) There exist upper semi-analytic functions $\hat{g}_k : \hat{\Gamma}_k \to R^*$ satisfying for every $p \in P(\Omega)$, $\pi \in \Pi$, $k = 0,\ldots,N-1$,

$$E^*[g(\omega_k, I_k) \mid \eta_k(p;\phi_k) = \bar{\xi}_k, I_k = \bar{I}_k] = \hat{g}_k(\bar{\xi}_k, \bar{I}_k)$$

for $P_k(\pi,p)$ almost every $(\bar{\xi}_k, \bar{I}_k)$ where the expectation (in the sense of an outer integral) is taken with respect to $P_k(\pi,p)$.

Condition (a) of the definition of a sufficient statistic guarantees that the incentive constraint set $U_k(\phi_k)$ can be recovered from $\eta_k(p;\phi_k)$. Indeed, for any $p \in P(\Omega)$, $\phi_k \in \Phi_k$, $k = 0,\ldots,N-1$

$$U_k(\phi_k) = \hat{U}_k[\eta_k(p;\phi_k)].$$

If $U_k(\phi_k) = C$ for every $\phi_k \in \Phi_k$, $k = 0,\ldots,N-1$, then condition (a) is satisfied with $\hat{\Gamma}_k = \Xi_k C$. This is the case of no incentive constraint. Condition (b) guarantees that the distribution of $\xi_{k+1}$ depends only on the values of $\xi_k$ and $I_k$. Thus, the variables $\xi_k$ can be used to construct the state space of the perfect state information model. Condition (c) guarantees that the net payoff to the principal corresponding to a contract can be computed from the distribution induced on the $(\xi_k, I_k)$ pairs.

7.3* <u>The Perfect State Information Model (PSI)</u>

If a sufficient statistic as defined above exists and its values can be computed from the knowledge of

$P(\Omega)$ and $\Phi_k$, then a perfect state information model can be defined in terms of $\Xi_k$. For notational purposes a $(\wedge)$ is used to denote objects in the PSI model.

Definition: Let the ISI model and a sufficient statistic $(\eta_o,\ldots,\eta_{N-1})$ be given. The perfect state information optimal incentive model (PSI) consists of the following:

$\Xi_k$, $k = 0,\ldots,N-1$ State space

$C$ Incentive space

$\hat{U}_k$, $k = 0,\ldots,N-1$ Incentive constraints

$\alpha$ Discount factor

$\hat{g}_k$, $k = 0,\ldots,N-1$ One-period net return to the principal

$\hat{t}_k$, $k = 0,\ldots,N-2$ State transition kernel as defined in Definition SS(b)

$N$ Horizon.

<u>Theorem 7.1</u>: Define $\varphi : P(\Omega) \to P(\Xi_o)$ by

$$\varphi(p)(\underline{\Xi}_o) = \int_\Omega s_o(\{Z_o : \eta_o(p;Z_o) \in \underline{\Xi}_o\}|\omega_o)p(d\omega_o)$$

for every $\underline{\Xi}_o \in \mathcal{B}_{\Xi_o}$.

Then $\varphi$ is Borel-measurable.

Proof: As defined, $\varphi(p)$ is the distribution of the initial state $\xi_o$ in PSI when the initial state $\omega_o$ in ISI has distribution $p$. Then

$$\psi_{\Xi_o}(\omega_o,p) = s_o(\{Z_o : \eta_o(p;Z_o) \in \Xi_o\}|\omega_o)$$

is Borel-measurable for every $\Xi_o \in \mathcal{B}_{\Xi_o}$ by Corollary 6.1.1. Define a Borel-measurable stochastic kernel on $\Omega$ given $P(\Omega)$ by $q(d\omega_o|p) = p(d\omega_o)$. Then

$$\varphi(p)(\Xi_o) = \int \psi_{\Xi_o}(\omega_o,p)q(d\omega_o|p).$$

By Theorems 6.1 and 6.4, $\varphi$ is Borel-measurable. QED

Theorem 7.2: If $\hat{\pi} = (\hat{\mu}_o,\ldots,\hat{\mu}_{N-1})$ is a contract for PSI, then the sequence

$$(\hat{\mu}_o[dI_o|\eta_o(p;\emptyset_o)],\ldots,\hat{\mu}_{N-1}[dI_{N-1}|\eta_o(p;\emptyset_o),I_o,$$
$$\ldots,I_{N-2},\eta_{N-1}(p;\emptyset_{N-1})]$$

where $\emptyset_k = (Z_o,I_o,\ldots,I_{k-1},Z_k)$ $k = 0,\ldots,N-1$ is a contract for ISI.

Proof: Condition (a) of the statistic $\{\eta_k\}$ to be sufficient guarantees that the incentive constraint set can be recovered from $\eta_k(p;\emptyset_k)$. Since

$$\Gamma_k = \{(\emptyset_k,I) : \emptyset_k \in \Phi_k,\ I \in U_k(\emptyset_k)\}$$
$$= \{(\emptyset_k,I) : [\eta_k(p;\emptyset_k),I] \in \hat{\Gamma}_k\}$$

and

$$\hat{U}_k(\xi_k) = (\hat{\Gamma}_k)_{\xi_k}.$$

Thus for and $p \in P(\Omega)$, $\emptyset_k \in \Phi_k$, $k = 0,\dots,N-1$

$$U_k(\emptyset_k) = \hat{U}_k[\eta_k(p,\emptyset_k)]$$

and the result follows. QED

Theorem 7.1 ensures that the initial probability measure on $\Xi$ can be induced from knowledge of the distribution of the initial payoff. Theorem 7.2 provides the very preliminary result that for every contract for PSI, a corresponding contract for ISI can be constructed. Obviously, the question to ask is whether the optimal contract for ISI can be induced from this optimal contract for PSI. Before exploring this question, a few definitions and notations are in order.

Definition: For $p \in P(\Omega)$, define the mapping $V_{p,k} : D_k \rightarrow \Xi_o C_o \cdots \Xi_k C_k$ by

$$V_{p,k}(\omega_o, Z_o, I_o, \dots, \omega_k, Z_k, I_k)$$
$$= [\eta_o(p;\emptyset_o), I_o, \dots, \eta_k(p;\emptyset_k), I_k]$$

where Theorem 7.2 holds. Thus, for $q \in P(\Xi_o)$ and $\hat{\pi} = (\hat{\mu}_o,\dots,\hat{\mu}_{N-1}) \in \hat{\Pi}$, there is a sequence of consistent probability measures $\hat{P}_k(\hat{\pi},q)$ generated on $\Xi_o C_o \cdots \Xi_k C_k$, $k = 0,\dots,N-1$, which are defined on measurable rectangles by

$$\hat{P}_k(\hat{\pi},q)(\underline{\Xi}_o\, \underline{C}_o \cdots \underline{\Xi}_k\, \underline{C}_k)$$

$$= \int_{\underline{\Xi}_o} \int_{\underline{C}_o} \cdots \int_{\underline{\Xi}_k} \hat{\mu}_k(\underline{C}_k | \xi_o, I_o, \ldots, I_{k+1}, \xi_k)\, \hat{t}_{k-1}(d\xi_k | \xi_{k-1}, I_{k-1}) \cdots \hat{\mu}_o(dI_o | \xi_o)\, q(d\xi_o)$$

where $\underline{\Xi}_k \in \mathcal{B}_{\Xi_k}$, $\underline{C}_k \in \mathcal{B}_{C_k}$.

When $(\hat{F}^+)$, $(\hat{F}^-)$, $(\hat{P})$, $(\hat{N})$ or $(\hat{D})$ holds, the net total return to the principal corresponding to a contract $\hat{\pi}$ for PSI at $\xi \in \Xi_o$ is

$$\hat{J}_{N,\hat{\pi}}(\xi) = \sum_{k=0}^{N-1} \alpha^k \int_{\Xi_o C_o \cdots \Xi_k C_k} \hat{g}_k(\xi_k, I_k)\, d\hat{P}_k(\hat{\pi}, p_\xi)$$

where $N \in [1,\infty]$ and the optimal return for PSI at $\xi \in \Xi_o$ is

$$\hat{J}^*_N(\xi) = \sup_{\hat{\pi} \in \hat{\Pi}} \hat{J}_{N,\hat{\pi}}(\xi).$$

## 7.4* Reduction of the Imperfect State Information Model

This section is devoted to study the relationships between net returns, optimal and nearly optimal contracts for the ISI and the PSI models. First, for a contract $\hat{\pi}$ for the PSI, these objects are related to the probability measures $P_k(\hat{\pi},p)$ as defined in Section 7.2 in the following manner.

<u>Lemma 7.1</u>: Suppose $p \in P(\Omega)$ and $\hat{\pi} \in \hat{\Pi}$. Then for $k = 0,\ldots,N-1$ and for every Borel set $B \subset \Xi_0 C_0 \ldots \Xi_k C_k$, we have

$$P_k(\hat{\pi},p)[V_{p,k}^{-1}(B)] = \hat{P}_k[\hat{\pi},\varphi(p)](B).$$

Proof: It suffices to prove if $\underline{\Xi}_0 \in \mathcal{B}_{\Xi_0}$, $\underline{C}_0 \in \mathcal{B}_{C_0},\ldots,\underline{\Xi}_k \in \mathcal{B}_{\Xi_k}$, $\underline{C}_k \in \mathcal{B}_{C_k}$ then

$$P_k(\hat{\pi},p)(\{\eta_0(p;\emptyset_0) \in \underline{\Xi}_0,\ I_0 \in C_0,\ldots,\eta_k(p;\emptyset_k) \in \underline{\Xi}_k,\ I_k \in \underline{C}_k\})$$
$$= \hat{P}_k[\hat{\pi},\varphi(p)](\underline{\Xi}_0\ \underline{C}_0,\ldots,\underline{\Xi}_k\ \underline{C}_k)$$

where

$$\{\eta_0(p;\emptyset_0) \in \underline{\Xi}_0,\ I_0 \in C_0,\ldots,\eta_k(p;\emptyset_k) \in \Xi_k,\ I_k \in \underline{C}_k\}$$
$$= \{(\omega_0,Z_0,I_0,\ldots,\omega_k,Z_k,I_k) : \eta_0(p;\emptyset_0) \in \underline{\Xi}_0,$$
$$I_0 \in C_0,\ldots,\eta_k(p;\emptyset_k) \in \underline{\Xi}_k,\ I_k \in \underline{C}_k\}$$

$$\emptyset_j = (Z_0,I_0,\ldots,I_{j-1},Z_j).$$

For $k = 0$, the result clearly holds by the definitions of $P_0(\pi,p)$ on $D_0$, $\varphi(p)(\Xi_0)$ and $\hat{P}_0(\hat{\pi},q)$ on $\Xi_0 C_0$. If $P_k(\hat{\pi},p)[V_{p,k}^{-1}(B)] = \hat{P}_k[\hat{\pi},\varphi(p)](B)$ for $k < N$ and since

$$\hat{t}_k(\underline{\Xi}_{k+1}|\bar{\xi}_k,\bar{I}_k) = P_{k+1}(\pi,p)[\eta_{k+1}(p;\emptyset_{k+1}) \in \underline{\Xi}_{k+1}|$$
$$\eta_k(p;\emptyset_k) = \bar{\xi}_k,\ I_k = \bar{I}_k]$$

then

$$P_{k+1}(\hat{\pi},p)(\{\eta_o(p;\emptyset_o) \in \underline{\Xi}_o, I_o \in C_o,$$

$$\ldots, \eta_{k+1}(p;\emptyset_{k+1}) \in \underline{\Xi}_{k+1}, I_{k+1} \in \underline{C}_{k+1}\})$$

$$= \int_{\{\eta_o(p;\emptyset_o)\in\underline{\Xi}_o, J_o\in s_o,\ldots,\eta_k(p;\emptyset_k)\in\underline{\Xi}_k, J_k\in\underline{C}_k} \int_{\underline{\Xi}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|\xi_{k+1})$$

$$\hat{t}_k(d\xi_{k+1}|\eta_k(p;\emptyset_k), I_k)\, dP_k(\hat{\pi},p)$$

$$= \int_{\underline{\Xi}_o \underline{C}_o \cdots \underline{\Xi}_k \underline{C}_k} \int_{\underline{\Xi}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|\xi_{k+1})\hat{t}_k(d\xi_{k+1}|\xi_k, I_k)\, d\hat{P}_k[\hat{\pi},\varphi(p)]$$

$$= \hat{P}_{k+1}[\pi\varphi(p)](\underline{\Xi}_o\ \underline{C}_o \cdots \underline{\Xi}_{k+1}\ \underline{C}_{k+1}). \quad \text{QED}$$

The next theorem establishes the relation between the total net return functions to the principal for the two models for a given contract $\hat{\pi}$.

<u>Theorem 7.3</u>: $(F^+,\hat{F}^+)(F^-,\hat{F}^-)(P,\hat{P})(N,\hat{N})(D,\hat{D})$. For every $p \in P(\Omega)$ and $\hat{\pi} \in \hat{\Pi}$, we have

$$\hat{J}_{N,\hat{\pi}}(p) = \int_{\Xi_o} J_{N,\hat{\pi}}(\xi_o)\varphi(p)(d\xi_o).$$

Proof:

$$\int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi)\varphi(p)(d\xi_o)$$

$$= \int_{\Xi_o} \sum_{k=0}^{N-1} \alpha^k \int_{\Xi_o C_o \cdots \Xi_k C_k} \hat{g}_k(\xi_k, I_k)\, d\hat{P}_k(\hat{\pi},p_\xi)\varphi(p)(d\xi_o).$$

By $(F^+)$ or $(F^-)$ if $N < \infty$, by monotone convergence theorem if $N = \infty$ and under (P)(N) and by bounded convergence theorem when $N = \infty$ and under (D)

$$= \sum_{k=0}^{N-1} \alpha^k \int_{\Xi_o} \int_{\Xi_o C_o \ldots \Xi_k C_k} \hat{g}_k(\xi_k, I_k)\, d\hat{P}_k(\hat{\pi}, p_\xi)\, \varphi(p)\,(d\xi_o).$$

By definition of $\hat{P}_k(\hat{\pi}, p_\xi)$ and $\varphi(p)$

$$= \sum_{k=0}^{N-1} \alpha^k \int_{\Xi_o C_o \ldots \Xi_k C_k} \hat{g}_k(\xi_k, I_k)\, d\hat{P}_k[\hat{\pi}, \varphi(p)].$$

By Lemma 7.1 and condition (c) of a sufficient statistic

$$= \sum_{k=0}^{N-1} \alpha^k \int_{D_k} g(\omega_k, I_k)\, dP_k(\hat{\pi}, p)$$

$$= J_{N,\hat{\pi}}(p). \quad \text{QED}$$

Before proceeding to prove the correspondence of the optimal return functions and optimal contracts, a few other results are studied. The following lemma defines the optimal return function for PSI in terms of the initial payoff probability measure p. Next, the relationship between the optimal return functions for the models for $p \in P(\Omega)$ is explored.

<u>Lemma 7.2</u>: $(\hat{F}^+)(\hat{F}^-)(\hat{P})(\hat{N})(\hat{D})$. For every $p \in P(\Omega)$

$$\int_{\Xi_o} \hat{J}_N^*(\xi_o)\, \varphi(p)\,(d\xi_o)$$

$$= \sup_{\hat{\pi} \in \hat{\Pi}} \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\, \varphi(p)\,(d\xi_o).$$

Proof: For $p \in P(\Omega)$ and $\hat{\pi} \in \hat{\Pi}$

$$\int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o) \geq \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varphi(p)(d\xi_o)$$

$$\geq \sup_{\hat{\pi}\in\hat{\Pi}} \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varphi(p)(d\xi_o).$$

Let $\varepsilon > 0$ and let $\hat{\pi}' \in \hat{\Pi}$ be $\varepsilon$-optimal. If

$$\varphi(p)(\{\xi_o | \hat{J}_N^*(\xi_o) = \infty\}) = 0.$$

Then

$$\int_{\Xi_o} \hat{J}_{N,\hat{\pi}'}(\xi_o)\varphi(p)(d\xi_o) \geq \int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o) - \varepsilon$$

$$\therefore \sup_{\hat{\pi}\in\hat{\Pi}} \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varphi(p)(d\xi_o) \geq \int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o)$$

If

$$\varphi(p)(\{\xi_o : \hat{J}_N^*(\xi_o) = \infty\}) > 0.$$

Then

$$\int_{\Xi_o} \hat{J}_{N,\hat{\pi}'}(\xi_o)\varphi(p)(d\xi_o)$$

$$\geq \int_{\{\xi_o : \hat{J}_N^*(\xi_o)<\infty\}} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o) - \varepsilon + \varphi(p)(\{\xi_o : \hat{J}_N^*(\xi_o) = \infty\})/\varepsilon.$$

If

$$\int_{\{\xi_o : \hat{J}_N^*(\xi_o)<\infty\}} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o) = -\infty,$$

then

$$\int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o) = -\infty$$

and

$$\sup_{\hat{\pi}\in\hat{\Pi}} \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varpi(\xi_o) \geq \int_{\Xi_o} \hat{J}_N^*(\xi_o)\varpi(p)(d\xi_o).$$

Otherwise, the right-hand is finite for any $\epsilon > 0$

and

$$\sup_{\hat{\pi}\in\hat{\Pi}} \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varpi(\xi_o) = \infty$$

$$\therefore \int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o) = \sup_{\hat{\pi}\in\hat{\Pi}} \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varphi(p)(d\xi_o).$$

QED

<u>Theorem 7.4</u>: $(F^+,\hat{F}^+)(F^-,\hat{F}^-)(P,\hat{P})(N,\hat{N})(D,\hat{D})$. For every $p \in P(\Omega)$

$$J_N^*(p) \geq \int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o).$$

Proof: By Theorem 7.3

$$J_N^*(p) = \sup_{\pi\in\Pi} J_{N,\pi}(p) \geq \sup_{\hat{\pi}\in\hat{\Pi}} J_{N,\hat{\pi}}(p)$$

$$= \sup_{\hat{\pi}\in\hat{\Pi}} \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varphi(p)(d\xi_o).$$

By Lemma 7.2

$$= \int_{\Xi_o} J_N^*(\xi_o)\varphi(p)(d\xi_o). \quad \text{QED}$$

The next result shows the correspondence of contracts for the two models.

Lemma 7.3: $(F^+,\hat{F}^+)(F^-,\hat{F}^-)(P,\hat{P})(N,\hat{N})(D,\hat{D})$. Let $p \in P(\Omega)$ and $\pi \in \Pi$ be given, there exists $\hat{\pi} \in \hat{\Pi}$ such that

$$J_{N,\pi}(p) = \int_{\Xi_0} \hat{J}_{N,\hat{\pi}}(\xi_0)\varphi(p)(d\xi_0).$$

Proof: Let $p \in P(\Omega)$ and $\pi = (\mu_0,\ldots,\mu_{N-1}) \in \Pi$ be given. For $k = 0,\ldots,N-1$, let $Q_k(\pi,p)$ be the probability measure on $\Xi_k C_k$ defined on measurable rectangles to be

$$Q_k(\pi,p)(\underline{\Xi}_k\ \underline{C}_k) = P_k(\pi,p)(\{\eta_k(p;\emptyset_k) \in \underline{\Xi}_k,\ I_k \in C_k\})$$

$\exists$ a Borel-measurable stochastic kernel $\hat{\mu}_k(dI_k|\xi_k)$ on $C_k$ given $\Xi_k$ such that for every Borel set $B \subset \Xi_k C_k$,

$$Q_k(\pi,p)(B) = \int_{\Xi_k C_k} \hat{\mu}_k(B_{\xi_k}|\xi_k)\,dQ_k(\pi,p).$$

In particular,

$$\begin{aligned} 1 &= P_k(\pi,p)(\{\emptyset_k,I_k) \in \Gamma_k\}) \\ &= P_k(\pi,p)(\{[\eta_k(p;\emptyset_k),I_k] \in \hat{\Gamma}_k\}) \\ &= Q_k(\pi,p)(\hat{\Gamma}_k) \\ &= \int_{\Xi_k C_k} \hat{\mu}_k(\hat{U}_k(\xi_k)|\xi_k)\,dQ_k(\pi,p). \end{aligned}$$

By altering $\hat{\mu}_k(dI_k|\xi_k)$ on a set of measure zero if necessary, we may assume that

$$Q_k(\pi,p)(B) = \int_{\Xi_k C_k} \hat{\mu}_k(B_{\xi_k}|\xi_k)\,dQ_k(\pi,p)$$

and $\hat{\mu}_k(\hat{U}_k(\xi_k)|\xi_k) = 1$ for every $\xi_k \in \Xi_k$. Let $\hat{\pi} = (\hat{\mu}_o,\ldots,\hat{\mu}_{N-1})$. Then $\hat{\pi}$ is a contract for PSI. Since

$$\varphi(p)(\underline{\Xi}_o) = \int_\Omega s_o(\{z_o : \eta_o(p;z_o) \in \underline{\Xi}_o\}|\omega_o)p(d\omega_o)$$
$$\forall\ \underline{\Xi}_o \in \mathcal{B}_{\Xi_o}\ .$$

Clearly, the marginal of $Q_o(\pi,p)$ on $\Xi_o$ is $\varphi(p)$ and for $k = 0$, by Lemma 7.1

$$Q_o(\pi,p)(\underline{\Xi}_o\ \underline{C}_o) = \hat{P}_o[\hat{\pi},\varphi(p)](\{\xi_o \in \underline{\Xi}_o,\ I_o \in \underline{C}_k\}).$$

Now assume that

$$Q_k(\pi,p)(\underline{\Xi}_k\ \underline{C}_k) = \hat{P}_k[\hat{\pi},\varphi(p)](\{\xi_k \in \underline{\Xi}_k,\ I_k \in \underline{C}_k\}).$$

Then

$$Q_{k+1}(\pi,p)(\underline{\Xi}_{k+1}\ \underline{C}_{k+1})$$
$$= \int_{\underline{\Xi}_{k+1}\underline{C}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|\xi_{k+1})\,dQ_{k+1}(\pi,p)$$
$$= \int_{\{\eta_{k+1}(p;\emptyset_{k+1})\in\underline{\Xi}_{k+1}\}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|\eta_{k+1}(p;\emptyset_{k+1}))\,dP_{k+1}(\pi,p).$$

By condition (b) of Definition SS

$$= \int_{D_k} \int_{\underline{\Xi}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1} | \xi_{k+1}) \hat{t}(d\xi_{k+1} | \eta_{k+1}(p; \emptyset_{k+1}), I_k) \, dP_k(\pi, p)$$

$$= \int_{\underline{\Xi}_k \underline{C}_k} \int_{\underline{\Xi}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1} | \xi_{k+1}) \hat{t}(d\xi_{k+1} | \xi_k, I_k) \, dQ_k(\pi, p)$$

$$= \int_{\Xi_o C_o \cdots \Xi_k C_k} \int_{\underline{\Xi}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1} | \xi_{k+1}) \hat{t}(d\xi_{k+1} | \xi_i, I_k) \, d\hat{P}_k[\hat{\pi}, \varphi(p)]$$

$$= \hat{P}_{k+1}[\hat{\pi}, \varphi(p)](\{\xi_{k+1} \in \underline{\Xi}_{k+1}, I_k \in \underline{C}_{k+1}\})$$

$\therefore$ For $\underline{\Xi}_k \in \mathcal{B}_{\Xi_k}$, $\underline{C}_k \in \mathcal{B}_C$, $k = 0, \ldots, N-1$

$$Q_k(\pi, p)(\underline{\Xi}_k \, \underline{C}_k) = \hat{P}_k[\hat{\pi}, \varphi(p)](\{\xi_k \in \underline{\Xi}_k, \; I_k \in \underline{C}_k\})$$

$$= P_k(\pi, p)(\{\eta_k(p; \emptyset_k) \in \underline{\Xi}_k, \; I_k \in \underline{C}_k\}).$$

By Theorem 7.3

$$J_{N,\pi}(p) = \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o) \varphi(p)(d\xi_o). \qquad \text{QED}$$

Definition: Let $q \in P(\Xi_o)$ and $\varepsilon > 0$, a contract $\hat{\pi} \in \hat{\Pi}$ is said to be weakly $q-\varepsilon$-optimal if

$$\int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o) q(d\xi_o)$$

$$\geq \begin{cases} \int_{\Xi_o} \hat{J}^*_N(\xi_o) q(d\xi_o) - \varepsilon & \text{if } \int_{\Xi_o} \hat{J}^*_N(\xi_o) q(d\xi_o) < \infty \\ \frac{1}{\varepsilon} & \text{if } \int_{\Xi_o} \hat{J}^*_N(\xi_o) q(d\xi_o) = \infty \end{cases}$$

The contact $\hat{\pi}$ is said to be q-optimal if

$$q(\{\xi_o \in \Xi_o : \hat{J}_{N,\hat{\pi}}(\xi_o) = \hat{J}^*_N(\xi_o)\}) = 1.$$

Lemma 7.4: Let $J : \Omega \to R^*$ be upper semianalytic. Then for $\varepsilon > 0$, there exists $\mu \in U(C|\Omega)$ such that $T_\mu(J)(\omega) \geq T(J)(\omega) - \varepsilon$ for every $\omega \in \Omega$, where $T(J)(\omega) - \varepsilon$ may be $\infty$.

Proof: By Theorem 6.14, there are universally measurable selectors $\mu_m : \Omega \to C$ such that for $m = 1,2,...$ and $\omega \in \Omega$, $\mu_m \in U(\omega)$ and

$$T_{\mu_m}(J)(\omega) \geq \begin{cases} T(J)(\omega) - \varepsilon & \text{if } T(J)(\omega) < \infty \\ 2^m & \text{if } T(J)(\omega) = \infty \end{cases}$$

Let $\mu(dI|\omega)$ assign mass one to $\mu_1(\omega)$ if $T(J)(\omega) < \infty$ and assign mass $1/2^m$ to $\mu_m(\omega)$, $m = 1,2,...$ if $T(J)(\omega) = \infty$. For each $\underline{C} \in \mathcal{B}_C$

$$\mu(\underline{C}|\omega) = \begin{cases} \chi_{\underline{C}}[\mu_1(\omega)] & \text{if } T(J)(\omega) < \infty \\ \sum_{m=1}^{\infty} \frac{1}{2^m} \chi_{\underline{C}}[\mu_m(\omega)] & \text{if } T(J)(\omega) = \infty \end{cases}$$

is a universally measurable function of $\omega$.

$\therefore$ $\mu$ is a universally measurable stochastic kernel with the desired properties. QED

Lemma 7.5: $(F^-)$ If $J_o : \Omega \to R^*$ is identically zero, then $T^k(J_o)(\omega) < \infty$ for every $\omega \in \Omega$, $K = 1,...,N$.

Proof: Suppose for some $K \leq N$ and $\omega_o \in \Omega$ that for every $\omega \in \Omega$

$$T^j(J_o)(\omega) < \infty \quad j = 0,...,K-1$$

and

$$T^k(J_o)(\omega_o) = \infty$$

$\exists$ universally measurable selectors $\mu_j : \Omega \to C$ $j = 1,\ldots,K-1$ such that $\mu_j(\omega) \in U(\omega)$ and

$$(T\mu_{K-j}T^{j-1})(J_o)(\omega) \geqq T^j(J_o)(\omega) - 1$$

$$j = 1,\ldots,K-1 \quad \forall\ \omega \in \Omega.$$

Then

$$(T\mu_1 \ldots T\mu_{K-1})(J_o) \geqq (T\mu_1 \ldots T\mu_{K-2})[T(J_o) - 1]$$

$$\geqq (T\mu_1 \ldots T\mu_{K-3})[T^2(J_o) - 1 - \alpha]$$

$$\geqq T^{K-1}(J_o) - (1 + \alpha + \ldots + \alpha^{K-2}).$$

By Lemma 7.4, there is a stochastic kernel $\mu_o \in U(C|\Omega)$ such that $(T\mu_o T^{K-1})(J_o)(\omega_o) = \infty$. Then

$$(T\mu_o T\mu_1 \ldots T\mu_{K-1})(J_o)(\omega_o) \geqq T\mu_o[T^{K-1}(J_o) - 1 - \alpha - \ldots - \alpha^{K-2}](\omega_o)$$

$$= \infty.$$

Choose any $\mu \in U(C|\Omega)$, let $\pi = (\mu_o,\ldots,\mu_{K-1},\mu_1,\ldots,\mu)$ so that $\pi \in \Pi$

$$\sum_{k=0}^{K-1} \alpha^k \int g dq_k(\pi,p_{\omega_o}) = J_{K,\pi}(\omega_o)$$

$$= (T\mu_o \ldots T\mu_{K-1})(J_o)(\omega_o) = \infty$$

$\therefore$ for some $k \leqq K-1$ $\int g dq_k(\pi,p_{\omega_o}) = \infty$.

This contradicts $(F^-)$. QED

<u>Theorem 7.5</u>: $(F^+, \hat{F}^+)(F^-, \hat{F}^-)(P, \hat{P})(N, \hat{N})(D, \hat{D})$

$$J_N^*(p) = \int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o) \quad \forall\ p \in P(\Omega).$$

Furthermore, if $\hat{\pi}$ is optimal, $\varphi(p)$-optimal or weakly $\varphi(p)$ - $\varepsilon$-optimal for PSI, then $\hat{\pi}$ is optimal, optimal at $p$ or $\varepsilon$-optimal at $p$ respectively for ISI If $\hat{\pi}$ is $\varepsilon$-optimal for PSI and $(F^-, \hat{F}^-)$, $(N, \hat{N})$ or $(D, \hat{D})$ holds, then $\hat{\pi}$ is also $\varepsilon$-optimal for ISI.

Proof:
$$\begin{aligned} J_N^*(p) &= \sup_{\pi\in\Pi} J_{N,\pi}(p) \\ &\geq \sup_{\hat{\pi}\in\hat{\Pi}} \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varphi(p)(d\xi_o) \\ &= \int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o). \end{aligned}$$

By Lemma 7.3, then

$$J_N^*(p) = \int_{\Xi_o} \hat{J}_N^*(\xi_o)\varphi(p)(d\xi_o) \quad \forall\ p \in P(\Omega).$$

Let $\hat{\pi}$ be $\varepsilon$-optimal for PSI. Clearly, under $(N, \hat{N})$, $(D, \hat{D})$

$$\hat{J}_N^*(\xi_o) < \infty \quad \forall\ \xi_o \in \Xi_o$$

$$\therefore\ \hat{J}_{N,\hat{\pi}}(\xi_o) \geq \hat{J}_N^*(\xi_o) - \varepsilon \quad \forall\ \xi_o \in \Xi_o.$$

Under $(F^-, \hat{F}^-)$, by Lemma 7.5, $\hat{J}_N^*(\xi_o) < \infty$, and again the above holds. From Theorem 7.3 and the first part of this theorem

$$J_{N,\hat{\pi}}(p) = \int_{\Xi_o} \hat{J}_{N,\hat{\pi}}(\xi_o)\varphi(p)(d\xi_o)$$

$$\geq \int_{\Xi_o} \hat{J}^*_N(\xi_o)\varphi(p)(d\xi_o) - \varepsilon$$

$$= J^*_N(p) - \varepsilon$$

$\therefore\ \hat{\pi}$ is $\varepsilon$-optimal for ISI.

Similar arguments hold for the rest of this theorem. QED

## 7.5* Existence of Statistics Sufficient for Contracting

This section presents two statistics which are shown to be sufficient for contracting. The first statistic is derived by a process called filtering. To aid the discussion of filtering, there is the following basic lemma.

Lemma 7.6: For the ISI model, there exist Borel-measurable stochastic kernels $r_o(d\omega_o|p;z_o)$ on $\Omega$ given $P(\Omega)Z$ and $r(d\omega|p;I,z)$ on $\Omega$ given $P(\Omega)CZ$ which satisfy

(A) $$\int_{\underline{\Omega}_o} s_o(\underline{Z}_o|\omega_o)p(d\omega_o) = \int_{\Omega}\int_{\underline{Z}_o} r_o(\underline{\Omega}_o|p;z_o)s_o(dz_o|\omega_o)p(d\omega_o)$$

$$\forall\ \underline{\Omega}_o \in \mathcal{B}_\Omega,\ \underline{Z}_o \in \mathcal{B}_Z,\ p \in P(\Omega)$$

(B) $$\int_{\underline{\Omega}} s(\underline{Z}|I,\omega)p(d\omega) = \int_{\Omega}\int_{\underline{Z}} r(\underline{\Omega}|p;I,z)s(dz|I,\omega)p(dx)$$

$$\forall\ \underline{\Omega} \in \mathcal{B}_\Omega,\ \underline{Z} \in \mathcal{B}_Z,\ p \in P(\Omega),\ I \in C.$$

Proof: For fixed $(p;I) \in P(\Omega)C$, define a probability measure $q$ on $\Omega Z$ by specifying its values on measurable rectangles to be (Theorem 6.3)

$$q(\underline{\Omega}\,\underline{Z}\,|p;I) = \int_{\underline{\Omega}} s(\underline{Z}|I,\omega)p(d\omega).$$

By Theorem 6.1 and 6.4, $q(d(\omega,z)|p;I)$ is a Borel-measurable stochastic kernel on $\Omega Z$ given $P(\Omega)C$. By Corollary 6.2.1, this stochastic kernel can be decomposed into its marginal on $Z$ given $P(\Omega)C$ and a Borel-measurable stochastic kernel $r(d\omega|p;I,z)$ on $\Omega$ given $P(\Omega)CZ$ such that the theorem holds.

The existence of $r_o(d\omega_o|p;z_o)$ is proved in a similar manner. QED

The system starts with the initial wealth outcome $\omega_o$ which has a priori distribution $p$. After $Z_o$ is observed, the distribution is "up-dated". The up-dated distribution is called a posteriori distribution and will be shown in the next lemma to be just $r_o(d\omega_o|p;z_o)$. At the $k^{th}$ stage, $k \geq 1$, there will be some a priori distribution $p_k'$ of $\omega_k$ based on $\emptyset_{k-1} = (z_o,I_o,\ldots,I_{k-2},z_{k-1})$. Then, a contract $I_{k-1}$ is negotiated, some $z_k$ is observed, and an a posteriori distribution of $\omega_k$ conditioned on $(\emptyset_{k-1},I_{k-1},z_k)$ is computed. This distribution is just $r(d\omega|p_k';I_{k-1},z_k)$.

The process of passing from an a priori to an a posterior distribution in this manner is called filtering. The process is formalized in the following definitions.

Definition: The function $\bar{f} : P(\Omega)C \to P(\Omega)$ is defined by

(C) $\bar{f}(p;I)(\underline{\Omega}) = \int t(\underline{\Omega}|\omega,I)p(d\omega) \quad \forall\ \underline{\Omega} \in \mathcal{B}_\Omega$

is called the one stage prediction equation.

By Theorems 6.1 and 6.4, $\bar{f}$ is Borel-measurable.

Definition: Given a sequence $\emptyset_k \in \Phi_k$ such that $\emptyset_{k+1} = (\emptyset_k, I_k, z_{k+1})$, $k = 0,\ldots,N-2$ and given $p \in P(\Omega)$, define recursively

(D) $p_o(p;\emptyset_o) = r_o(d\omega_o|p;z_o)$

(E) $p_{k+1}(p;\emptyset_{k+1}) = r(d\omega|\bar{f}[p_k(p;\emptyset_k),I_k];I_k,z_{k+1})$

$k = 0,\ldots,N-2.$

Thus defined, for each $k$, $p_k : P(\Omega)\Phi_k \to P(\Omega)$ is Borel-measurable. Equations (A)-(E) are called the filtering equations corresponding to the ISI model.

<u>Lemma 7.7</u>: Let the ISI modle be given. For any $p \in P(\Omega)$, $\pi = (\mu_o,\ldots,\mu_{N-1})$, $\pi \in \Pi$ and $\underline{\Omega}_k \in \mathcal{B}_\Omega$, then

$$p_k(\pi,p)[\omega_k \in \underline{\Omega}_k|\emptyset_k] = p_k(p;\emptyset_k)(\underline{\Omega}_k)$$

for $P_k(\pi,p)$ almost every $\emptyset_k$, $k = 0,\ldots,N-1$.

Proof: For any $\underline{\Omega}_o \in \mathcal{B}_\Omega$ and $\underline{Z}_o \in \mathcal{B}_Z$

$$\int_{\{z_o \in \underline{Z}_o\}} p_o(p;z_o)(\underline{\Omega}_o)\,dP_o(\pi,p) = \int_{\{z_o \in \underline{Z}_o\}} r_o(\underline{\Omega}_o|p;z_o)\,dP_o(\pi,p)$$

$$= \int_\Omega \int_{\underline{Z}_o} r_o(\underline{\Omega}_o|p;z_o)\,s_o(dz_o|\omega_o)\,p(d\omega_o)$$

$$= \int_{\underline{\Omega}_o} s_o(\underline{Z}_o|\omega_o)\,p(d\omega_o)$$

$$= P_o(\pi,p)(\{\omega_o \in \underline{\Omega}_o,\ z_o \in \underline{Z}_o\}).$$

The theorem follows for $k = 0$ by the definition of conditional probability. Now assume the theorem holds for $k$. For any $\underline{\Phi}_k \in \mathcal{B}_{\Phi_k}$, $\underline{C}_k \in \mathcal{B}_C$, $\underline{Z}_{k+1} \in \mathcal{B}_Z$ and $\underline{\Omega}_{k+1} \in \mathcal{B}_\Omega$. Then

$$\int_{\{\phi_k \in \underline{\Phi}_k, I_k \in \underline{C}_k, z_{k+1} \in \underline{Z}_{k+1}\}} p_{k+1}(p;\phi_k,I_k,z_{k+1})(\underline{\Omega}_{k+1})\,dP_{k+1}(\pi,p)$$

$$= \int_{\{\phi_k \in \Phi_k\}} \int_{\underline{C}_k} \int_{\underline{\Omega}_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p;\phi_k,z_{k+1})(\underline{\Omega}_{k+1})\, s(dz_{k+1}|I_k,\omega_{k+1})\,t(d\omega_{k+1}|\omega_k,I_k)\, \mu_k(dI_k|p;\phi_k)\,dP_k(\pi,p)$$

$$= \int_{\{\phi_k \in \underline{\Phi}_k\}} \int_{\underline{\Omega}_k} \int_{\underline{C}_k} \int_{\underline{\Omega}_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p;\phi_k,J_k,z_{k+1})(\underline{\Omega}_{k+1})\, s(dz_{k+1}|I_k,\omega_{k+1})\,t(d\omega_{k+1}|\omega_k,I_k)\, \mu_k(dI_k|p;\phi_k)[p_k(p;\phi_k)(d\omega_k)]dP_k(\pi,p)$$

$$= \int_{\{\phi_k \in \Phi_k\}} \int_{\underline{C}_k} \int_{\underline{\Omega}_k} \int_{\underline{\Omega}_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p;\phi_k,I_k,z_{k+1})(\underline{\Omega}_{k+1})\, s(dz_{k+1}|I_k,\omega_{k+1})\,t(d\omega_{k+1}|\omega_k,I_k)\, [p_k(p;\phi_k)(d\omega_k)]\mu_k(dI_k|p;\phi_k)\,dP_k(\pi,p)$$

$$= \int_{\{\emptyset_k \in \Phi_k\}} \int_{\underline{C}_k} \int_{\underline{\Omega}_{k+1}} \int_{\underline{Z}_{k+1}} r(\underline{\Omega}_{k+1} | \bar{f}[p_k(p;\emptyset_k), I_k]; I_k, z_{k+1}) \; s(dz_{k+1} | I_k, \omega_{k+1}) \bar{f}[p_k(p;\emptyset_k), I_k](d\omega_{k+1}) \; \mu_k(dI_k | p;\emptyset_k) \, dP_k(\pi,p)$$

$$= \int_{\{\emptyset_k \in \Phi_k\}} \int_{\underline{C}_k} \int_{\underline{\Omega}_{k+1}} s(\underline{Z}_{k+1} | I_k, \omega_{k+1}) \bar{f}[p_k(p;\emptyset_k), I_k](d\omega_{k+1}) \; \mu_k(dI_k | p;\emptyset_k) \, dP_k(\pi,p)$$

$$= \int_{\{\emptyset_k \in \Phi_k\}} \int_{\underline{C}_k} \int_{\underline{\Omega}_k} \int_{\underline{\Omega}_{k+1}} s(\underline{Z}_{k+1} | I_k, \omega_{k+1}) t(d\omega_{k+1} | \omega_k) [p_k(p;\emptyset_k)(d\omega_k)] \; \mu_k(dI_k | p;\emptyset_k) \, dP_k(\pi,p)$$

$$= \int_{\{\emptyset_k \in \Phi_k\}} \int_{\underline{\Omega}_k} \int_{\underline{C}_k} \int_{\underline{\Omega}_{k+1}} s(\underline{Z}_{k+1} | I_k, \omega_{k+1}) t(d\omega_{k+1} | \omega_k, I_k) \; \mu_k(dI_k | p;\emptyset_k) [p_k(p;\emptyset_k)(d\omega_k)] dP_k(\pi,p)$$

$$= \int_{\{\emptyset_k \in \Phi_k\}} \int_{\underline{C}_k} \int_{\underline{\Omega}_{k+1}} s(\underline{Z}_{k+1} | I_k, \omega_{k+1}) t(d\omega_{k+1} | \omega_k, I_k) \; \mu_k(dI_k | p;\emptyset_k) \, dP_k(\pi,p)$$

$$= P_{k+1}(\pi,p)(\{\emptyset_k \in \Phi_k, \; I_k \in \underline{C}_k, \; \omega_{k+1} \in \underline{\Omega}_{k+1}, \; z_{k+1} \in \underline{Z}_{k+1}\}).$$

By the definition of conditional probability

$$P_{k+1}(\pi,p)[\omega_{k+1} \in \underline{\Omega}_{k+1} | \emptyset_{k+1}] = p_{k+1}(p;\emptyset_{k+1})(\underline{\Omega}_{k+1})$$

for $P_{k+1}(\pi,p)$ almost every $\emptyset_k$. QED

<u>Theorem 7.6</u>: For the ISI model, assume that $U_k(\omega) = C$ for every $\omega \in \Omega$ and $k = 0,\ldots,N-1$. Then the sequence $[p_0(p;\emptyset_0),\ldots,p_{N-1}(p;\emptyset_{N-1})]$ defined by (D) and (E) is a sufficient statistic and the resulting perfect state information model is stationary.

Proof: Let $\Xi_k$ in the Definition SS be $P(\Omega)$, $k = 0,\ldots,N-1$. Since $p_k : P(\Omega)\Phi_k \to P(\Omega)$ is Borel-measurable for each $k$ $(p_0,\ldots,p_{N-1})$ is a statistic. Let $\hat{\Gamma}_k = P(\Omega)C$, $k = 0,\ldots,N-1$. Then condition (a) of Definition SS is satisfied. For $\xi \in P(\Omega)$, $I \in C$ and $\underline{\Xi} \in \mathcal{B}_{P(\Omega)}$, define

$$\underline{Z}(\xi,I,\underline{\Xi}) = \{z \in Z : r[d\omega|\bar{f}(\xi,I);I,z] \in \underline{\Xi}\}$$

$$\hat{t}(\underline{\Xi}|\xi,I) = \int_\Omega \int_\Omega s[\underline{Z}(\xi,I,\underline{\Xi})|I,\omega']t(d\omega'|\omega,I)\xi(d\omega).$$

But $\underline{Z}(\xi,I,\underline{\Xi})$ is the $(\xi,I)$-section of the inverse image of $\underline{\Xi}$ under a Borel-measurable function. The stochastic kernel

$$\lambda(\underline{Z}|\omega,I) = \int_\Omega s(\underline{Z}|I,\omega')t(d\omega'|\omega,I)$$

is Borel-measurable by Theorems 6.1 and 6.4

$$\therefore \Lambda(\underline{Z}|\xi,I) = \int_\Omega \int_\Omega s(\underline{Z}|I,\omega')t(d\omega'|\omega,I)\xi(d\omega)$$

$$= \int_\Omega \lambda(\underline{Z}|\omega,I)\xi(d\omega)$$

is Borel-measurable by the same theorems. By Theorem 6.1 and Corollary 6.1.1, $\hat{t}(d\xi'|\xi,I)$ is a Borel-measurable stochastic kernel on $P(\Omega)$ given $P(\Omega)C$. For $\pi \in \Pi$, $p \in P(\Omega)$, $\underline{\Xi} \in \mathcal{B}_{P(\Omega)}$ and $k = 0,1,\ldots,N-2$, by Lemma 7.7

$$P_{k+1}(\pi,p)[p_{k+1}(p;\phi_{k+1}) \in \underline{\Xi} | p_k(p;\phi_k) = \bar{\xi}_k,\ I_k = \bar{I}_k\}$$

$$= P_{k+1}(\pi,p)[z_{k+1} \in \underline{Z}(\bar{\xi}_k,\bar{I}_k,\underline{\Xi} | p_k(p;\phi_k) = \bar{\xi}_k,\ I_k = \bar{I}_k]$$

$$= E\{P_{k+1}(\pi,p)[z_{k+1} \in \underline{Z}(\bar{\xi}_k,\bar{I}_k,\underline{\Xi}) | \phi_k, I_k] | p_k(p;\phi_k) = \bar{\xi}_k, I_k = \bar{I}_k\}$$

$$= E\{\int_\Omega \int_\Omega s[\underline{Z}(\bar{\xi}_k,\bar{I}_k,\underline{\Xi}) | I_k, \omega_{k+1}]$$

$$t(d\omega_{k+1} | \omega_k, I_k)[p_k(p;\phi_k)(d\omega_k)] | p_k(p;\phi_k) = \bar{\xi}_k, I_k = \bar{I}_k\}$$

$$= \hat{t}(\underline{\Xi} | \bar{\xi}_k, \bar{I}_k)$$

for $P_k(\pi,p)$ almost every $(\bar{\xi}_k,\bar{I}_k)$ where the expectations are with respect to $P_{k+1}(\pi,p)$.

∴ Condition (B) is satisfied.

For $\pi \in \Pi$, $p \in P(\Omega)$, and $k = 0,\ldots,N-1$. By Lemma 7.7

$$E[g(\omega_k,I_k) | p_k(p;\phi_k) = \bar{\xi}_k,\ I_k = \bar{I}_k]$$

$$= E\{E[g(\omega_k,I_k) | \phi_k, I_k] | p_k(p;\phi_k) = \bar{\xi}_k,\ I_k = \bar{I}_k\}$$

$$= E\{\int_\Omega g(\omega_k,I_k) p_k(p;\phi_k)(d\omega_k) | p_k(p;\phi_k) = \bar{\xi}_k,\ I_k = \bar{I}_k\}$$

$$= \int_\Omega g(\omega_k,\bar{I}_k)\bar{\xi}_k(d\omega_k)$$

for $P_k(\pi,p)$ almost every $(\bar{\xi}_k,\bar{I}_k)$ where the expectations are with respect to $P_k(\pi,p)$. The function $\hat{g} : P(\Omega)C \to R^*$ defined by

$$\hat{g}(\bar{\xi},\bar{I}) = \int_\Omega g(\omega,\bar{I})\bar{\xi}(d\omega)$$

is upper semianalytic by Theorem 6.12

∴ $\hat{g}$ satisfies condition (C). QED

Theorem 7.7: Let the ISI model be given. The sequence of identity mappings on $P(\Omega)\Phi_k$, $k = 0,\ldots,N-1$ is a sufficient statistic.

Proof: Let $\Xi_k$ in Definition SS by $P(\Omega)\Phi_k$, $k = 0,\ldots,N-1$ and let $\eta_k$ be the identity mapping on $P(\Omega)\Phi_k$. Then $(\eta_0,\ldots,\eta_{N-1})$ is a statistic. Let $\hat{\Gamma}_k = P(\Omega)\Gamma_k$, $k = 0,\ldots,N-1$. Condition (A) is satisfied. If $\underline{\Xi}_{k+1} \in \mathcal{B}_{P(\Omega)\Phi_{k+1}}$, $\bar{\xi}_k \in P(\Omega)\Phi_k$, and $\bar{I}_k \in C_k$, define

$$(\underline{\Xi}_{k+1})_{(\bar{\xi}_k,\bar{I}_k)} = \{z_{k+1} \in Z : (\bar{p};\bar{z}_0,\bar{I}_0,\ldots,\bar{I}_{k-1},\bar{z}_k,\bar{I}_k,z_{k+1}) \in \underline{\Xi}_{k+1}\}$$

where $\bar{\xi}_k = (\bar{p};\bar{z}_0,\bar{I}_0,\ldots,\bar{I}_{k+1},\bar{z}_k\}$. Now, define for $k = 0,\ldots,N-2$, the stochastic kernel $\hat{t}_k(d\xi_{k+1}|\bar{\xi}_k,\bar{I}_k)$ on $P(\Omega)\Phi_{k+1}$ given $P(\Omega)\Phi_k C$ by

$$\hat{t}_k(\underline{\Xi}_{k+1}|\bar{\xi}_k,\bar{I}_k) = \int_{\Omega_{k+1}} s[(\underline{\Xi}_{k+1})_{(\bar{\xi}_k,\bar{I}_k)}|\bar{I}_k,\omega_{k+1}] \, t(d\omega_{k+1}|\omega_k,\bar{I}_k)p_k(\bar{\xi}_k)(d\omega_k)$$

$$\forall \ \underline{\Xi}_{k+1} \in \mathcal{B}_{P(\Omega)\Phi_{k+1}}$$

where $p_k(\bar{\xi}_k)$ is as defined by (D) and (E). Using similar arguments as in Theorem 7.6, it can be shown that $\hat{t}_k$ is Borel-measurable. By Lemma 7.7

$$P_{k+1}(\pi,p)[\eta_{k+1}(p;\emptyset_{k+1}) \in \underline{\Xi}_{k+1} | \eta_k(p;\emptyset_k) = \bar{\xi}_k,\ I_k = \bar{I}_k]$$

$$= P_{k+1}(\pi,p)[(\bar{\xi}_k,\bar{I}_k,z_{k+1}) \in \underline{\Xi}_{k+1}]$$

$$= \int_{\Omega_{k+1}} s[(\underline{\Xi}_{k+1})_{(\bar{\xi}_k,\bar{I}_k)} | \bar{I}_k, \omega_{k+1}] t(d\omega_{k+1} | \omega_k, \bar{I}_k) p_k(\bar{\xi}_k)(d\omega_k)$$

$$= \hat{t}(\underline{\Xi}_{k+1} | \bar{\xi}_k, \bar{I}_k)$$

for $P_k(\pi,p)$ almost every $(\bar{\xi}_k,\bar{I}_k)$. Condition (B) is satisfied. For $k = 0,\ldots,N-1$, define

$\hat{g}_k : P(\Omega)\Phi_k C \to R^*$ by

$$\hat{g}_k(\bar{\xi}_k,\bar{I}_k) = \int_{\Omega_k} g(\omega_k,\bar{I}_k) p_k(\bar{\xi}_k)(d\omega_k).$$

By Theorem 6.12, $\hat{g}_k$ is upper semianalytic for each $k$

For $p \in P(\Omega)$, $\pi \in \Pi$ and $k = 0,\ldots,N-1$ from Lemma 7.7

$$E[g(\omega_k,I_k) | \eta(p;\emptyset_k) = \bar{\xi}_k,\ I_k = \bar{I}_k]$$

$$= \int_{\Omega_k} g(\omega_k,\bar{I}_k) p_k(\bar{\xi}_k)(d\omega_k)$$

$$= \hat{g}_k(\bar{\xi}_k,\bar{I}_k)$$

for $P_k(\pi,p)$ almost every $(\bar{\xi}_k,\bar{I}_k)$ where the expectation is with respect to $P_k(\pi,p)$

∴ Condition (C) is satisfied. QED

# APPENDICES

## APPENDIX ONE

This appendix develops the rudiments of probability measures on Borel spaces. It tries to summarize the basic facts of the subject. All results in this section are available in the literature. They are collected here for easy reference.

It is understood that throughout the appendix $X$ is a metrizable topological space with $\mathcal{B}_X$ as the Borel $\sigma$-algebra on $X$. The space of probability measures on $(X,\mathcal{B}_X)$ is denoted by $P(X)$. $C(X)$ is the Banach space of bounded, real-valued continuous functions on $X$ with the supremum norm for any metric $d$ on $X$ consistent with its topology. $U_d(X)$ is the space of bounded, real-valued functions on $X$ which are uniformly continuous with respect to $d$. A probability measure $p \in P(X)$ determines a linear functional $l_p : C(X) \to R$ defined by $l_p(f) = \int f dp$ and conversely, a function $f \in C(X)$ determines a real-valued function

$\theta_f : P(X) \to R$ defined by $\theta_f(p) = \int f dp$. These relationships and the metrizability of the space $X$ enable one to show several properties of $P(X)$. In particular, it can be proved that there is a natural topology on $P(X)$, the weakest topology with respect to which every mapping of the form of $\theta$ is continuous, under which $P(X)$ is a Borel space whenever $X$ is a Borel space.

Definition: Let $X$ be a metrizable space. A probability measure $p \in P(X)$ is said to be regular if for every $B \in \mathcal{B}_X$,

$$p(B) = \sup\{p(F) : F \subset B,\ F \text{ closed}\}$$
$$= \inf\{p(G) : B \subset G,\ G \text{ open}\}.$$

Theorem A.1: Let $X$ be a metrizable space. Every probability measure in $P(X)$ is regular.

Proof: Let $p \in P(X)$ be given and $\mathcal{S}$ be the collection of $B \in \mathcal{B}_X$ such that

$$p(B) = \sup\{p(F) : F \subset B,\ F \text{ closed}\}$$
$$= \inf\{p(G) : B \subset G,\ G \text{ open}\}.$$

Let $H \subset X$ be open, $\exists$ an increasing sequence of closed sets $\{F_n\}$ such that $H = \bigcup_{n=1}^{\infty} F_n$. Thus

$$\begin{aligned}&\inf\{p(G) : H \subset G,\ G \text{ open}\}\\ &= p(H)\\ &= \lim_{n\to\infty} p(F_n)\\ &\leqq \sup\{p(F) : F \subset H,\ F \text{ closed}\}\\ &\leqq p(H)\end{aligned}$$

$\therefore$ $p(H) = \sup\{p(F) : F \subset H,\ F \text{ closed}\}$ and $H \in \mathcal{B}$

$\therefore$ $\mathcal{B}$ contains every open subset of $X$.

Suppose $B \in \mathcal{B}$, then

$$\begin{aligned}p(B^c) = 1 - p(B) &= 1 - \sup\{p(F) : F \subset B,\ F \text{ closed}\}\\ &= \inf\{p(G) : B^c \subset G,\ G \text{ open}\}.\end{aligned}$$

Similarly

$$p(B^c) = \sup\{p(F) : F \subset B^c,\ F \text{ closed}\}$$

$\therefore$ $\mathcal{B}$ is closed under complementation.

Let $\{B_n\} \subset \mathcal{B}$. Choose $\varepsilon > 0$ and $F_n \subset B_n \subset G_n$ such that $F_n$ is closed, $G_n$ is open and $p(G_n - F_n) \leqq \varepsilon/2^n$. Then

$$\begin{aligned}\bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} G_n &= \left(\bigcup_{n=1}^{\infty} F_n\right) \cup \left[\bigcup_{n=1}^{\infty} (G_n - F_n)\right]\\ &\subset \left(\bigcup_{n=1}^{\infty} B_n\right) \cup \left[\bigcup_{n=1}^{\infty} (G_n - F_n)\right].\end{aligned}$$

Thus $p\left(\bigcup_{n=1}^{\infty} G_n\right) \leqq p\left(\bigcup_{n=1}^{\infty} B_n\right) + \varepsilon$. This implies

$$p(\bigcup_{n=1}^{\infty} B_n) = \inf\{p(G) : \bigcup_{n=1}^{\infty} B_n \subset G, \ G \text{ open}\}.$$

Also $p(\bigcup_{n=1}^{\infty} B_n) \leqq p(\bigcup_{n=1}^{\infty} F_n) + \varepsilon$ and

$p(\bigcup_{n=1}^{\infty} B_n) \leqq p(\bigcup_{n=1}^{N} F_n) + 2\varepsilon$ for $N$ sufficiently large.

However, the finite union $\bigcup_{n=1}^{N} F_n$ is a closed subest of $\bigcup_{n=1}^{\infty} B_n$. Thus

$$p(\bigcup_{n=1}^{\infty} B_n) = \sup\{p(F) : F \subset \bigcup_{n=1}^{\infty} B_n, \ F \text{ closed}\}$$

$\therefore$ $\mathcal{S}$ is closed under countable unions.

Hence $\mathcal{S}$ is a $\sigma$-algebra and $\mathcal{S} = \mathcal{B}_X$. QED

<u>Theorem A.2</u>: Let $X$ be a metrizable space and $d$ a metric on $X$ consistent with its topology. If $p_1, p_2 \in P(X)$ and

$$\int g dp_1 = \int g dp_2 \quad \forall\, g \in U_d(X)$$

then $p_1 = p_2$.

Proof: Let $F$ be any closed proper subset of $X$ and $G_n = \{x \in X : d(x,F) < \frac{1}{n}\}$.

For sufficiently large $n$, $F$ and $\sim G_n$ are disjoint non-empty closed sets for which

$\inf_{x \in F, y \in \sim G_n} d(x,y) > 0$. By Urysohn's Lemma,

$\exists\, f_n \in U_d(X)$ such that $f_n(x) = 0$ for $x \in \sim G_n$,

$f_n(x) = 1$ for $x \in F$ and $0 \leqq f_n(x) \leqq 1 \ \forall \ x \in X$.
Then

$$p_1(F) \leqq \int f_n dp_1 = \int f_n dp_2 \leqq p_2(G_n)$$

$$p_1(F) \leqq p_2(\bigcap_{n=1}^{\infty} G_n) = p_2(F).$$

Reversing $p_1$ and $p_2$, we obtain $p_1(F) = p_2(F)$. By Theorem A.1, this implies $p_1(B) = p_2(B)$ for every $B \in \mathcal{B}_X$. QED

These two theorems essentially says that a probability measure on a metrizable space is completely determined by its values on the open or closed sets. Also, a probability measure $p$ on a metric space $(X,d)$ is completely determined by the values $\int gdp$ where $g$ ranges over $U_d(X)$. Next, attention is paid to the development of the topology $\mathcal{T}[C(X)]$, the so-called weak topology on $P(X)$. However, the space $C(X)$ is too large to be manipulated easily, one would need a countable set $D \subset C(X)$ such that $\mathcal{T}(D) = \mathcal{T}[C(X)]$. Such a set $D$ is produced by the next three lemmas.

Definition: Let $\epsilon > 0$, $p \in P(X)$ and $f \in C(X)$,

$$V_\epsilon(p;f) = \{q \in P(X) : |\int fdq - \int fdp| < \epsilon\}.$$

Definition: Let $D \subset C(X)$. Define the collection of subsets of $P(X)$:

$$\mathcal{V}(D) = \{V_\epsilon(p;f) : \epsilon > 0, \ p \in P(X), \ f \in D\}.$$

Let $\mathscr{T}(D)$ be the weak topology on $P(X)$ which contains $\mathscr{V}(D)$, i.e., the topology for which $\mathscr{V}(D)$ is a subbase.

Lemma A.1: Let $X$ be a metrizable space and $D \subset C(X)$. Let $\{p_\alpha\}$ be a net in $P(X)$ and $p \in p(x)$. Then $p_\alpha \to p$ relative to the topology $\mathscr{T}(D)$ if and only if $\int f dp_\alpha \to \int f dp$ for every $f \in D$.

Proof: Let $p_\alpha \to p$ and $f \in D$. Let $\epsilon > 0$ be given, and let $\beta$ be such that for $\alpha \geqq \beta$ implies $p_\alpha \in V_\epsilon(p;f)$

$\therefore \int f dp_\alpha \to \int f dp$.

Conversely, let $\int f dp_\alpha \to \int f dp$ for every $f \in D$.

Suppose $G \in \mathscr{T}(D)$ contains $p$. Then $p$ is contained in some basic open set $\bigcap_{k=1}^{n} V_{\epsilon_k}(p;f_k) \subset G$ where $\epsilon_k > 0$, $f_k \in D$ and $k = 1,\ldots,n$. Let $\beta$ be such that for all $\alpha \geqq \beta$

$$\left|\int f_k dp_\alpha - \int f_k d_p\right| < \epsilon_k \quad k = 1,\ldots,n$$

$\therefore p_\alpha \in G$ for $\alpha \geqq \beta$

$\therefore p_\alpha \to p$. QED

Lemma A.2: Let $X$ be a metrizable space and $d$ a metric on $X$ consistent with its topology. If $f \in C(X)$, then there exists sequences $\{g_n\}$ and $\{h_n\}$ such that $g_n \uparrow f$ and $h_n \uparrow f$.

Proof: Let $b \in R$ and $x_o \in X$ be such that

$$b \leqq f(x) \leqq f(x_o) < \infty \quad \text{for every} \quad x \in X.$$

Define $g_n(x) = \inf_{y \in X}[f(y) + nd(x,y)]$. Thus for every $x \in X$

$$b \leqq g_n(x) \leqq f(x) + nd(x,x)$$
$$\leqq f(x_0) + nd(x,x_0) < \infty$$

$$\therefore \quad b \leqq g_1 \leqq g_2 \leqq \cdots \leqq f \quad \text{and} \quad \lim_{n\to\infty} g_n \leqq f$$

For every $x,y,z \in X$

$$f(y) + nd(x,y) \leqq f(y) + nd(y,z) + nd(x,z)$$

$$\therefore \quad g_n(x) \leqq g_n(z) + nd(x,z).$$

Thus

$$|g_n(x) - g_n(z)| \leqq nd(x,z)$$

$$\therefore \quad g_n \in U_d(X) \quad \text{for each } n.$$

Let $\epsilon > 0$ and $\{y_n\} \subset X$ be such that

$$f(y_n) + nd(x,y_n) \leqq g_n(x) + \epsilon.$$

As $n \to \infty$, either $g_n \uparrow \infty$ or $y_n \to x$. If

$$g_n \uparrow \infty \Rightarrow \lim_{n\to\infty} g_n \geqq f$$

$$\therefore \quad g_n \uparrow f.$$

If $y_n \to x$, since $f$ is continuous

$$f(x) = \lim_{n\to\infty} f(y_n)$$
$$\leqq \lim_{n\to\infty} g_n(x) + \epsilon$$
$$\therefore \quad \lim_{n\to\infty} g_n(x) = f(x). \quad \text{QED}$$

<u>Lemma A.3</u>: Let X be a metrizable space and d a metric on X consistent with its topology. Then $\mathcal{T}[C(X)] = \mathcal{T}[U_d(X)]$.

Proof: Since $U_d(X) \subset C(X)$, $\mathcal{V}[U_d(X) \subset \mathcal{V}[C(X)]$. This implies $\mathcal{T}[U_d(X)] \subset \mathcal{T}[C(X)]$. Let $p_0 \in V_\epsilon(p;f)$ $\exists$ $\epsilon_0 > 0$ such that $V_{\epsilon_0}(p_0;f) \subset V_\epsilon(p;f)$. By Lemma A.2, $\exists$ $g,h \in U_d(X)$ such that $g \leqq f \leqq h$ and $\int fdp_0 < \int gdp_0 + \epsilon_0/2$, $\int hdp_0 < \int fdp_0 + \epsilon_0/2$. Let $q \in V_{\epsilon_0/2}(p_0;g) \cap V_{\epsilon_0/2}(p_0;h)$, then

$$\int fdp_0 < \int gdp_0 + \epsilon_0/2 < \int gdq + \epsilon_0$$
$$\leqq \int fdq + \epsilon_0$$

and

$$\int fdq \leqq \int hdq < \int hdp_0 + \epsilon_0/2 < \int fdp_0 + \epsilon_0$$

$$\therefore \quad \left|\int fdq - \int fdp_0\right| < \epsilon_0$$

$$\therefore \quad q \in V_{\epsilon_0}(p_0;f)$$

and

$$V_{\epsilon_0/2}(p_0;g) \cap V_{\epsilon_0/2}(p_0;h) \subset V_\epsilon(p;f).$$

Since $p_0 \in V_\epsilon(p;f) \in \mathcal{V}[C(X)]$ this implies $V_\epsilon(p;f)$ is open in the $\mathcal{T}[U_d(X)]$ topology and

$$\mathcal{T}[C(X)] \subset \mathcal{T}[U_d(X)]$$

$$\therefore \quad \mathcal{T}[C(X)] = \mathcal{T}[U_d(X)]. \quad \text{QED}$$

<u>Lemma A.4</u>: Let $X$ be a metrizable space and $d$ a metric on $X$ consistent with its topology. If $D$ is dense in $U_d(X)$ then $\mathcal{T}[U_d(X)] = \mathcal{T}(D)$.

Proof: Clearly $\mathcal{T}(D) \subset \mathcal{T}[U_d(X)]$. Let

$$V_\epsilon(p;g) \in \mathcal{V}[U_d(X)] \quad \text{and} \quad p_0 \in V_\epsilon(p;g).$$

Let

$$\epsilon_0 = \epsilon - \left|\int gdp_0 - \int gdp\right| > 0.$$

Let $h \in D$ be such that $\|g-h\| < \epsilon_0/3$. For any $q \in V_{\epsilon_0/3}(p_0;h)$

$$\begin{aligned}\left|\int gdq - \int gdp\right| &\leqq \left|\int gdq - \int hdq\right| + \left|\int hdq - \int hdp_0\right| \\ &\quad + \left|\int hdp_0 - \int gdp_0\right| + \left|\int gdp_0 - \int gdp\right| \\ &< \epsilon_0/3 + \epsilon_0/3 + \epsilon_0/3 \\ &\quad + \left|\int gdp_0 - \int gdp\right| = \epsilon\end{aligned}$$

$$\therefore \quad V_{\epsilon_0/3}(p_0;h) \subset V_\epsilon(p;q).$$

This implies

$$\mathcal{T}[U_d(X)] \subset \mathcal{T}(D)$$

$$\therefore \quad \mathcal{T}[U_d(X)] = \mathcal{T}(D). \qquad \text{QED}$$

<u>Theorem A.3</u>: Let $X$ be a separable metrizable space. There exists a metric $d$ on $X$ consistent with its topology and a countable dense subset $D$ of $U_d(X)$ such that $\mathcal{T}(D)$ is the weak topology $\mathcal{T}[C(X)]$ on $P(X)$.

Proof: By Urysohn's theorem, $X$ can be homeomorphically embedded into a subset of the Hilbert cube. Since the Hilbert cube is compact by Tychonff's theorem, it is totally bounded.

$\therefore \exists$ a totally bounded metrization $d$ on $X$. This implies that $(X,d)$ can be isometrically embedded as a dense subset of a compact metric space $(X_d, d_1)$ where $X \subset X_d$. Let $g \in U_d(X)$, $g$ has a unique extension $\hat{g} \in C(X_d)$ such that $\|g\| = \|\hat{g}\|$. The mapping $g \to \hat{g}$ is linear and norm-preserving. Since $C(X_d)$ is separable, this implies $U_d(X)$ is separable.

$\therefore \exists$ a countable dense set $D$ in $U_d(X)$. The result follows from Lemmas A.3 and A.4. QED

From this point on, whenever $X$ is metrizable, it is understood that $P(X)$ is a topological space with the weak topology $\mathcal{T}[C(X)]$.

<u>Theorem A.4</u>: If $X$ is a separable metrizable space, then $p(x)$ is separable and metrizable.

Proof: Let $d$ be a metric on $X$ consistent with its topology and $D$ a countable subset of $U_d(X)$ such that $\mathcal{T}(D)$ is the weak topology on $P(X)$. Let $R^\infty$ be the product of countably many copies of $R$. Define $\varphi : P(X) \to R^\infty$ by

$$\varphi(p) = (\int g_1 dp, \int g_2 dp, \ldots)$$

where $\{g_1, g_2, \ldots\}$ is an enumeration of $D$. Suppose that $\varphi(p_1) = \varphi(p_2)$, then

$$\int g_k dp_1 = \int g_k dp_2 \quad \text{for every } g_k \in D.$$

Let $g \in U_d(X)$. $\exists$ a sequence $\{g_k\} \subset D$ such that $\|g_{k_i} - g\| \to 0$ as $i \to \infty$. Then

$$\begin{aligned} &\left|\int g dp_1 - \int g dp_2\right| \\ &\leqq \limsup_{i\to\infty}\left|\int (g - g_{k_i}) dp_1\right| + \limsup_{i\to\infty}\left|\int g_{k_i} dp_1 - \int g_{k_i} dp_2\right| \\ &\qquad + \limsup_{i\to\infty}\left|\int (g_{k_i} - g) dp_2\right| \\ &\leqq 2 \limsup_{i\to\infty}\|g_{k_i} - g\| = 0 \end{aligned}$$

$$\therefore \int g dp_1 = \int g dp_2.$$

By Theorem A.2, $p_1 = p_2$

$\therefore$ $\varphi$ is one-to-one.

By Lemma A.1, for each $g_k \in D$, the mapping $p \to \int g_k dp$ is continuous

$\therefore$ $\varphi$ is continuous.

Let $\{p_\alpha\}$ be a net in $P(X)$ such that $\varphi(p_\alpha) \to \varphi(p)$ for some $p \in P(X)$. Then

$$\int g_k dp_\alpha \to \int g_k dp \quad \text{for every} \quad g_k \in D.$$

By Lemma A.1, $p_\alpha \to p$

$\therefore\ \varphi^{-1}$ is continuous. The $\varphi$ is a homeomorphism. $R^\infty$ is metrizable and separable $\Rightarrow P(X)$ is metrizable and separable. QED

<u>Theorem A.5</u>: Let $X$ be a separable metrizable space and let $d$ be a metric on $X$ consistent with its topology. Let $\{p_n\}$ be a sequence in $P(X)$ and $p \in P(X)$. The following statements are equivalent:

(a) $p_n \to p$

(b) $\int f dp_n \to \int f dp$ for every $f \in C(X)$

(c) $\int g dp_n \to \int g dp$ for every $g \in U_d(X)$

(d) $\limsup_{n\to\infty} p_n(F) \leq p(F)$ for every closed set $F \subset X$

(e) $\liminf_{n\to\infty} p_n(G) \geq p(G)$ for every open set $G \subset X$.

Proof: The equivalence of (a), (b) and (c) follows from Lemmas A.1 and A.3. The equivalence of (d) and (e) follows by complementation. To show (b) implies (d), let $F$ be a closed proper nonempty subset of $X$. Let $G_k = \{x \in X : d(x,F) < \frac{1}{k}\}$. $F$ and $\sim G_k$ are disjoint nonempty sets for $k$ sufficiently large. $\exists\ f_k \in C(X)$

such that $f_k(x) = 1$ for $x \in F$, $f_k(x) = 0$ for $x \in \sim G_k$ and $0 \leqq f_k(x) \leqq 1$ for every $x \in X$. Thus

$$\limsup_{n\to\infty} p_n(F) \leqq \lim_{n\to\infty} \int f_k dp_n = \int f_k dp \leqq p(G_k).$$

(d) follows by letting $k \to \infty$.

To show (d) implies (b), let $f \in C(X)$ and assume without loss of generality that $0 \leqq f \leqq 1$. Let $K$ be a positive integer and define

$$F_k = \{x \in X : f_k \geqq \tfrac{k}{K}\}, \quad k = 0,\ldots,K.$$

Define $\varphi : X \to [0,1]$ by $\varphi(x) = \sum_{k=0}^{K} (k/K)\chi_{F_k - F_{k+1}}(x)$ where $F_{K+1} = \emptyset$. Then

$$f - \frac{1}{K} \leqq \varphi \leqq f.$$

For any $q \in P(X)$,

$$\int \varphi dq = \sum_{k=0}^{K} (\tfrac{k}{K}) q(F_k - F_{k+1}) = \frac{1}{K} \sum_{k=1}^{K} q(F_k)$$

$$\begin{aligned}\limsup_{n\to\infty} \int f dp_n - (\tfrac{1}{K}) &\leqq \limsup_{n\to\infty} \int \varphi dp_n \\ &= \frac{1}{K} \limsup_{n\to\infty} \sum_{k=1}^{K} p_n(F_k) \\ &\leqq \frac{1}{K} \sum_{k=1}^{K} p(F_k) = \int \varphi dp \\ &\leqq \int f dp.\end{aligned}$$

This implies

$$\limsup_{n\to\infty} \int f dp_n \leqq \int f dp \quad \text{for every} \quad f \in C(X).$$

But the above argument holds for $-f$

$$\liminf_{n\to\infty} \int f dp_n = -\limsup_{n\to\infty} \int (-f) dp_n$$

$$\geqq -\int (-f) dp = \int f dp$$

$\therefore \int f dp_n \to \int f dp$ for every $f \in C(X)$. QED

The above two theorems guarantee that when $X$ is separable and metrizable, the topology on $P(X)$ can be characterized in terms of convergent sequences rather than nets. Theorem A.5 gives several conditions which are equivalent to convergence in $P(X)$.

<u>Corollary A.5.1</u>: Let $X$ be a metrizable space. The mapping $\delta : X \to P(X)$ defined by $\delta(x) = p_x$ is a homeomorphism.

Proof: Clearly $\delta$ is one-to-one.

Let $\{x_n\}$ be a sequence in $X$ and $x \in X$. If $x_n \to x$ and $G$ is an open subset of $X$, then either

(i) $x \in G$, implies $x_n \in G$ for large $n$

$\therefore \liminf_{n\to\infty} p_{x_n}(G) = 1 = p_x(G)$ or

(ii) $x \notin G$, then $\liminf_{n\to\infty} p_{x_n}(G) \geqq 0 = p_x(G)$.

By Theorem A.5, this implies $p_{x_n} \to p_x$

$\therefore \delta$ is continuous.

On the other hand, if $p_{x_n} \to p_x$ and $G$ is an open neighborhood of $x$, since $\liminf_{n\to\infty} p_{x_n}(G) \geqq p_x(G) = 1$. Then $x_n \in G$ for sufficiently large $n$

$\therefore x_n \to x$

$\therefore \delta$ is a homeomorphism. QED

<u>Theorem A.6</u>: If $X$ is a compact metrizable space, then $P(X)$ is a compact metrizable space.

Proof: If $X$ is a compact metrizable space, it is separable and $C(X)$ is separable.

Let $\{f_k\}$ be a countable set in $C(X)$ such that $f_1 \equiv 1$, $\|f_k\| \leqq 1$ for every $k$ and $\{f_k\}$ is dense in the unit sphere $\{f \in C(X) : \|f\| < 1\}$.

Define $\varphi : P(X) \to [-1,1]^\infty$ by $\varphi(p) = (\int f_1 dp, \int f_2 dp, \ldots)$. $\varphi$ can be easily verified as a homeomorphism. Suppose $\{p_n\}$ is a sequence in $P(X)$ and $\varphi(p_n) \to (\alpha_1, \alpha_2, \ldots) \in [-1,1]^\infty$. Let $\epsilon > 0$ be given and $f \in C(X)$ with $\|f\| \leqq 1$, $\exists f_k$ with $\|f - f_k\| < \frac{\epsilon}{3}$. Also $\exists n_0 \geqq 0$ such that $|\int f_k dp_n - \int f_k dp_m| < \frac{\epsilon}{3}$ whenever $n,m \geqq n_0$. Then

$$|\int f dp_n - \int f dp_m| \leqq |\int f dp_n - \int f_k dp_n| + |\int f_k dp_n - \int f_k dp_m| + |\int f_k dp_m - \int f dp_m| < \epsilon$$

$\therefore \{\int f dp_n\}$ is Cauchy in $[-1,1]$.

Let $E(f)$ be the limit of such a sequence. If $\|f\| > 1$, define $E(f) = \|f\| \, E(\frac{f}{\|f\|})$. Thus $E$ is a linear function on $C(X)$, $E(f) \geqq 0$ whenever $f \geqq 0$, $|E(f)| \leqq \|f\|$ for every $f \in C(X)$ and $E(f_1) = 1$. Let $\{h_n\}$ be a sequence in $C(X)$ and $h_n(X) \downarrow 0$ for every $x \in X$. For each $\epsilon > 0$, the set $K_n(\epsilon) = \{x \in X : h_n(X) > \epsilon\}$ is compact and $\bigcap_{n=1}^{\infty} K_n(\epsilon) = \emptyset$.

$\therefore$ For $n$ sufficiently large, $\|h_n\| \downarrow 0$ which implies $E(h_n) \downarrow 0$.

$\exists$ a unique probability measure on $\sigma[\bigcup_{f \in C(X)} f^{-1}(\mathcal{B}_R)]$ which satisfies $E(f) = \int f dp$ for every $f \in C(X)$ since $E$ is a Daniell integral (see Royden [1968])

$\therefore p \in P(X)$

and

$$\alpha_k = \lim_{n\to\infty} \int f_k dp_n = E(f_k) = \int f_k dp \qquad k = 1,2,\ldots$$

$\therefore \varphi(p_n) \to \varphi(p)$ and $\varphi[P(X)]$ is closed on $[-1,1]^{\infty}$.

Since $[-1,1]^{\infty}$ is compact, $P(X)$ is compact. QED

<u>Lemma A.5</u>: Let $X$ and $Y$ be separable metrizable spaces and $\varphi : X \to Y$ a homeomorphism. Define $\psi: P(X) \to P(Y)$ by

$$\psi(p)(B) = p[\varphi^{-1}(B)] \;\; \forall \; B \in \mathcal{B}_Y$$

$\psi$ is a homeomorphism.

Proof: Let $p_1, p_2 \in P(X)$ and $p_1 \neq p_2$. Since $p_1$ and $p_2$ are regular, $\exists$ an open set $G \subset X$ for which $p_1(G) \neq p_2(G)$

$$\varphi(G) \text{ is relatively open in } \varphi(X).$$

Thus $\varphi(G) = \varphi(X) \cap B$, where $B$ is open in $Y$ and

$$\psi(p_1)(B) = p_1(G) \neq p_2(G) = \psi(p_2)(B)$$

$\therefore$ $\psi$ is one-to-one.

Let $\{p_n\}$ be a sequence in $P(X)$ and $p \in P(X)$. If $p_n \to p$, since $\varphi^{-1}(H)$ is open in $X$ for every open set $H \subset Y$. By Theorem A.5,

$$\liminf_{n\to\infty} \psi(p_n)(H) = \liminf_{n\to\infty} p_n[\varphi^{-1}(H)]$$
$$\geq p[\varphi^{-1}(H)] = \psi(p_2)(H)$$

$\therefore$ $\psi(p_n) \to \psi(p)$ and $\psi$ is continuous.

Reversing the arguments with $\{p_n\}$ and $p$ such that $\psi(p_n) \to \psi(p)$ will show that $p_n \to p$

$\therefore$ $\psi^{-1}$ is continuous. QED

<u>Theorem A.7</u> (Urysohn's Theorem): Every separable metrizable space is homeomorphic to a subset of the Hilbert cube $\mathcal{X}$.

<u>Theorem A.8</u> (Alexandroff's Theorem): Let $X$ be a topologically complete space, $Z$ a metrizalbe space,

and $\varphi : X \to Z$ a homeomorphism. The $\varphi(X)$ is a $G_\delta$-subset of $Z$. Conversely, if $Y$ is a $G_\delta$-subset of $Z$ and $Z$ is topologically complete, then $Y$ is topologically complete.

<u>Theorem A.9</u>: If $X$ is a topologically complete spearable space, then $P(X)$ is topologically complete and separable.

Proof: By Theorem A.7, $\exists$ a homeomorphism $\varphi : X \to \mathcal{H}$ and the mapping $\psi$ in Lemma A.5 with $\mathcal{H}$ replacing $Y$ is a homeomorphism from $P(X)$ to $P(\mathcal{H})$. By Theorem A.8, $\varphi(X)$ is a $G_\delta$-subset of $\mathcal{H}$

$$\psi[P(X)] = \{p \in P(\mathcal{H}) : p[\mathcal{H} - \varphi(X)] = 0\}.$$

Since $\mathcal{H}$ is compact, by Theorem A.6, $P(\mathcal{H})$ is compact.

Let $G_1 \supset G_2 \supset \ldots$ such that $\varphi(X) = \bigcap_{n=1}^{\infty} G_n$. Then

$$\begin{aligned}\psi[P(X)] &= \bigcap_{n=1}^{\infty} \{p \in P(\mathcal{H}) : p(\mathcal{H} - G_n) = 0\} \\ &= \bigcap_{n=1}^{\infty} \bigcap_{k=1}^{\infty} \{p \in P(\mathcal{H}) : p(\mathcal{H} - G_n) < \tfrac{1}{k}\}.\end{aligned}$$

But by Theorem A.5 (d), for any closed set $F$ and $c \in R$, the set $\{p \in P(\mathcal{H}) : p(F) \geq c\}$ is closed and this implies $\{p \in P(\mathcal{H}) : p(\mathcal{H} - G_n) < \frac{1}{k}\}$ is open

$\therefore$ $\psi[P(X)]$ is a $G_\delta$-subset of $P(\mathcal{H})$.

Again by Theorem A.8, this implies $P(X)$ is topologically complete. QED

The next step is to establish the fact that $\mathcal{B}_{P(X)}$ is the smallest σ-algebra with respect to which $\theta_B$ is measurable for every $B \in \mathcal{B}_X$. To do that, a useful aid, the concept of a Dynkin system, is needed.

Definition: Let $X$ be a set and $\mathcal{D}$ a class of subsets of $X$. $\mathcal{D}$ is a Dynkin system if the following conditions hold:

(a) $X \in \mathcal{D}$

(b) If $A, B \in \mathcal{D}$ and $B \subset A$, then $A - B \in \mathcal{D}$

(c) If $A_1, A_2, \ldots \in \mathcal{D}$ and $A_1 \subset A_2 \subset \ldots$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$.

The following is a well-known result in measure theory, which is quoted without proof.

<u>Theorem A.10</u> (Dynkin System Theorem): Let $\mathcal{J}$ be a class of subsets of a set $X$, and assume $\mathcal{J}$ is closed under finite intersections. If $\mathcal{D}$ is a Dynkin system containing $\mathcal{J}$, then $\mathcal{D}$ also contains $\sigma(\mathcal{J})$.

<u>Theorem A.11</u>: Let $X$ be a separable metrizable space and $\mathcal{E}$ a collection of subsets of $X$ which generates $\mathcal{B}_X$ and is closed under finite intersections. Then $\mathcal{B}_{P(X)}$ is the smallest σ-algebra with respect to which all functions of the form $\theta_E(p) = P(E)$, $E \in \mathcal{E}$

are measurable from P(X) to [0,1], i.e.,

$$\mathcal{B}_{P(X)} = \sigma[\bigcup_{E \in \mathcal{E}} \theta_E^{-1}(\mathcal{B}_R)].$$

Proof: Let $\mathcal{F}$ be the smallest $\sigma$-algebra with respect to which $\theta_E$ is measurable for every $E \in \mathcal{E}$.

Let $\mathcal{D} = \{B \in \mathcal{B}_X : \theta_B$ is $\mathcal{B}_{P(X)}$-measurable$\}$. Clearly, $\mathcal{D}$ is a Dynkin system, and by Theorem A.10, $\mathcal{D} = \mathcal{B}_X$.

Let $\mathcal{D}' = \{B \in \mathcal{B}_X : \theta_B$ is $\mathcal{F}$-measurable$\}$. Again $\mathcal{D}'$ is a Dynkin system. Since $\mathcal{E} \subset \mathcal{D}'$, $\mathcal{D}' = \mathcal{B}_X$. Thus $\theta_f(p) = \int f dp$ is $\mathcal{F}$-measurable for f is the indicator of a Borel set. $\theta_f$ is $\mathcal{F}$-measurable when f is a Borel-measurable simple function.

Let $f \in C(X)$, $\exists$ a sequence of simple functions $f_n$ uniformly bounded below such that $f_n \uparrow f$. By the monotone convergence theorem $\theta_{f_n} \uparrow \theta_f$. Thus $\theta_f$ is $\mathcal{F}$-measurable. Then for $\epsilon > 0$, $p \in P(X)$, $f \in C(X)$,

$$V_\epsilon(p;f) \text{ is } \mathcal{F}\text{-measurable.}$$

This implies $\mathcal{B}_{P(X)} = \mathcal{F}$. QED

<u>Theorem A.12</u>: If X is a Borel space, then P(X) is a Borel space.

Proof: Let $\varphi$ be a homeomorphism mapping X onto a Borel subset of a topologically complete separable space Y. Then by Lemma A.5, P(X) is homeomorphic to the Borel set $\{p \in P(Y) : p[\varphi(x)] = 1\}$. But $P(Y)$ is topologically complete and separable by Theorem A.9, $P(X)$ is a Borel space. QED

In this appendix, it has been shown that the space $P(X)$ can be taken to be a topological space with the weak topology $\mathcal{T}[C(X)]$. It inherits most of the properties of the space $X$. When $X$ is separable and metrizable, $P(X)$ is separable and metrizable; when $X$ is separable and topologically complete, $P(X)$ is separable and topologically complete; and when $X$ is a Borel space, $P(X)$ is a Borel space.

## APPENDIX TWO

The properties of analytic sets are summarized in this appendix. The analysis will stop with a collection of equivalent definitions of the set. Again, there are no new results in this appendix. The reader may be referred to some standard mathematics text for a more elaborate treatment of the topic.

The definitions of a paved space, Suslin scheme and analytic sets are given in Chapter VI* and will not be repeated here. First, some preliminary results are stated without proof. (The reader may request the author for proof of the following theorem and its corollaries.)

<u>Theorem A.13</u>: Let X be a space with pairings $\mathcal{P}$ and $\mathcal{Q}$ such that $\mathcal{P} \subset \mathcal{Q}$.

(a) $\mathcal{S}(\mathcal{P}) \subset \mathcal{S}(\mathcal{Q})$

(b) $\mathcal{S}(\mathcal{P})_{\delta} = \mathcal{S}(\mathcal{P})$

(c) $\mathcal{S}(\mathcal{P})_{\sigma} = \mathcal{S}(\mathcal{P})$

(d) $\mathcal{P} \subset \mathcal{S}(\mathcal{P})$

(e) $\mathcal{S}(\mathcal{P}) = \mathcal{S}[\mathcal{S}(\mathcal{P})]$

<u>Corollary A.13.1</u>: Let $(X,\mathcal{P})$ be a paved space and suppose that the complement of each set in $\mathcal{P}$ is in $\mathcal{S}(\mathcal{P})$. Then $\sigma(\mathcal{P}) \subset \mathcal{S}(\mathcal{P})$.

<u>Corollary A.13.2</u>: Let X be a Borel space. The countable intersections and unions of analytic subsets of X are analytic.

<u>Theorem A.14</u>: Let X be a Borel space. Then every Borel subset of X is analytic. In fact, the class of analytic sets $\mathcal{S}(\mathcal{T}_X)$ is equal to $\mathcal{S}(\mathcal{B}_X)$.

Proof: Every open subset of X is an $F_\sigma$, so every open set is analytic. Corollary A.13.1 implies $\mathcal{B}_X \subset \mathcal{S}(\mathcal{T}_X)$. By Theorem A.13 (a) and (e), since $\mathcal{T}_X \subset \mathcal{B}_X$

$$\mathcal{S}(\mathcal{T}_X) \subset \mathcal{S}(\mathcal{B}_X) \subset \mathcal{S}[\mathcal{S}(\mathcal{T}_X)] = \mathcal{S}(\mathcal{T}_X)$$

$$\therefore \quad \mathcal{S}(\mathcal{T}_X) = \mathcal{S}(\mathcal{B}_X). \qquad \text{QED}$$

If the Borel space X is countable, then every subset of X is both analytic and Borel-measurable. If X is uncountable, however, the class of analytic subsets of X is strictly larger than $_X$. Note that an immediate consequence of Theorem A.14 is that if Y is a Borel subset of the Borel space X, then the analytic subsets of Y are the analytic subsets of X contained in Y. A generalization of this fact is the following.

Corollary A.14.1: Let $X$ and $Y$ be Borel spaces and $\varphi : X \to Y$ a Borel isomorphism. Then $A \subset X$ is analytic if and only if $\varphi(A) \subset Y$ is analytic.

Proof: Let $\varphi : X \to Y$ be a Borel isomorphism and $A \subset X$ is analytic. Then $A = N(S)$ where $S$ is a Suslin scheme for $\mathcal{F}_X$. Let $\varphi \circ S$ be the Suslin scheme for $\mathcal{B}_Y$ defined by

$$(\varphi \circ S)(s) = \varphi[S(s)]$$

Thus $\varphi(A) = N(\varphi \circ S)$.

$\therefore$ By Theorem A.14, $\varphi(A)$ is analytic.

If $\varphi(A) \subset Y$ is analytic, $A \subset X$ is analytic by a similar argument. QED

Definition: Let $(X,\mathcal{P})$ be a paved space and $S$ a Suslin scheme for $\mathcal{P}$. The Suslin scheme $S$ is regular if for each $n \in N$ and $(\sigma_1, \sigma_2, \ldots, \sigma_{n+1}) \in \Sigma$, then $S(\sigma_1, \sigma_2, \ldots, \sigma_n) \supset S(\sigma_1, \sigma_2, \ldots, \sigma_n, \sigma_{n+1})$.

Lemma A.6: Let $(X,d)$ be a separable metric space and $S$ a Suslin scheme for $\mathcal{F}_X$. Then there exists a regular Suslin scheme $R$ for $\mathcal{F}_X$ such that $N(R) = N(S)$ and, for every $z = (\zeta_1, \zeta_2, \ldots) \in \mathfrak{N}$

$$\lim_{n\to\infty} \operatorname{diam} R(\zeta_1, \zeta_2, \ldots, \zeta_n) = 0 \quad \text{if}$$

$$R(\zeta_1, \zeta_2, \ldots, \zeta_n) \neq \emptyset \ \forall\, n.$$

Proof: By the Lindelörf property, for each positive integer $k$, $X$ can be covered by a countable collection of open balls of the form $B_{kj} = \{x \in X : d(x, x_{kj}) < 1/k\}$ $j = 1, 2, \cdots$. For $(\bar{\zeta}_1, \zeta_1, \bar{\zeta}_2, \zeta_2, \ldots) \in \mathfrak{N}$, define

$$R(\bar{\zeta}_1) = \bar{B}_{1\bar{\zeta}_1}$$

$$R(\bar{\zeta}_1, \zeta_1) = R(\bar{\zeta}_1) \cap S(\zeta_1)$$

$$R(\bar{\zeta}_1, \zeta_1, \bar{\zeta}_2) = R(\bar{\zeta}_1, \zeta_1) \cap \bar{B}_{2\bar{\zeta}_2}$$

$$R(\bar{\zeta}_1, \zeta_1, \bar{\zeta}_2, \zeta_2) = R(\bar{\zeta}_1, \zeta_1, \bar{\zeta}_2) \cap S(\zeta_1, \zeta_2) \quad \text{etc.}$$

Thus

$$\bigcap_{<(\bar{\zeta}_1, \zeta_1, \bar{\zeta}_2, \zeta_2 \ldots)} R(s) = \left[\bigcap_{k=1}^{\infty} \bar{B}_{k\bar{\zeta}_k}\right] \cap \left[\bigcap_{s<z}^{\infty} S(s)\right]$$

where $z = (\zeta_1, \zeta_2, \ldots)$. Clearly $R$ is a regular Suslin scheme for $\mathcal{F}_X$

$$\lim_{n\to\infty} \operatorname{diam} R(\zeta_1, \zeta_2, \ldots, \zeta_n) = 0 \quad \text{if}$$

$$R(\zeta_1, \zeta_2, \ldots, \zeta_n) \neq \emptyset \quad \forall\, n.$$

Let $x \in N(R)$ $\exists$ $(\bar{\zeta}_1, \zeta_1, \bar{\zeta}_2, \zeta_2, \ldots) \in \mathfrak{N}$ such that

$$x \in \bigcap_{s<(\bar{\zeta}_1, \zeta_1, \bar{\zeta}_2, \zeta_2)} R(s)$$

Thus

$$x \in \bigcap_{s<(\zeta_1, \zeta_2, \ldots)} S(s) \subset N(S)$$

$$\therefore \quad N(R) \subset N(S).$$

Now let $x \in N(S)$, $\exists\ (\zeta_1, \zeta_2, \ldots) \in \mathfrak{N}$ such that

$$x \in \bigcap_{s<(\zeta_1,\zeta_2,\ldots)} S(s).$$

Since $\{B_{kj} : j = 1,2,\ldots\}$ covers $X$, $\exists$ for each $k$ a positive integer $\bar{\zeta}_k$ for which $x \in B_{k\bar{\zeta}_k}$. Thus

$$x \in \bigcap_{k=1}^{\infty} B_{k\bar{\zeta}_k}$$

$$x \in \bigcap_{s<(\bar{\zeta}_1,\zeta_1,\bar{\zeta}_2,\zeta_2,\ldots)} R(s) \subset N(R).$$

Thus $N(S) \subset N(R)$

$\therefore$ $N(R) = N(S)$. QED

Lemma A.7: Let $(x,d)$ be a complete separable metric space. If $A \subset X$ is a nonempty analytic set, then there exist a closed subset $\mathfrak{N}_1$ of $\mathfrak{N}$ and a continuous function $f : \mathfrak{N}_1 \to X$ such that $A = f(\mathfrak{N}_1)$. Conversely, if $\mathfrak{N}_1 \subset \mathfrak{N}$ is closed and $f : \mathfrak{N}_1 \to X$ is continuous, then $f(\mathfrak{N}_1)$ is analytic.

Proof: Let $A = N(R)$ be nonempty, where $R$ is a regular Suslin scheme for $\mathcal{F}_X$ satisfying Lemma A.6. Define

$$\mathfrak{N}_1 = \{z \in \mathfrak{N} : \bigcap_{s<z} R(s) \neq \emptyset\}.$$

Let $z = (\zeta_1, \zeta_2, \ldots)$ be in $\mathfrak{N}$. Let $R(\zeta_1, \zeta_2, \ldots, \zeta_n) \neq \emptyset$ for each $n$, $\exists\ x_n \in R(\zeta_1, \zeta_2, \ldots, \zeta_n)$. By Lemma A.6,

$\{x_n\}$ is Cauchy and $x_n \to x$ where $x \in X$ since $X$ is complete.

$R$ is regular, for each $n$, $\{x_m : m \geqq n\} \subset R(\zeta_1, \zeta_2, \ldots, \zeta_n)$. Thus $x \in R(\zeta_1, \zeta_2, \ldots, \zeta_n)$

$$x \in \bigcap_{s<z} R(s).$$

Now, let $z \in \mathfrak{N} - \mathfrak{N}_1$. Then for some $s_n < z$, $R(s_n) = \emptyset$. If $z \in \{\omega \in \mathfrak{N} : s_n < \omega\}$ an open neighborhood, then $\{\omega \in \mathfrak{N} : S_n < \omega\} \subset \mathfrak{N} - \mathfrak{N}_1$. Thus $\mathfrak{N} - \mathfrak{N}_1$ is open and $\mathfrak{N}_1$ is closed. For $z \in \mathfrak{N}_1$, define $f(z)$ to be the unique point in $\bigcap_{s<z} R(s)$. Let $\{z_k\} \subset \mathfrak{N}_1$ and $z_k \to z_0$ where $z_0 = (\zeta_1, \zeta_2, \ldots) \in \mathfrak{N}_1$. By Lemma A.8, given $\varepsilon > 0$ $\exists$ $s_n < z_0$ such that $\operatorname{diam} R(s_n) < \varepsilon$. For $k$ sufficiently large, $z_k \in \{z \in \mathfrak{N} : s_n < z\}$

$$f(z_k) \in R(s_n)$$

$$\therefore \quad d(f(z_k), f(z_0)) \leqq \operatorname{diam} R(s_n) < \varepsilon$$

$\therefore$ $f$ is continuous.

Now, suppose $\mathfrak{N}_1 \subset \mathfrak{N}$ is closed and $f : \mathfrak{N}_1 \to \mathfrak{N}$ is continuous. Define a regular Suslin scheme $R$ for $\mathcal{F}_X$ by $R(s) = \overline{f(\{z \in \mathfrak{N}_1 : s < z\})}$ where $R(s) = \emptyset$ if $\{z \in \mathfrak{N}_1 : s < z\} = \emptyset$. If $z \in \mathfrak{N}_1$, then

$$f(z) \in \bigcap_{s<z} R(s) \subset N(R)$$

$$\therefore \quad f(\mathfrak{N}_1) \subset N(R)$$

Let $x \in N(R)$, then for some $z_0 = (\zeta_1, \zeta_2, \ldots) \in \mathfrak{N}$,

$$x \in \bigcap_{s < z_0} R(s).$$

For each $n$, $x \in \overline{f(\{z \in \mathfrak{N}_1 : (\zeta_1, \zeta_2, \ldots, \zeta_n) < z\})}$. Given $\varepsilon > 0$, $\exists\, z_n \in \mathfrak{N}_1$ such that $(\zeta_1, \zeta_2, \ldots, \zeta_n) < z_n$ and $d(x, f(z_n)) < \varepsilon$. Then $z_n$ must converge to $z_0$ as $n \to \infty$. Since $\mathfrak{N}_1$ is closed, $z_0 \in \mathfrak{N}_1$, $f$ is continuous implying

$$d(x, f(z_0)) \leqq \varepsilon$$

$$\therefore \quad f(z_0) = x$$

and

$$x \in f(\mathfrak{N}_1),\ N(R) \subset f(\mathfrak{N}_1)$$

$$\therefore \quad N(R) = f(\mathfrak{N}_1). \quad \text{QED}$$

So far, analytic sets are characterized as the continuous images of closed subsets of $\mathfrak{N}$. The following lemma and theorem allows one to get an even sharper characterization.

Lemma A.8: If $\mathfrak{N}_1$ is a nonempty closed subset of $\mathfrak{N}$, then there exists a continuous function $g : \mathfrak{N} \to \mathfrak{N}$ such that $\mathfrak{N}_1 = g(\mathfrak{N})$.

Proof: Let $\mathfrak{N}_1$ be covered by a countable collection of nonempty closed sets $\{S(\zeta_1) : \zeta_1 \in N\}$

satisfying

$$\mathfrak{N}_1 \supset S(\zeta_1), \quad \text{diam } S(\zeta_1) \leqq 1 \quad \zeta_1 = 1,2,\ldots$$

where d is a metric on $\mathfrak{N}$ consistent with its topology. Cover each $S(\zeta_1)$ with a countable collection of non-empty closed sets $\{S(\zeta_1,\zeta_2) : \zeta_2 \in N\}$ satisfying

$$S(\zeta_1) \supset S(\zeta_1,\zeta_2), \quad \text{diam } S(\zeta_1,\zeta_2) \leqq \frac{1}{2}, \quad \zeta_2 = 1,2,\ldots .$$

Thus, for any $(\zeta_1,\zeta_2,\ldots,\zeta_{n-1})$

$$S(\zeta_1,\zeta_2,\ldots,\zeta_n) \neq \emptyset \quad \zeta_n = 1,2,\ldots$$

$$S(\zeta_1,\zeta_2,\ldots,\zeta_{n-1}) = \bigcup_{\zeta_n=1}^{\infty} S(\zeta_1,\zeta_2,\ldots,\zeta_n)$$

$$S(\zeta_1,\zeta_2,\ldots,\zeta_{n-1}) \supset S(\zeta_1,\zeta_2,\ldots,\zeta_{n-1},\zeta_n) \quad \zeta_n = 1,2,\ldots$$

$$\text{diam}(\zeta_1,\zeta_2,\ldots,\zeta_n) \leqq \frac{1}{n} \quad \zeta_n = 1,2,\cdots.$$

Since $\mathfrak{N}$ is complete, for each $z \in \mathfrak{N}$, $\bigcap_{s<z} S(s)$ consists of a single point.

Define $g(z)$ to be this point. Thus $N(S) = g(\mathfrak{N}) = \mathfrak{N}_1$. By the arguments of Lemma A.7, g is continuous. QED

Theorem A.15: Let X be a Borel space. A nonempty set $A \subset X$ is analytic if and only if $A = f(\mathfrak{N})$ for some continuous function $f : \mathfrak{N} \to X$.

Proof: If X is complete, the theorem follows from Lemmas A.7 and A.8. If X is not complete, it is homeomorphic to a Borel subset of a complete separable space and the result follows from Corollary A.14.1. QED

The above theorem gives a very useful characterization of nonempty analytic sets in terms of continuous functions and the Baire null space $\mathfrak{N}$. The Baire null space has a simple description and its topology allows considerable flexibility.

<u>Lemma A.9</u>: The space $\mathfrak{N}$ is homeomorphic to any finite or countably infinite product of copies of itself.

Proof: We prove the lemma for the case of a countable infinite product. Let $\pi_1, \pi_2, \ldots$ be a partition of $N$ into infinitely many infinite sets. Define $\varphi : \mathfrak{N} \to \mathfrak{N}\mathfrak{N}\mathfrak{N}\ldots$ by $\varphi(z) = (z_1, z_2, \ldots)$ where $z_k$ consists of the components of $z$ with indices in $\pi_k$. Clearly $\varphi$ is one-to-one and onto. Since convergence in a product space is componentwise, $\varphi$ is a homeomorphism. QED

<u>Theorem A.16</u>: Let $X_1, X_2, \ldots$ be a sequence of Borel spaces and $A_k$ an analytic subset of $X_k$, $k = 1,2,\ldots$ . Then the sets $A_1A_2\ldots$ and $A_1A_2\ldots A_n$, $n = 1,2,\ldots$ are analytic subsets of $X_1X_2\ldots$ and $X_1X_2\ldots X_n$ respectively.

Proof: Let $f_k : \mathfrak{N} \to \mathfrak{N}_k$ be continuous such that $A_k = f_k(\mathfrak{N})$, $k = 1,2,\cdots$. Let $\varphi$ be defined as in Lemma A.9 and $F : \mathfrak{N}\mathfrak{N}\ldots \to X_1X_2\ldots$ be defined by

$F(z_1, z_2, \ldots) = (f_1(z_1), f_2(z_2), \ldots)$. So $F \circ \varphi$ is continuous and maps $\mathfrak{N}$ onto $A_1 A_2 \cdots$. Similar arguments for the finite products. QED

Another consequence of Theorem A.15 is that the continuous image of an analytic set, in particular, the projection of an analytic set, is analytic. It is formalized in the following theorem.

Theorem A.17: Let $X$ and $Y$ be Borel spaces and $A$ an analytic subset of $XY$. Then $\text{proj}_X(A)$ is analytic. Conversely, given any analytic set $C \subset X$ and any uncountable Borel space $Y$, there is a Borel set $B \subset XY$ such that $C = \text{proj}_X(B)$. If $Y = \mathfrak{N}$, $B$ can be chosen to be closed.

Proof: If $A = f(\mathfrak{N}) \subset XY$ where $f$ is continuous $\text{proj}_X(A) = (\text{proj}_X \circ f)(\mathfrak{N})$ is analytic by Theorem A.20. Now, suppose $C = f(\mathfrak{N}) \subset X$ is nonempty and analytic.

$C = \text{proj}_X[\widetilde{Gr}(f)]$ where

$\widetilde{Gr}(f) = \{(f(z), z) \in X\mathfrak{N} : z \in \mathfrak{N}\}$.

Since $f$ is continuous, $\widetilde{Gr}(f)$ is closed. If $Y$ is any uncountable Borel space, then $\exists$ a Borel isomorphism $\varphi$ from $\mathfrak{N}$ onto $Y$.

Define a mapping $\Phi$ by $\Phi(x,z) = (x, \varphi(z))$. Then $\Phi(x,z)$ is a Borel isomorphism from $X\mathfrak{N}$ onto $XY$

$\therefore$ $C = \text{proj}_X(\Phi[\widetilde{gr}(f)])$. QED

<u>Theorem A.18</u>: Let $X$ and $Y$ be Borel spaces and $f: X \to Y$ a Borel-measurable function. If $A \subset X$ is analytic, then $f(A)$ is analytic. If $B \subset Y$ is analytic, then $f^{-1}(B)$ is analytic.

Proof: Suppose $A \subset X$ is analytic. By Theorem A.17, $\exists$ a Borel set $B \subset X\mathfrak{N}$ such that

$$A = proj_X(B).$$

Define $\psi: B \to Y$ by $\psi(x,z) = f(x)$. Clearly $\psi$ is Borel measurable. This implies $gr(\psi) \in \mathcal{B}_{X\mathfrak{N}Y}$

$$f(A) = proj_Y[gr(\psi)].$$

By Theorem A.17, $f(A)$ is analytic. If $B \subset Y$ is analytic, then $B = N(S)$ where $S$ is some Suslin scheme for $\mathcal{T}_Y$.

$f^{-1}(B) = N(f^{-1} \circ S)$ where $f^{-1} \circ S$ is the Suslin scheme for $\mathcal{B}_X$ defined by $(f^{-1} \circ S)(s) = f^{-1}[S(s)]\ \forall\ s \in \Sigma$, $f^{-1}(B)$ is analytic by Theorem A.14. QED

Now, the above results can be summarized in the following theorem.

<u>Theorem A.19</u>: Let $X$ be a Borel space. The following definitions of the collection of analytic subsets of $X$ are equivalent:

(a) $\mathcal{S}(\mathcal{T}_X)$

(b) $\mathcal{S}(\mathcal{B}_X)$

(c) the empty set and the images of $\mathfrak{N}$ under continuous functions from $\mathfrak{N}$ into $X$

(d) the projections into $X$ of the closed subsets of $X\mathfrak{N}$

(e) the projections into $X$ of the Borel subsets of $XY$, where $Y$ is an uncountable Borel space

(f) the images of Borel subsets of $Y$ under Borel-measurable functions from $Y$ into $X$, where $Y$ is an uncountable Borel space.

Proof: Only (f) needs to be shown.

If $Y$ is an uncountable Borel space and $f : Y \to X$ is Borel-measurable. For every $B \in \mathcal{B}_Y$, $f(B)$ is analytic in $X$ by Theorem A.18. Let $\varphi$ be a Borel isomorphism from $Y$ onto $X\mathfrak{N}$ and let $F \subset X\mathfrak{N}$ be closed such that $\text{proj}_X(F) = A$. Define $B = \varphi^{-1}(F) \in \mathcal{B}_Y$. Then $(\text{proj}_X \circ \varphi)(B) = A$ is analytic. If $A = \emptyset$, then $f(\emptyset) = A$ for any Borel measurable $f : Y \to X$. QED

# BIBLIOGRAPHY

## BIBLIOGRAPHY


Akerloff, G., "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism", Quarterly Journal of Economics, August 1977.

Alchian, A., and H. Demetz, "Production, Information Costs, and Economic Organization", American Economic Review, December 1972.

Amershi, A., and J. Butterworth, "The Theory of Agency With Diverse Beliefs", Working Paper, University of British Columbia, 1979.

Ash, R., Real Analysis and Probability, Academic Press, New York, 1972.

Åström, K.J., "Optimal Contral of Markov Processes With Incomplete State Information", Journal of Mathematical Analysis and Applications, 1965.

Baiman, S., "Discussion of Auditing, Incentive and Truthful Reporting", Journal of Accounting Research, Supplement, 1979.

Baiman, S., and J. Demski, "Economically Optimal Performance Evaluation and Control Systems", Journal of Accounting Research, Supplement, 1980.

Bertsekas, D., and S. Shreve, Stochastic Optimal Control: The Discrete Time Case, Academic Press, New York, 1978.

Blackwell, D., "Positive Dynamic Programming", Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1965a.

Blackwell, D., "Discounted Dynamic Programming", Annuals of Mathematical Statistics, 1965b.

Blackwell, D., "Borel-Programmable Functions", Annals of Probability, 1978.

Blackwell, D., D. Freedman, and M. Orkin, "The Optimal Reward Operator in Dynamic Programming", Annals of Probability, 1974.

Dahlquist, G., and A. Björk, Numerical Methods, Prentice-Hall, Inc., New Jersey, 1974.

Denardo, E.V., "Contraction Mappings in the Theory Underlying Dynamic Programming", SIAM Review, 1967.

Demski, J., and G. Feltham, "Economic Incentive in Budgetary Control Systems", Accounting Review, April 1978.

Fama, E., "Agency Problems and the Theory of the Firm", Journal of Political Economy, April 1980.

Feller, W., An Introduction to Probability Theory and its Applications, Volume II, John Wiley and Sons, New York, 1971.

Freedman, D., "The Optimal Reward Operator in Special Classes of Dynamic Programming Problems", Annals of Probability, 1974.

Furukawa, N., "Markovian Decision Processes with Compact Action Spaces", Annals of Mathematical Statistics, 1972.

Gjesdal, F., "Accounting in Agencies", Working Paper, Stanford University, 1976.

Glowinski, R., J.L. Lions, and R. Trémolières, Numerical Analysis of Variation Inequalities, North-Holland Publishing Company, Amsterdam, 1981.

Harris, M., and A. Raviv, "Optimal Incentive Contracts with Imperfect Information", Journal of Economic Theory, December 1979.

Himmelberg, C.J., T. Parthasarathy, and F.S. Van Vleck, "Optimal Plans for Dynamic Programming Problems", Mathematical Operation Research, 1976.

Hirschleifer, J., "The Private and Social Value of Information and the Reward to Incentive Activity", American Economic Review, September 1977.

Holmstrom, B., "On Incentives and Control in Organizations", Unpublished Ph.D. Dissertation, Stanford University, 1977.

Holmstrom, B., "Moral Hazard and Observability", The Bell Journal of Economics, Spring 1979.

Jensen, M., and W. Meckling, "Theory of the Firm: Managerical Behavior, Agency Costs and Ownership Structure", Journal of Financial Economic, October 1976.

Juskevič, A.A., "Reduction of a Controlled Markov Model with Incomplete Data to a Problem with Complete Information in the Case of Borel State and Control Spaces", Theory of Probability Applications, 1976.

Kobayashi, T., "Equilibrium Contracts for Syndicates with Differential Information", Econometrica, November 1980.

Lambert, R., "Managerial Incentives and Short-Run vs Long-Run Optimization", Unpublished Ph.D. Dissertation, Stanford University, 1981.

Luenberger, D.G., Optimization by Vector Space Methods, John Wiley and Sons, New York, 1969.

Maitra, A., "Discounted Dynamic Programming on Compact Metric Spaces", Sankhya, 1968.

Marchuk, G.I., Methods of Numerical Mathematics, Second Edition, Springer-Verlag, New York, 1982.

Mirrless, J., "Notes on Welfare Economics, Information, and Uncertainty", M. Balch, D. McFadden and S. Wu (eds.), Essays on Economics Behavior Under Uncertainty, North Holland Publishing Company, Amsterdam, 1974.

Myerson, R., "Incentive Compatibility and the Bargaining Problem", Econometrica, January 1979.

Ng, D., and J. Stoeckenius, "Auditing, Incentives and Truthful Reporting", Journal of Accounting Research, Supplement 1979.

Radner, R., "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship", Econometrica, September 1981.

Rockafellar, R.T., "Integral Functionals, Normal Integrands and Measurable Selections", Nonlinear Operators and the Calculus of Variations, Springer-Verlag, New York, 1976.

Ross, S., "The Economic Theory of Agency: The Principal's Problem", American Economic Review, May 1973.

Ross, S., "On the Economic Theory of Agency and the Principle of Similarity", M. Balch, D. McFadden and S. Wu (eds.), Essays on Economic Behavior Under Uncertainty, North-Holland Publishing Company, Amsterdam, 1974.

Rudin, W., Functional Analysis, McGraw-Hill Book Company, New York, 1973.

Rudin, W., Real and Complex Analysis, McGraw-Hill Book Company, New York, 1974.

Savage, L., The Foundations of Statistics, John Wiley and Sons, New York, 1954.

Sawaregi, Y., and T. Yoshikawa, "Discrete-Time Markovian Decision Process with Incomplete State Information", Annals of Mathematical Statistics, 1970.

Schäl, M., "On Continuous Dynamic Programming with Discrete Time Parameter", Z. Wahrecheinlichkeitstheorie und Verw. Gebiete, 1972.

Shavell, S., "Risk Sharing and Incentives in the Principal and Agent Relationship", The Bell Journal of Economics, Spring 1979.

Spence, A.M., Market Signalling Information Transfer in Hiring and Screening Processes, Harvard University Press, Boston, 1974.

Spence, A.M., and R. Zeckhauser, "Insurance, Information and Individual Action", American Economic Review, May 1971.

Strauch, R.E., "Negative Dynamic Programming", Annals of Mathematical Statistics, 1966.

Striebel, C., Optimal Control of Discrete Time Stochastic Systems, Springer-Verlag, New York, 1975.

Wilson, R., "On the Theory of Syndicates", Econometrica, January 1968.

# GENERAL REFERENCES

## GENERAL REFERENCES

Arrow, K., Essays in the Theory of Risk-Bearing, Chicago Press, 1970.

Arrow, K., "Political and Economic Evaluation of Social Effects and Externalities", M. Intriligator (ed.) Frontiers of Quantitative Economics, North-Holland Publishing Company, Amsterdam, 1971.

Bellman, R., Dynamic Programming, Princeton University Press, New Jersey, 1957.

Bertsekas, D.P., Dynamic Programming and Stochastic Control, Academic Press, New York, 1976.

Blackwell, D., "On Stationary Policies", Journal of Royal Statistics Society, 1970.

Demski, J., "Cost Allocation Games", S. Moriarity (ed.) Joint Cost Allocations-Proceedings of the University of Oklahoma Conference on Cost Allocations, Center for Economic and Management Research, Normal, Oklahoma, 1981.

Feltham, G., "Optimal Incentive Contracts: Penalties, Costly Information and Multiple Workers", Working Paper, University of British Columbia, 1977.

Fleming, W., and R. Rishel, Deterministic and Stochastic Optimal Control, Springer-Verlag, 1975.

Hinderer, K., Foundations of Nonstationary Dynamic Programming with Discrete Time Parameter, Springer-Verlag, 1970.

Hoffman-Jorgensen, J., The Theory of Analytic Spaces, Aarhus Universitet, Aarhus, Denmark, 1970.

Huston, V., and J. Pym, Applications of Functional Analysis and Operator Theory, Academic Press, New York, 1980.

Ioffe, A.D., and V.M. Tihomirov, Theory of Extremal Problem, North-Holland Publishing Company, Amsterdam, 1979.

Kreps, D.M., and E. Porteus, "Dynamic Choice Theory and Dynamic Programming", Econometrica, January 1979.

Kreps, D.M., and R. Wilson, "Sequential Equilibria", Econometrica, July 1982.

Larsen, R., Functional Analysis, Marcel Dekker Inc., New York, 1973.

Mirrless, J., "The Optimal Structure of Incentive and Authority Within an Organization", The Bell Journal of Economics, Spring 1976.

Ng, D., "An Information Economics Analysis of Financial Reporting and External Reporting", Accounting Review, October 1978.

Ng, D., "Supply and Demand of Auditing Services and the Nature of Regulations in Auditing", S. Davidson (ed.) The Accounting Establishment in Perspective, Arthur Young and Co., 1979.

Radner, R., "Competitive Equilibrium Under Uncertainty", Econometrica, 1968.

Royden, H.L., Real Analysis, Macmillan, New York, 1968.

Scott, W., "A Bayesian Approach to Asset Valuation and Audit Size", Journal of Accounting Research, Autumn 1973.

Stiglitz, J., "Incentives and Risk Sharing in Sharecropping", Review of Economic Studies, 1974.

Stiglitz, J., "Incentives, Risk and Information: Notes Towards a Theory of Hierarchy", The Bell Journal of Economics, Autumn 1975.

Yoshida, K., Functional Analysis, Springer-Verlag, New York, 1980.