NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

UNIVERSITY OF ALBERTA

PRODUCTION AND MONITORING INCENTIVES IN CHINA'S COLLECTIVE FARMING: THEORY AND EVIDENCE

by

Dong Xiao-yuan

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

EDMONTON, ALBERTA

Spring 1991

ISBN 0-315-66722-2

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Dong Xiao-yuan

TITLE OF THESIS: Production and Monitoring Incentives in China's Collective Farming: Theory and Evidence

DEGREE FOR WHICH THESIS WAS PRESENTED: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: Spring 1991

Dong Xiao yuan

(Student's Signature)

Department of Economics

(Student's Permanent Address)

Heilongjiang University

Harbin, China

Date: April 5, 1991

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptence, a thesis entitled **PRODUCTION AND MONITORING INCENTIVES IN CHINA'S COLLECTIVE FARMING: THEORY AND EVIDENCE** submitted by Dong xiao-yuan in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Greg Dow

Adolf Buse

Terry Veeman

Brian Evans

Louis Putterman

Date: Apr. 4. 1991

DEDICATION

To My Parents

Dong Guixin and Li Yuying

ABSTRACT

Since 1978, China's agriculture has undergone dramatic economic reform. One important change was the dismantling of collective farming units after nearly a quarter century of practice in China. The widely held explanation of this change is that collective farms cannot generate adequate work incentives because of the special problems of labour supervision in team farming. This argument, however, focuses narrowly on the technological aspects of team farming, but ignores broader problems of optimal incentive design.

This thesis develops a theoretical and empirical model to address the incentive problems of China's rural teams. A simple model of static mutual monitoring is developed. The model endogenizes the determination of the size of penalties as well as the choice of monitoring intensity in the context of team optimization. The processes of formal and informal mutual monitoring, and individual and group incentives, are then integrated by extending the static monitoring model to a production team with an infinite time horizon. It is demonstrated that the possibility of rent is a necessary condition for an effective monitoring system. Any policy that represses the financial returns to the team will make labour supervision more difficult.

The theoretical model provides several testable hypotheses: (i) Chinese collective farmers faced a free-rider problem; (ii) teams invested resources in labour monitoring to cope with the resulting incentive problems in contrast to the prevailing belief that supervision in Chinese production teams was negligible; (iii) the recognition of mutual interdependence among individual workers has significant impact on the worker's effort decisions; (iv) the internal incentive problems of Chinese collective farms were not the simple outcome of team farming per se, but rather resulted from dysfunctional economic and political policies adopted during the commune era.

A 1970s' panel data set from a commune in China is used to test these hypotheses. The static and dynamic versions of an implicit function determining the team's optimal allocation of labour time between production and monitoring are estimated by the nonlinear two-stage least squares method. The empirical results provide a broad pattern of support for

the above hypotheses. These results reject the conventional belief that no effective monitoring system can be implemented in a farming team with shared residual claims. By highlighting the complexity of the incentive problems in China, the findings of this thesis have important implications for the broad issues concerning the choice of proper economic reform and development strategies in China.

ACKNOWLEDGEMENT

I owe both theoretical and personal debt to my supervisor, Professor Greg Dow, who took my research interest so seriously and spent countless hours reading and commenting upon this thesis draft. Without his outstanding knowledge and guidance, this thesis would have been inconceivable.

I am grateful to Professor Adolf Buse for his essential guidance and inspiration that helped shape the empirical work of this thesis, to Professor Terrence S. Veeman for his advice and clarification of conceptual issues in the field of economic development, and also to Professor Brian Evans for his endorsement and valuable comments.

Professor Louis Putterman of Brown University provided me with the data for this study. He has been a constant source of encouragement and criticism. His insightful comments have helped me place this dissertation and my thinking in broad perspective.

On behalf of my family, I particularly wish to thank Professors Terrence and Michele Veeman, who kept my morale up not only through their financial support every summer during my Ph.D program but also through their constant encouragement.

The academic training and financial support of the Department of Economics, University of Alberta is greatly appreciated. I would like to thank Dr. Melville McMillan, Dr. Michael Percy, Dr. Beverly Dahlby, Louise Edwards, Leslie Wayne, and Margaret Howell for their encouragement and help during my study in the department. I also wish to acknowledge the help from Alan Sharpe with respect to computer programming.

I owe a special debt to Heilongjiang University, China for the opportunity and financial support which enabled me to study in Canada. I am also thankful to my teachers and colleagues in the Department of Economics, Heilongjiang University for their support and inspiration.

I am grateful to my brother and sisters, Yubin, Zhuangli and Xiuli, for fulfilling my duty of taking care of our parents. I am deeply thankful to my parents in law, Jinshi Shen and Fenglan Zhang, for their encouragement and support.

I could not have survived my graduate school and this dissertation without my husband, Haiying Shen, whose understanding, patience and sacrifice are beyond telling. I benefitted greatly from his insights drawn upon his working experience in team management. My gratitude to my daughter, Jia, who has tolerated a life without Mom's due attention, cannot be adequately expressed.

Table of Contents

List of Tables

Table | Page

Chapter 1

Introduction

Since 1978, China's agriculture has undergone dramatic economic reform. One important change is the dismantling of collective farming units and restoration of farming by households after nearly a quarter century of practice in China. It is widely recognized that this dramatic institutional change took place because collective farms had failed to generate adequate work incentives. However, there are sharp disagreements concerning the causes of the incentive problems in Chinese collective farms. Nolan (1988) and Lin (1987,1988) attribute the incentive problems to the special difficulty of labour supervision in team farming, indicating that collective farms are not a viable alternative to private ownership for farm modernisation. Nevertheless, Lardy (1983), Blecher (1985) and Putterman (1985, 1987) argue that despite the difficulty of labour monitoring, democratic producers' cooperatives need not be inferior to family farms. The incentive failure of Chinese collective farms was the outcome of the nation's bureaucratic controls, extractive government policies and anti-incentive ideology. Because China's rural institutional experiment is of historic importance, this thesis develops a theoretical and empirical model to assess these competing arguments concerning the incentives of team farming in China.

The thesis proceeds as follows. An historical background for the incentive problems of China's collective farms is provided in chapter 2. I discuss why China adopted collective farms as an alternative to private household farms for rural development and why such an institutional choice gained considerable popular support from Chinese peasants in the mid-1950s. I further document how the performance of Chinese collective farms was influenced by policy fluctuations during the period 1953-77, and discuss why in many circumstances the collective farms were operating quite inefficiently and were unable to provide adequate incentives for the peasants. The conclusion following the historical review is that Chinese collective farms were not inherently inefficient rural institutions; it was the adoption of dysfunctional policy regimes that aggravated the difficulties of labour supervision facing Chinese rural teams and undermined the peasants' incentives to work hard.

Chapter 3 reviews the literature on the incentive systems of worker teams. I identify three issues that have not been dealt with adequately in the labour supervision literature. The failure to handle these issues properly leads to an unduly pessimistic view of the possibilities for monitoring when peasants hold shared residual claims. The first involves comparison between central monitoring and mutual monitoring. The pessimist impicitly assumes that monitoring must be undertaken by a specialized agent. However, mutual monitoring among workers helps reduce monitoring cost by narrowing the information asymmetry between the principal and agents that exists under a central monitoring system. Hence, the neglect of mutual monitoring leads to an exaggeration of the difficulties of labour supervision in group farming. The second problem with existing discussions of labour supervision is that they usually take an ad hoc approach to the income distribution scheme. The formulation of the payment rules and the specification of the monitoring system have often been treated as separated problems. This has led to excessive emphasis on the role of monitoring per se as compared with constraints on the rewards or penalties to which individual workers may be subjected. As a result, the influence of the policy regime on the internal incentives is often ignored. Finally, the monitoring problem is usually discussed within the framework of a one-shot noncooperative game. However, most production teams involve long-term relationships. It is well known that under certain conditions economic agents can achieve much better actions in a long-term relationship than in a one-shot game.

In chapter 4, a simple model of static mutual monitoring is formulated to address the question of work incentives in a production team. The major innovations of the chapter are twofold. First, I specify a formal model of mutual monitoring in which each worker engages in two tasks: productive labour and monitoring the effort of the other workers. Second, in contrast to the ad hoc choice of an income distribution scheme found in most models of labour managed firms, this model integrates the selection of a payment scheme with the choice of monitoring intensity in a team's optimization program. It is demonstrated that the possibility of rent is a necessary condition for an effective monitoring system. Any policy that represses the financial returns to the team will make labour supervision more difficult.

In chapter 5, a panel data set from a commune in China in the 1970s is used to examine the technical, economic and ideological constraints on the incentives of China's production teams. An implicit function determining the team's optimal allocation of labour time between production and monitoring is estimated by the nonlinear two-stage least squares method. Several hypotheses are tested. These hypotheses are: (i) Chinese collective farmers faced a free-riding problem; (ii) teams made rational decisions with regard to monitoring intensity, in contrast to the prevailing belief that supervision in Chinese production teams was negligible; and (iii) the internal incentive problems of Chinese collective farms were not the simple outcome of team farming per se, but rather resulted from dysfunctional economic and political policies adopted during the commune era. The empirical results provide a broad pattern of support for the above hypotheses.

In chapter 6, a simple model of dynamic mutual monitoring in a worker team is developed. What is unique in this model is its integration of formal and informal labour supervision. The pure dynamic monitoring system and the pure static monitoring system that have been discussed in the literature are special cases of this general monitoring model. Compared with established dynamic monitoring models in the literature, the present set-up is argued to be more realistic, because it is observed that formal labour supervision is conducted in most worker teams. What we attempt to show is that the group incentives and the long-term relationships among workers can supplement the static formal monitoring system and improve the efficacy of labour supervision in cooperative production.

In chapter 7, the dynamic incentive system developed in chapter 6 is estimated with the 1970s' panel data from a commune in China. The effect of dynamic monitoring upon the incentives of Chinese production teams is examined. It is found that the recognition of mutual interdependence among individual workers had a significant impact on the workers' effort decisions. The fear that shirking would be punished by informal retaliation helped reduce the team's monitoring cost, and therefore permitted more time to be allocated to physical production activities. The hypotheses that were tested in chapter 5 are re-examined under the dynamic specification. The test results are, however, inconclusive.

Finally, in chapter 8 the results are summarized and the implications for the broad issues concerning the choice of economic institutions and development strategies are discussed.

Chapter 2

Historical Review

2.1 Introduction

The Chinese commune system was the largest experiment with cooperative production in human history. As an alternative approach to industrialization and rural development, the Chinese collective farming system was shaped by China's factor endowments and communist ideology. The evolution of the system reflected China's changing development strategies. Therefore, the performance of team farming in China must be evaluated from an historical perspective.

The chapter is organized as follows. In section 2, China's economic setting and rural institutions prior to 1949 are introduced, and the factors that determined China's development strategies are identified. In section 3, I address the issue of why cooperative farming gained popular support from Chinese peasants in the mid-1950s, by reviewing the historical events of land reform and collectivization. In section 4, the policy fluctuation during the period 1958-77 is documented. Section 5 analyzes how the incentive system of Chinese production teams developed in response to changes in the external environment. Section 6 summarizes and concludes the discussion by assessing the general performance of the Chinese commune system in the process of industrialization.

2.2 Economic Setting and Agricultural Policies

2.2.1 Economic Setting and Rural Institutions prior to 1949

China had been a traditional agrarian country prior to 1949. It was estimated that the GNP per capita of China was less than US $50 in the 1940s (Dernberger and Fasenfest, 1978). This was much lower than the income level at which the western developed countries started their industrialization process in the late 18th and early 19th centuries (Eckstein, 1978). China's productive capacity had suffered severe damage during the Japanese invasion and the civil war. Even in the 1930s industrial output accounted for only 10 percent of China's gross national product. Most of that was concentrated in textiles and other light industries. There was virtually no machine-building industry and only a limited amount of repair and assembly work. But in 1949 industrial output fell to only one-half its prewar peak. The industrial workers in 1949 accounted for only 1.5% of the national population[1].

Agriculture was even more backward, with its increasing rural population coupled with a scarcity of arable land. The growing agrarian population and the inheritance system contributed to dwindling farm size and excessive land fragmentation. Most farms were too small for efficient operation.[2] The limited arable land, however, faced great variability of precipitation, uneven distribution of surface water and uncertain or insufficient provision of underground water (Prybyla, 1978). In consequence, throughout history Chinese farms had been plagued by the twin calamities of flood and drought. Agricultural production fluctuated over time as a result.

The basic rural institution was the land-tenure system. An extremely unequal distribution of land-ownership and high rental rates had been typical of China's land-tenure system. The landlords and rich peasants made up less than 10% of the population, but owned more than 70% of the total arable land, while poor peasants formed 70% of the village population, but owned only 10% of the land. The land rent ranged from 50 to 80% of the crop (Xiu, 1982). The land rental income made little contribution to rural growth, because three-quarters of all rented out land was owned by the absentee landlords, who were military or government officials, moneylenders and businessmen in commerce and industry (Perkins, 1969). By extracting economic surplus from the peasants, the unequal land distribution exacerbated agricultural stagnation and chronic and deep-seated rural poverty (Riskin, 1975).

[1]See China Handbook Editorial Committee, 1984, **Economy**, Foreign Languages Press, Beijing.

[2]It was estimated that in the 1930s, 33% of Chinese farms had less than 1.6 acres, 25% had 1.6 to 3.3 acres, and only 8% were larger than 8.4 acres (Chen, 1982). According to a sample of 16,786 farms in 22 provinces, the average number of parcels per farm was 5.6 (Buck, 1937).

As a result, the land problem represented a fundamental political issue in China's modern history. In the bourgeois revolution, Dr. Sun zhongshan, founder of the Nationalist Party, proposed the guideline of "land to the tillers". Unfortunately, little was accomplished in this task before the end of Civil War, and the issue was taken over by the Chinese communists.

2.2.2 The Party's Development Strategies

In 1949 the Chinese communists gained control over the country. The new government's approach to the development of agriculture stemmed in part from communist ideology and the Stalin-type development strategy. Following the logic of Marxism, after the proletariat takes over power, the immediate tasks are threefold. First, a centrally planned economy should be established to avoid the "anarchy" of the market system. Second, private ownership of the means of production should be abolished, for it is thought to be the primary source of class conflict. Third, the proletariat should try its best to accelerate the speed of economic growth to demonstrate the superiority of the socialist system. Marx did not take sufficiently into account the special features of agriculture. He considered the development of agriculture in the same way he considered the development of industry, and asserted that mechanization would revolutionize agricultural production and the future would belong to large agricultural enterprises (Wittfogel, 1971).

Lenin led the first socialist revolution in Russia, a backward capitalist country with a large population of individual peasants. Lenin foresaw the conflict between private farms and central planning, and the polarization between the rich and poor due to the private ownership of land. Hence, he contended that in countries where communists came to power, peasants could be led to socialism without passing through a capitalist stage of development.

Stalin actually conducted the socialist transformation and the socialist construction of the Soviet economy. In the Stalinist regime, the major economic objective was rapid industrialization to catch up with industrially developed countries. Therefore, the main emphasis was placed on growth of a modern heavy industrial sector; and agriculture was viewed as an accommodating agent (Nove, 1971).

Despite the influence of the Marxist-Leninist ideology and the Soviet development model, China's acute economic realities were the ultimate factor that determined the Chinese communist government's development strategies. Industrialization of a vast agrarian economy such as China requires a huge amount of capital investment. Foreign aid is an important source which has helped many developing countries break domestic economic constraints and start industrialization. However, due to political and historical reasons, sizable foreign aid was not available for China until the early 1980s. As a result, the Chinese government had no alternative but to rely on domestic sources to start industrialization. In other words, the success of industrialization in China crucially depended upon the ability to transfer a surplus from the dominant agricultural sector for investment and capital accumulation in the industrial sector.

Nevertheless, the Chinese communists could not afford to copy exactly the Stalinist strategy, for agriculture's restrictive hold on industrialization progressively tightened in China. Chinese per capita food availability in 1952 was less than half that of the Soviet Union in 1928. So the Chinese agricultural problem was one of production, while the Soviet problem had much to do with marketing. Therefore, the Soviet agricultural policy in the early stage of the Stalinist regime was extractive, whereas the Chinese leaders found it necessary to adopt a policy which was both extractive and developmental (Tang, 1985). The Chinese government needed a delicate economic policy which balanced industrialization and rural development properly so that agriculture was able to grow fast enough to meet the demand of industrialization and the basic human needs of its mass population as well. This can never be an easy task given the harsh realities of poverty and the scarcity of land, capital and skilled manpower imposed by China's factor endowments.

2.3 Land Reform and Collectivization

2.3.1 Land Reform and Its Consequences

As soon as the Chinese communists came to power, they kept their promise to the peasants and carried out a program of land reform. The land reform redistributed 115.5 million acres of land (46.5% of the total cultivated land) to the peasants. The burden of annual land rental payments relieved accounted to 35 million tons of grain, 31% of the total grain output in 1949. The land reform created incentives for increasing agricultural production and consolidated the Party's political power in the rural areas.

However, the land reform brought about a series of new problems. The land reform granted equal shares of land to peasant households and increased the number of landholders by more than 50 million. The total number of household farms in 1952 was 110 million (Chen, 1982). In consequence, average farm size was reduced and land fragmentation was aggravated. The egalitarian land redistribution also caused serious shortages of farm inputs for small owner-cultivators. According to an investigation of 15,000 peasant households in 23 provinces in 1954, each household averaged 0.82 hectare of farmland, and there were approximately two draft animals for every three households, one plow for every two households and one waterwheel for every ten households (Yan, 1981). Small scale family farming was also threatened by natural disasters, illness and death in the family. Hence, many household farms were not able to be self-sustaining over time.

As a result, there was a strong tendency toward class stratification in the rural sector. Some poor peasants were soon forced into debt and land tended to become once again concentrated in the hands of rich peasants. It was reported that one-third of the poor peasants in 10 townships in Hubei, Human, and Jiangxi provinces and half of them in several townships in Guangdong province were in debt. The interest rate on these loans ranged from 10 to 100% (Yan, 1981). A survey of 143 villages by the Xinxian prefecture communist party committee in Shanxi province showed that 19.5% of total peasant households sold their land and houses in 1949 immediately after the land reform, and the number jumped to 25.25% in

1952 (Shi, 1987). One feature of the re-stratification was that among the rising newly rich peasants, the majority were rural party cadres. A significant proportion of rural party cadres engaged in labour hiring, usury, speculation, and profiteering with their political privilege. This damaged the prestige of the party with Chinese peasants (Shi, 1987).

The constraint of agriculture on economic development became apparent as the First Five Year Plan was initiated in 1953. The major objective of FYP-1 was the rapid development of a heavy industrial base. The expansion of modern industry and the urban population caused an excess demand for grain and raw materials. Many light industries failed to run at full capacity due to the serious shortage of agricultural raw materials.

In sum, the land reform provided primarily a means for overthrowing the former ruling class in the rural areas, but did not offer a solution for the conflict between economic growth and equity in the rural area.

2.3.2 Traditional Practices of Mutual Aid and Cooperation

Mutual aid and cooperation among peasants has a long history in China. In old times, labour exchange and draft animal sharing among groups of peasants and cooperation during busy sowing and harvesting seasons were adopted to overcome shortages of means of production, funds and labour (Myers, 1975). There were occasionally examples where a few peasant households with more or less the same factor endowments pooled land together and conducted unified farm management (Xiao, 1987).

However, mutual aid and cooperation did not become a common and regular practice until the communist revolution took place. In 1920-30s, mutual-aid working groups were formed by peasants in the revolutionary base areas of Jiangxi province. During World War II, such organizations as "labour-exchange teams", "teams for the exchange and hiring of labour", and "peasant cooperatives" appeared in Northern Shanxi. By the year 1949, numerous mutual aid teams had been established voluntarily in North and Northwest China. By the end of 1952, there were over 8 million mutual aid teams; over 65 percent of the peasants in the old liberated areas and 25 percent in the new had been organized on a mutual aid basis (Adler, 1957). The teams were formed primarily to exchange draft animals, implements, and manpower during the busy season. Compensation for labour, implements, and draft animals borrowed from other member households was settled by members on mutually agreeable terms. Mutual aid teams helped poor peasants break the shortage of manpower and draft animals and thus increased their output.

In the later 1940s, some "spontaneous" producers' cooperatives came into being. These cooperatives were set up by the peasants of their own accord without any government initiation or approval. These cooperatives engaged in farming, food processing, transportation, carpentry, blacksmithing, etc. The internal management and distribution structures of these coops varied with production technology and peasants' commitment. For instance, 32 peasant households in Shanliu village in Guang Rao county in Shangdong province collectively reclaimed 740 mu of land in 1946. To operate the land on an efficient scale, the peasants put the newly reclaimed land under unified management, while each household also had its own original private farm land. After a year, having noticed that some peasants did not devote as much effort to the coops' land as they did to their privately owned land, the peasants decided to put all farmland under collective operation (Xiao, 1987). These "spontaneous cooperatives" were generally run very well, producing higher outputs and income. The practice of mutual-aid and cooperation indicated that cooperative production could be a promising option to achieve rural growth and equity for Chinese peasants.

2.3.3 Success in Collectivization

Seizing upon the traditional practice as the first step, the Chinese leadership initiated its collectivization program in 1953. Learning from Russia's experience, the CCP leader, Mao Zhedong, emphasized policies of stage-by-stage progression, voluntary membership, and mutual benefit. The original plan was to accomplish co-operativization in 15 years with one-third of the cooperatives established in each five-year plan period. These policies allowed peasants gradually to adjust themselves to the new system and learn how to manage farms on a large scale.

Moreover, attention was paid to balanced growth between industry and agriculture. The speed of industrialization was controlled so as to lessen exploitation of the peasants. The capital accumulation rate during the period was 24.2 percent of national income, lower than most of the rates in later periods. China's agricultural tax rates were set relatively low compared with the Stalinist regime (Dernberger, 1982). The Chinese government manipulated a "scissors" gap between the prices for industrial and agricultural products to transfer economic surplus from agriculture. Yet to provide production incentives for the peasants, the prices of procurement quotas were raised at an annual rate of 4 % from 1952 to 1957 (Perkins and Yusuf, 1984).

Furthermore, the state provided producers' cooperatives with technical, material and financial supports. During FYP-1, state direct investment in agriculture amounted to 4.18 billion yuan or 9.7% of total state investment funds. The state also facilitated 7.6 billion yuan in rural credit supplies for production and investment at low interest rates. In 1954, 4,549 agricultural technological stations were established to help producers' cooperatives adopt new farm technology. Meanwhile, 50,000 rural accountants were trained to improve the management of the newly established co-ops (Wu, 1985).

With the avoidance of Russia's mistakes, collectivization in China succeeded rather smoothly. By the end of 1953, 15,000 elementary producers' cooperatives had been established. Elementary agricultural producers' cooperatives were several times larger than the mutual aid teams. The system was characterized by a pooling of land as shares under a unified management. Income was distributed by both land shares and work effort. The average size of elementary producers' cooperatives was 20 to 30 households. Most of the cooperatives increased their production during the next few years. According to a survey of 16,000 households in 25 provinces in 1954, the average return of the cropland of cooperatives was 18.5% higher than that of the individual poor peasants, while the return of the rich peasant households farms was 14% more than that of the individual poor peasants (Yan, 1981). Cooperative farming was thus attractive not only to poor peasants but also to well-to-do peasants.

In consequence, the rate of cooperativization exceeded the party's expectation, although the party had dissolved numerous new cooperatives in the spring of 1955 to ensure the quality of cooperatives. By the fall of 1955, 634,000 elementary cooperatives had been established, 41.3 times the figure in 1953, and 16,920,000 households joined the coops, 61.5 times the number in 1953. It was reported that per acre productivity on cooperative farms in 1955 was 10 per cent higher than on individual peasant farms in the case of rice, 7 percent higher for wheat, 19 percent higher for soyabeans and 26 per cent higher for cotton (Adler, 1957).

By the end of 1955, 529 advanced producers' cooperatives were established, organizing 40,000 households. Advanced producer cooperatives represented a higher level of collectivization similar to the Soviet Kolkhoz. In this system, peasants' land and other principal means of production were transferred from private to collective ownership, and payment for land shares and other means of production was abolished. On the average, an advanced cooperative embraced 158 peasant households, about four to five times the size of elementary cooperatives (Chen,1982). Most of these new advanced producer cooperatives made further improvement in their production (Xiao, 1987).

The success in collectivization led to an overestimate of the peasants' enthusiasm for socialism on the part of the CCP leadership. In his "preface to the socialist upsurge in China's countryside" in December 1955, Mao Zhedong modified the original schedule of collectivization and proposed that elementary agricultural cooperativization be basically accomplished by 1956 and that advanced agricultural cooperativization be basically accomplished by 1959 or 1960. This fueled the already high speed of collectivization. The peasant households organized into agricultural producers' cooperatives leaped from 60% of the total in 1955 to well over 96% by the end of 1956, of which households in advanced agricultural coops accounted for 87.8%(Yan, 1981). Collectivization was then basically completed.

The overly hasty pace of collectivization gave rise to a series of problems for newly established coops. Many new coops lacked managerial personnel in organising team production. Too fast a pace in collectivization inevitably led to the violation of the principles of voluntary membership and mutual benefits. In consequence, the number of draft animals fell and sideline production dropped in many rural areas. The party had to send work teams to rectify these mistakes in 1957.

Despite the haste and its undesirable consequences, collectivization in China was a remarkable achievement. The tremendous institutional changes did not bring about social disorder; in fact, agricultural production continued to increase in this period. It was estimated that without significant growth in modern inputs, 80 percent of the villages increased their production after cooperatives were established (Wu, 1985). During FYP-1 agricultural production grew at an annual rate of 4.5%. The total productivity of agriculture rose steadily over the period (Tang, 1980). As a consequence, peasants' income increased by 30% in 1957, compared to that in 1952. The experience of producers' cooperatives during the period had left Chinese peasants with a fine memory. Thirty years later when collective farms were dismantled, in an interview with a "Bejing Review" correspondent, a 43-year-old peasant, Wang Xin, made the following statement:

> "In 1951 the agricultural collectivization movement got underway in my village. We first got organized into mutual-aid production teams and then into elementary agricultural cooperatives, pooling our land and sharing the dividends. In 1956 we switched to the advanced agricultural co-operatives and put our farmland into public ownership."......"With the land under public ownership all the villagers met to discuss how to use their farmland and how to distribute the income. This was completely different from pre-liberation days when we had no land at all. During those years since every one worked hard and the government provided the co-operative with preferential loans and farm tools, production grew rapidly. The grain output for instance, grew from 2,250 kg per hectare before 1949 to 4,225 kg in 1956. I remember my family got more than enough wheat that year. We lived quite well during those years."(Bejing Review, 1984)

2.3.4 Economies of Scale, Incentives, and Policy Regimes

It has been a conventional belief in the west that cooperative production is not a viable alternative to private household farming because there are no economies of scale in a traditional agrarian economy. Also following the logic of Marxism that advanced productive forces are the prerequisite for socialist institutions, some Marxists in the socialist countries assert that producers' cooperatives are too advanced to be suitable for a poor country like

China. The success of collectivization in China provided evidence contrary to these popular arguments. If cooperative farming had offered no benefits to Chinese peasants at all, it could not have gained considerable popular support from Chinese peasants in the mid-1950s.

Despite the prevalence of household farming in the world, China's unfavourable factor endowments determined the special desirability of cooperative farming in the country. It is well known that China has nearly 25% of the world's population with only 7% of the world's cultivated land. With more than 85% of the population living on farms, the land-farm worker ratio of China is among the lowest in the world. Moreover, China is a large and undeveloped country; even over long time horizons it is unlikely that the economy would provide sufficient employment outside of agriculture to relieve the population pressure on farmland. These two factors limited the ability of owner-operated family farms to achieve a unimodal rural growth in China. As was evidenced by the land reform, household farming based on egalitarian land distribution implied exceedingly small land holdings and excessive land fragmentation. Nevertheless, to achieve an efficient scale of farm size, a large proportion of the rural population had to become landless and jobless. The resultant rural inequality might undermine the social stability which is the prerequisite for sustained economic growth. China perhaps is the last country in the world which can afford to waste its scarce land resources and pay no attention to rural inequality. Cooperative farming offered a practical option for Chinese peasants to obtain a rational factor combination without paying the cost of severe rural inequality.

The optimal size of farms is an elusive economic issue in the literature, yet there has been little doubt about the adverse impact of land parcellation on China's agriculture. Buck (1937) notes that, traditionally, Chinese agriculture suffered severely from a high degree of fragmentation, "the ownership by individuals of scattered pieces of land..... Such fragmentation has the disadvantages of using up land in boundaries, increasing the number of boundary disputes, consuming time to reach the plots, increasing the difficulties of irrigation, limiting the size of fields and hence the use of machinery, and making crop protection difficult." With the statistical data of North China for 1937-40, Dittrich and Myers (1971) show that the individual peasants were allocating their land and labour efficiently, but operated under increasing return to scale. In an empirical study on eight agrarian areas of China in the 1930s, Chinn (1977) notes that an increase in land fragmentation had a significantly negative impact on private farm productivity. The egalitarian land distribution made farm size even smaller; therefore, the scale economy in land utilization was substantial under joint farming. It was estimated that cooperative farming added 13 million acres to China's arable land simply by wiping out boundaries and unnecessary paths between fields (Adler, 1957). Larger units of farm management make more efficient land utilization possible.

The second advantage of cooperative farming was the better utilization of indivisible farm inputs. Given exceedingly small farm size, it is too costly for every household to own a complete set of draft animals and farm implements. Under the land tenure system, the indivisible farm capital was pooled by having the landlords supply draft animals and implements to his tenants or other farmers on a rental basis. However, the transaction costs of a rental market in capital equipment can be very high. Since it may be difficult to tell by inspection whether the renter has taken due care and attention, the problems of moral hazard are likely to be severe. Requiring a full compensation implies high rent which may drive small tillers into bankruptcy. Moreover, the demand for the services of farm capital inputs may be correlated due to the seasonality of agricultural production. In other words, in peak seasons, there will be excess demand and no farmers have surplus capacity, while in slack seasons, the situation is reversed. In both cases, the market will experience no trade. As a result, having their own draft animals has been a common dream of Chinese small cultivators. In cooperative farms, costly transactions for the services of capital equipment are replaced by an integrated farming arrangement.

Cooperative production also facilitates efficient utilization of the rural labour force. Georgescu-Roegen (1960) compares the labour-hiring decision of a capitalist farm with that of a family farm in an overpopulated economy where the marginal productivity of labour falls below the minimum subsistence income. Georgescu-Roegen contends that to maximize its profit, a capitalist farm will never hire labour beyond the point where the marginal productivity of labour is equal to the subsistence wage, whereas a household farm will utilize family labour to the point where the marginal productivity equals zero in order to maximize output. If the capitalist farms were prevalent in an overpopulated economy, there would be sizable rural unemployment unless wages were below subsistence. Given the fact that unemployed labour still consumes social resources, capitalist farms could create considerable waste of labour resources. Chao (1970) suggests that due to the inability to dismiss members, China's producers' cooperatives can preserve the merits of a household farm with respect to labour utilization. According to Chao's argument, cooperatives are not only able to fully utilize the rural labour force to maximize current production, but also hold the promise of mobilizing the rural labour force for rural capital formation and rural industry, which individual household farms were not able to conduct. Additionally, cooperative farming encourages labour coordination, division and specialization.

Producer cooperatives also provide a mechanism for dispersing risk among the team members. In the case of small private household farming, the well-being of peasant households depends upon the state of the family members and their animals. Illness and death of the family's labour suppliers and animals during the peak season not only directly impose financial burdens on the family but can also destroy the crops. This type of risk can be handled by the rural credit and insurance network. However, in poor countries, the rural credit and insurance markets are highly localized, so that the suppliers of credit and insurance schemes in rural areas face high risks and operational costs. Consequently, the service charges for credit and insurance schemes are too high for many small family farms. With collective management for manpower and draft animals, producers' cooperatives offer an automatic mechanism of pooling this type of uncorrelated risk among individual peasants. Moreover, with a large operating unit, producers' cooperatives are more able to absorb losses stemming from risky innovation and thus less reluctant to adopt new farming technology.

Another advantage of cooperative farming is its ability to provide public goods such as flood control, irrigation, experimentation with new varieties of crops, and rural health care and education. Collective ownership overcomes the major economic obstacles to these types of cooperation by independent peasants. It has been argued that the services of agro-technicians, doctors and school teachers can be provided by the local government and financed through taxation. Nevertheless, the cost of this pure governmental approach might also be rather high in a poor country. With a huge gap between the living standards of the urban and rural sectors, the financial return to specialized agro-technicians, doctors and teachers must be high enough that these intellectuals would like to serve the rural populace. Without sizable government financial subsidies, a large proportion of the rural population could not afford these services. By pooling limited financial resources and making a proper division of labour, producers' cooperatives are able to finance the training of the local peasant technicians, barefoot doctors and part-time school teachers, and then make the services of technological diffusion, health care and education available to the rural population at a relatively low cost.

Cooperative farming internalizes numerous market transactions, the cost of which is rather high in a backward country. However, there were agency costs associated with cooperative production. A crucial issue facing producers' cooperatives is how to provide adequate incentives for all individual participants. The internal incentive system of a collective farm in a poor nation is subject to two handicaps. First, the metering of individual contributions is costly due to the nature of agricultural production. Second, low per capita income may narrow the feasible reward scale for discriminating with respect to labour input performance. However, the institutional design of cooperative production introduces a group incentive mechanism which can supplement the inadequate individual incentives.

An effective collective incentive system depends on the following conditions. First, the team size should not be too large, so that the team members have the incentive and ability to make sure each does his or her share of the work. Second, a collective farm must be able to control its own production process and to reap the benefit of improving team production. Third, the workers must form stable long-term relationships. Fourth, there should be sufficient coordination and communication among team workers. For these reasons, the internal incentive problem of a farming team is not a merely technical issue. To overcome the potential disadvantage of team farming in labour supervision, considerable attention must be paid to the material incentives for individual workers as well as for teams as groups.

The experience in the later period of collectivization revealed certain unhealthy tendencies in the CCP's economic policies, which were detrimental to the establishment of effective internal incentive systems in Chinese collective farms. First, the party was not able to make a proper distinction between class-stratification on the one hand, and income inequality necessary for preserving optimal work incentives, on the other. The fear of re-stratification in the rural area could spontaneously lead to rejection of material incentives and encouraging egalitarianism in producers' cooperatives. Second, the party leadership did not quite understand the limitation of economies of scale in agricultural production. In the view of the party, advanced socialist production relations, characterized by large scale and a high degree of public ownership, could stimulate economic growth. Third, due to similar ideological reasons, rural collective ownership was regarded as an imperfect socialist production relation which was subject to transformation. This view not only encouraged political encroachment on collective ownership but also imposed institutional uncertainty on Chinese collective farms. Fourth, following the doctrine of Marxism, the party considered state intervention as the primary characteristic of socialism and local autonomy as "spontaneous capitalism". Lastly, to the party, cooperative farming was not only a stimulus to agricultural production but also a strategic tool to transfer agricultural surplus for industrialization. Excessive extraction of economic rent from collective farms could make both group and individual incentive systems infeasible. In sum, all these policy tendencies would thwart the consolidation and development of Chinese collective farms.

2.4 Policy Fluctuations During 1958-1977

2.4.1 Disaster During the Great Leap Forward

The Chinese communists did not draw on the experience of FYP-1 properly. The superior performance led to an over-estimation of the contribution of institutional changes to agricultural production and the peasants' socialist enthusiasm on the part of Chinese leaders. As a consequence, the Great Leap Forward was promoted in March 1958. The major objective of the Great Leap Forward was to simultaneously develop agriculture and industry and to surpass Great Britain in principal industrial output within fifteen years. To support this ambitious goal, the rate of accumulation shot up to 33.9 percent in 1958 and 43.8 percent in 1959. The sharp expansion of industry put great pressure on agriculture. To break this bottleneck, a further transformation from agricultural producers' cooperatives to people's communes took place at the end of 1958.

With their "large size and high degree of public ownership",[3] people's communes provided a mechanism for mobilizing labour for large scale projects to support the "Great Leap Forward". Nevertheless, a sudden expansion of the size of collective farms caused serious management problems. The problem was exacerbated by the so-called "three winds": a "communist wind", characterized by extreme egalitarianism and the uncompensated transfer of resources from one collective to another, or from one level of ownership to another; a "wind of boasting", a tendency among cadres to exaggerate production achievements; and a "wind of blind direction", a tendency among the authorities to issue arbitrary and uninformed orders to collective farms.[4] The tremendous waste and losses caused by these "three winds", together with the occurrence of serious natural disasters in 1959 and the deterioration of the Sino-Soviet relationship, contributed to the nation's agricultural crisis of 1959-61. During this period, total output of grain fell sharply by 15 percent per year; and there were some 15 million famine related deaths. The economy was pushed toward the brink of total collapse.

2.4.2 Economic Recovery during 1962-66

The agricultural crisis led to the readjustment of Chinese development strategy. The Chinese leadership realized that industrial development could not proceed without agricultural

[3] Each commune averaged 4,600 families, in contrast to the average size of 158 peasant households for a producers' cooperative.

[4] The phrase of "three winds" was utilized in Chinese official media as a criticism of incorrect past policies after the food crisis.

development, and that institutional reorganization alone could not solve China's agricultural problems. The CCP leaders adopted a new development policy, different from past ones in several respects.

First, there was a subtle shift of emphasis toward agriculture. The targets for steel and other heavy industrial production were lowered, and the rate of capital accumulation was curtailed to 22.7 percent. This allowed the state to take steps to reduce the burden of agriculture. In the period of 1963-65, the prices paid to peasants for agricultural products increased by 20-25 percent and the agricultural tax rate was reduced from 15 to 10 percent . A larger share of state investment was allocated to agriculture than before and more modern inputs were provided for agriculture (Lardy, 1983).

Second, within the agricultural sector, the focus was to be on grain. Another new guideline, "taking grain as the key link", has dominated Chinese agricultural development since 1962.

Third, the commune movement was reappraised. The size of the commune was drastically curtailed.[5] The three -level collective ownership by the commune, the production brigade, and the production team was implemented. [6] The production team, with an average of twenty to thirty peasant households was the basic unit of collective ownership and accounting in a people's commune. Moreover, various forms of the work point system, initiated by producers' cooperatives, were reintroduced to implement the principle of distribution to each according to his work rather than need. In the meantime, 5-10 percent of the cultivated land was returned to individual households to allow them to increase their income. These measures rectified the practices of the "three winds" and restored team autonomy.

With the implementation of the above policies, the national economy showed a turn for the better. During the three years from 1963 to 1965, the average annual increase in the

[5] The new commune averaged 1622 households in 1963, only one- third of that in 1959.

[6] The commune owned only small enterprises, motor vehicles, mills and farm-tool repair shops. The brigade owned some farm implements and facilities. The team owned virtually all land, draft animals and tools.

gross value of industrial and agricultural output was 15.7 percent, with industry at 17.9 percent and agriculture at 11.1 percent (Xiu, 1982). The recovery in output was also suggested by rising total factor productivity in agriculture during 1962-66 (Tang, 1980).

2.4.3 The Turmoil of the Cultural Revolution

The recovery of agricultural production, however, fuelled the party's aspiration for rapid industrialization. High targets for the development of heavy industry were set again and the rate of accumulation began to grow. By 1966 it reached 30.6 percent and never dropped below 30 percent during the period of 1966-1976. The high target of industrialization once again put great pressure on agriculture.

The burden of agriculture was aggravated by an excessive growth in the population. In the early 1950s some economists did advocate that the rate of population growth should be brought under control. Unfortunately, these proposals were criticized and suppressed, because the CCP leaders believed that overpopulation is a problem of capitalism that does not exist in a socialist country. In consequence, from 1949 to 1975 the annual natural growth rate of population exceeded 2.3 percent. The rapid growth in the population not only put great pressure on the food supply, but also placed considerable strains on the party's employment policies.

To support industrialization and meet the demands of the growing population for food, the Chinese central government adopted policies which aimed at maximizing agricultural output and procurement through the growth in the use of modern inputs and a large-scale mobilization of peasants on rural capital construction. Among agricultural products, there was a lopsided emphasis on grain. Local self-sufficiency in cereals became an important target for production units at all levels. To ensure full employment in the urban sector, the government implemented a strict system of resident registration management to control the flow of the labour force from the rural areas into the cities.

The Cultural Revolution started in 1966. The objective of the Cultural Revolution was to fight against the so-called "bureaucrat-capitalist class within the party and to prevent restoration of capitalism. The leftists of the CCP were adamant about the "correctness" of the commune movement and the Great Leap, and tried to re-introduce the previous policies through political struggle.

The policies adopted during the Cultural Revolution caused great damage to Chinese collective farms in three general respects. First of all, the price structure for industrial and agricultural products was manipulated to the disadvantage of agriculture. During the period, an increasing amount of modern inputs had been injected into the rural sector. From 1965 to 1977, the utilization of chemical fertilizers was extended from 0.63 million tonnes to 8.84 million tonnes, the combined horsepower of agricultural machinery power grew from 14.94 million hp to 159.75 million hp, and electricity used in rural areas was extended from 3,710 million kwh to 25,310 million kwh. Due to the existence of the price scissors between agricultural and industrial products, the exchange value of industrial products relative to agricultural products was higher in China than on the international market. As the use of modern inputs had greatly increased, agricultural production costs shot up.

Moreover, during the period 1949 to 1975, the labour force engaged in agriculture leaped from less than 200 million people to 300 million people, an increase of 50 percent. Instead of transferring rural surplus labour to the urban sector, some 16 million urban educated youths were sent to the countryside. More and more peasants were tied to work on the limited land, simultaneously with the growth of the use of modern inputs. Furthermore, the marginal returns to new investment in agriculture were diminishing because high yields had already been achieved in the most productive areas and the state stressed cereal production even on relatively marginal land. All these factors led to substantial increases in agricultural production cost per unit of output.

Despite continuous increases in agricultural production cost, there were few upward adjustments in the procurement prices of agricultural products. From 1966 to 1977 the price index of all farm products purchased by the state rose only 4 percent, less than 0.4 percent per year (Lardy, 1983). The procurement price adjustments could not match the increases in the production cost. A survey of 2,163 production units in 23 provinces, for example, revealed that from 1965 to 1977 grain yields increased by 36 percent but production costs went up by 54 percent and the value of the labour day declined from 0.70 yuan to 0.56 yuan, a reduction of around 20 percent (Chen, 1982). As a result, collective farms produced high grain yields but still had low income. The decline in the value of a day's work and stagnation of peasants' incomes had a significant impact on incentives for peasants to work hard and on peasant support for the collective generally (Watson, 1983).

Another factor was that during the Cultural Revolution the decision-making power of the production teams was greatly restricted and the peasants were deprived of their democratic rights. The agricultural policy adopted during the period conflicted with the interest of collective farmers, for the policy required the peasants to fulfill the state production targets at high costs. For such a policy to be carried out the state had to concentrate the decision-making power and tighten its control over production teams. Thus, the critical policy instruments became the sown acreage, the yield target and the input requirements for individual crops that were issued downward to production teams. The production teams had to comply even though the production plans were not suited to local conditions and reduced the teams' income, for the peasants were required to "farm for the state and the revolution rather than for themselves". For instance, to achieve the goal of regional grain self-sufficiency, forests, grassland and lakes were blindly destroyed so as to open up wasted land for grain production, and arable land suitable for production of high yield cotton, soybeans and sugar cane was also used for grain. This kind of practice generated rather low returns to collective farms and created a disruption of the ecological balance as well.

The state intervention not only restricted team autonomy, but also deprived team members of their democratic rights in dealing with the teams' own affairs. This occurred because the state production and procurement plan had to be enforced through the local cadres and party organs. Consequently, whether a cadre could continue being in power depended upon his performance in motivating the production team to fulfill state output and market targets; the democratic right of the members in the choice of the leadership of the team was simply neglected. Thus, too much power was given to local cadres, and the problem of corruption among commune and team leaders was aggravated. The freedom of collective farmers was further restricted by such policies as eliminating private plots, banning household side-line production and closing rural markets. Under this policy regime, collective farmers had few alternatives except to produce cheap agricultural products for the state.

The Cultural Revolution era ideology directly disrupted the internal management system of Chinese collective farms. The Cultural Revolution motivated the desire to set up higher forms of public ownership. In the early 1970s, nearly half of the communes in Jiangxi province and one quarter of the communes in Zhejiang province transferred the basic accounting units from production teams to production brigades (Wu, 1985). The expropriation of grain, money and manpower from the production teams and unpredictable institutional changes shook the collective ownership of these teams and encouraged short-run behavior. Many production teams deliberately adopted the strategy of " partitioning all and eating all" in their income distribution, and did not leave any income for collective accumulation (Domes, 1980). Under the anti-incentivist ideology, egalitarianism pervaded the income distribution system of the collective farms. [7]

In sum, with numerous external economic and political disturbances for more than a decade, collective farming in China had been seriously obstructed, and the peasants' inclination towards cooperativism had been undermined.

2.5 Evolution of the Internal Incentive System

2.5.1 Formulation of Group Material Incentive Systems

Ever since agricultural producers' cooperatives were established, Chinese peasants had searched for some simple and practicable systems of internal management. At the very beginning of collectivization, producers' cooperatives were small, consisting of 10-15 households; most of these coops adopted a time rate work point system. The team heads gave the peasants job assignments on the spot every morning, and provided close labour

[7] I will discuss this issue in detail in the next section.

supervision. Each peasant gained work points for his/her working days.

The transition from elementary producers' cooperatives to advanced cooperatives brought about a rapid expansion in the farmland and labour force available to the collectives. In this situation, the method of assigning jobs on the spot often led to idleness and chaos. The simple time rate work point system was then replaced by the group bonus and penalty systems. The most popular reward scheme adopted by advanced producer-coops was called "three contracts, one reward" i.e. contract describing jobs, output quotas and production costs, and rewarding overfulfillment of output quotas while penalizing unsatisfactory task fulfillment. Under the reward system, regular work teams were organized to take charge of a fixed section of land and to perform a perennial task. The cooperatives set the quantitative and qualitative targets for a contracted job and output, and also specified the production expense necessary for reaching fixed quotas in production tasks contracted to work teams. The work team was rewarded with work points in accordance with its task fulfillment. The work points were distributed to individual workers based on their contribution to the team task fulfillment (Wang et al, 1985).

Some coops also set individual task rates alongside the group material incentive system. In these coops, large-scale tasks, such as carrying manure to the fields, ploughing and raking the fields, sowing, harvesting, threshing, small-scale capital construction, and fighting natural calamities were organized by the production team, while some divisible tasks such as weeding and field management were assigned to individual cooperative members. Some cooperatives conducting diversified economic activities contracted the cultivation of cash crops, the raising of animals and poultry and some aquatic products, the management of orchards, vegetable gardening, etc. to individual members or households. These individual workers and households received work points according to the amount of output they handed over to the teams. With these flexible incentive schemes, producers' cooperatives were able to provide adequate incentives for individual workers and to take advantage of joint farming as well.

In Sepember 1957, the party central committee decided that agricultural producers' cooperatives should universally adopt the system of "three contracts and one reward". The implementation of the group incentive system successfully overcame idleness and waste, improved the quality of farm work, raised labour efficiency, and increased production.

Theoretically speaking, with imposition of proper group bonuses and penalties, optimal incentives can be obtained through workers self-discipline (Holmstrom, 1982 and MacLeod, 1987). Yet it should be noted that a number of factors worked to ensure the efficacy of the group incentive system during the period of 1953-1957. First of all, the majority of collective farmers were the poor peasants who used to be hired labourers or land tenants. By voluntarily joining collective farms, these peasants became the masters of production and started working for their own interests. This aroused the peasants' initiative. Secondly, at the early stage of collectivization the central planning system had just been built up to replace the market system, and the state influenced collective farms' decision-making mainly through the price mechanism and taxation rather than direct intervention. Thus, the collective units exercised a high degree of control over their own economic processes. Third, the agricultural policy adopted during the period was more developmental and less extractive; therefore, collective farms were able to increase their income by improving production. Fourth, the Chinese government paid attention to inspiring the peasants' collective consciousness through the appropriate propaganda. For instance, one of the popular slogans associated with collectivization was "only if I serve for other people, would others serve for me." This kind of slogan was acceptable and instructive in directing Chinese peasants to work for the collective interests. In consequence, the Chinese producers' cooperatives ran reasonably well in the 1950s.

2.5.2 Adoption of a wage system and emergence of household farming

The labour management system of Chinese collective farms was disrupted by the commune movement. To support the Great Leap, agricultural production was carried out in the form of "large formation warfare" and the labour forces of the communes were organized into military forms of organization, such as squads, platoons, companies, etc.. The commune members received fixed wages and ate in the commune kitchens. The egalitarian income distribution system of the communes resulted in free-rider problems. In the spring of 1959, the central government reaffirmed the system of "three contracts and one reward" in rural production brigades.

The poor harvest in 1959 combined with excessive state grain procurement initiated the devastating famine in the early 1960s. Since no feasible reward scheme could be implemented in the situation of famine, production teams in many regions instituted a system of contracting land to individual households as an emergency device to overcome famine and food shortage in 1961.

2.5.3 Restoration of the work point systems

In the spring of 1962, the central government issued the "Regulations for the work of the rural peoples' communes", known as the "Sixty Articles". The document made significant policy adjustments in the commune system. The production team with an average of twenty to thirty peasant households became the basic unit of production and accounting. The team was responsible for profit and loss and directly organised production and income distribution (Art. 20 and 21). The production team was required to practice the system of democratic management. Decisions about the team's production, income distribution and leadership were required to be made through discussion at a general meeting of the team members.

The Regulations also specified the general structure of the income distribution of production teams. Before income was distributed to member households, deductions were made from the total income to cover taxes, sales to the state grain procurement organization, seeding grains, the 1-2 percent of the grain for storage, the 2-3 percent used to provide welfare funds and the 3-5 percent used as public accumulation funds each year (Art. 36). The remainder, about 65-70 percent of the crop, was distributed to the households according to the number of "work points achieved".

The income to be distributed took the form of both cash and staple food stuffs produced. To guarantee basic food needs, the teams distributed grain on a per capita basis according to the numbers in a family. Grain and other staples distributed were offset against each household's work point earnings, and the balance was paid in cash. For those labour-poor households whose work points were insufficient to cover their basic grain distribution, the team allowed them to go into debt for the amount of the short-fall.

The work points were calculated based on work days and task fulfillment. The team fixed a standard for work points according to the technical level required by different types of labour, as well as the intensity and importance of job tasks. The basic work norms were discussed and approved at the general meeting of team members (Art. 30-35). The "regulations" emphasized that egalitarianism must be prevented.

In this spirit, five distinct types of material incentive shemes were implemented during the period of 1962-1966. First, the team conducted unified production and income distribution and assigned jobs with the specified quantitative and qualitative requirements to individual workers. Second, with unified production planning and income distribution, the team assigned the tasks to the work groups with "four fixed requirements", i.e. the team specified the production quota, the quality and the length of working time, and the group was to be rewarded a fixed amount of work points according to the fulfillment of the task. Third, fixed tasks on a section of land were contracted to groups or individuals, depending on the nature of the tasks, and rewards were determined through appraisal of the quality of the job. Fourth, a small group of individual households located in mountainous regions were assigned output quotas. Fifth, farm output quotas were contracted for individual households. Among these forms, the first three were the most popular ones (Wu, 1985).

Reducing the basic accounting unit to the team level and implementing the task rate work point systems improved internal incentives significantly. The improvement of the workers' incentives was suggested by the rapid recovery and growth of agricultural production during this period (Tang, 1980).

2.5.4 The Dazhai Work Point System

In 1964, a movement of "learning from Dazhai" was initiated in the rural areas. Dazhai, a small production brigade in a very backward area of Northwest China, was selected as the symbol of self-reliance and hard struggle to develop production. Dazhai adopted a time rate work point system associated with imposing peer group pressure. Under the Dazhai work point system, the team leader recorded the number of days an individual worked. At a monthly meeting, a "standard" person was selected according to criteria such as political thought, technical skill, labour intensity, and labour attitude, with whom the team members were compared in assigning work points through self-assessment and public discussion. The merit of the Dazhai work point system was simplification of the awarding of work points through promotion of collective consciousness and self-discipline.

The Dazhai system did not necessarily reject material incentives; it could function well in teams with strong leadership and workers' commitment (Perkins and Yusuf, 1984). In a case study of China's collective farms, Unger (1985) notes that when collective production contributed to a significant improvement in peasants' living standards, the stronger team members had not particularly minded that their less productive teammates were rewarded with a disproportionate share of the team income. In this case, the Dazhai work point system had operated reasonably well. Unfortunately, the Dazhai work point system was introduced into Chinese collective farms during the Cultural Revolution.

2.5.5 Political Struggle, Rural Poverty and Egalitarianism

The Cultural Revolution sabotaged refinement of the internal incentive system of Chinese collective farms. During the period, the material incentive schemes were criticised as "putting working points in command". The principle of "distribution according to work" was attacked as "bourgeois rights". As a result, the various forms of piece rate and task rate work point schemes were replaced by the Dazhai work point system.

The potential merit of the Dazhai method was, however, outweighed by the Cultural Revolution era ideology. Under the guideline of "taking class struggle as the key link", the class alignment in the rural areas went astray, and a series of man-made conflicts were created among cadres and peasants which had a disastrous effect on the solidarity of the peasants. Moreover, having experienced numerous political movements, the cadres of most teams were afraid of arousing disputes and spoiling relations among team members. Furthermore, when implementing the Dazhai method, political reliability rather than work contribution was introduced as the most important criterion for the allocation of work points. Lastly, the Dazhai method was used as a tool to fight selfishness. However, this incidentally destroyed the self-enforcing mechanism of an effective group incentive system, because free-riders in the team no longer feared retaliation. As a result, egalitarianism dominated the income distribution of most teams. A predominant practice of allocating work points was to assign a fixed number of work points to each worker for each day worked according to his/her sex and perceived strength. The scale of daily work points usually ranged from 6 to 10.

Egalitarianism in Chinese collective farms was reinforced by the chronic rural poverty. Under the extractive agricultural policies, the value of a day's labour in Chinese collective farms fell by 33% between 1957 and 1977 (Zhang, 1980). As a result, many production teams possessed little economic rent, and were not able to distribute the team's cash and staple according to work points adequately. More precisely, many teams could distribute little cash above the basic grain ration and also had few welfare funds to finance overdrawn grain by the labour-poor families. As a result, the peasants' indebtedness diluted the income distributable for work points. According to Nolan (1983), income distributed according to work points did not easily cover 30 to 40% of the distributed collective income in many poor areas. [1] Under the work point system the raw number of work points gained by each household was quite different, for each household possessed a different capacity for labour supply. Yet the income

[1]The problem was noticed by the CCP leaders. In the document of the Central Committee of the CCP issued at the end of 1971, it was stated, "At present, some communes, brigades and teams have quite a few overdrawing households, rendering normal distribution impossible. They have even used up the accumulation of the collective, partitioned the national loans, thus seriously battering the aggressiveness of the commune member's collective production, and affecting the consolidation and development of the collective economy." (Domes, 1980)

distributed to each household did not differ much. The rural poverty induced excessive egalitarianism in the income distribution of collective farms, which dampened peasants' work incentives.

2.6 Concluding Remarks

Despite traversing a tortuous road, Chinese collective farms made a remarkable contribution to China's economic development. The collective farms provided a tremendous amount of financial resources for industrialization. It has been estimated that from 1949 to 1978, the price scissors between industrial and agricultural products had extracted 600 billion yuan for the state from the peasants, or 45 percent of their cash income during this period (Jiang and Lou, 1987). Thus, with little foreign financial support, China's industrial fixed assets had increased 26-fold from 1952 to 1979. The share of the industrial sector to gross national product increased by more than 30 percentage points from 23.1 in 1952 to 53.6 in 1978. The corresponding share of agriculture declined by nearly 25 percentage points from 57.7 to 32.8 during the same period.[9] The development of modern industry made it possible to equip agriculture with modern technologies.

The collective farms also provided huge amounts of labour for large-scale capital construction aimed at water conservancy. With the state's financial assistance, new dikes and embankments as long as 164,000 kilometers, four times the circumference of the earth, had been finished. More than 86,000 reservoirs with a total capacity of 400 billion cubic meters had been built (Xue, 1982). The total irrigated area of the country increased from 20 million hectares in 1952 to 44.96 million hectares in 1978. As a result, Chinese agriculture became much less vulnerable to cyclical weather anomalies (Kueh, 1986). The remarkable improvement in land infrastructure removed the major obstacle to accelerating agricultural output growth through modern technologies.

[9]Kueh (1989) points out,"This sharp rise in industry's share in China's national income is a rare historical phenomenon. For example, during the first four or five decades of their drive to modern industrialization, the industrial share rose by only 11 percent in Britain(1801-41); and 22 percent in Japan (1878/82-1923/27)."

During the period 1962-80, China's cereal production grew at a respectable 4.5 percent per year in contrast to the annual growth rate of world total cereal production of 3.1 percent. There was a significant improvement in productivity per unit area. In terms of sown acreage, grain yields rose from 1.03 tons per hectare in 1949 to 2.78 tons per hectare in 1979, and cotton yields from 158 to 488 kg per hectare. In 1980, China produced 18.4 percent of the world cereal output with only 7 percent of the total arable land on the Earth. China's cereal production was 120 percent of the cereal output produced by the U.S.A, 178 percent of that by the Soviet Union and 224 percent of that by India, whereas China's arable land was only 61 percent of that of the U.S.A, 49.6 percent of the Soviet Union and 67.8 percent of India (Ma, 1989).

The Chinese collective farms also provided an institutional vehicle through which great achievements in improving rural health and education were made. According to Perkins and Yusuf (1984), Chinese peasants had a life expectancy at birth of thirty to thirty-five years in the 1930s. Infant mortality for the rural populace was around 200 per 1000. In 1979, life expectancy was nearly sixty-nine years, having increased by an average of 1.5 years per calender year since 1949. This rate was three times the standard for developing countries. The infant mortality rate dropped to around 50 per 1000 for the nation as a whole. The literacy of rural young adults increased from 20 percent in 1949 to 70 percent in 1979 (Perkins and Yusuf, 1984). In 1979, there were 250,000 part-time peasant technicians and 17,622 agricultural technology diffusion stations in the rural areas, in contrast to zero and 10 respectively in 1950. In consequence, a great number of technological innovations had been made in the rural areas since 1953 (Xue, 1982).

The collective farms also promoted rural industrialization. Up to 1979, there were 767,098 commune- and brigade-run enterprises in industry, 82,142 in transportation and 49,651 in construction, employing 22.5 million workers. This unprecedented progress made under the commune system in the development of industry, rural infrastructure and human capital in fact provided the backbone of accelerated growth during the post-1978 reform era in China.

However, the heavy investment in industry and rural infrastructure was not accomplished without the sacrifice of peasants' current consumption. The growth of the income of peasants had been quite slow. It was estimated that the total per capita income of peasants was 73 yuan in 1956 and 113 yuan in 1977 (Lardy, 1983), and the income distributed from collective sources of a peasant in nearly one-fourth of production teams was below the 50-yuan poverty line (Document No. 1, 1979), and as much as one-eighth of the rural population could not produce enough grain to feed themselves (Watson, 1983). The stagnation of peasants' income had a significant impact on incentives for peasants to work hard and on peasants' support for the collective farming system.

Judging by its performance, cooperative farming in China as a special approach to maximize the speed of industrialization, and minimize the human cost that many developed countries have paid for their industrialization, can hardly be accused of a failure, given China's harsh economic realities. Nevertheless, there is no doubt that Chinese collective farms did not operate at full efficiency. In many circumstances the system was operating quite inefficiently.

However, the efficiency and incentive problems facing the Chinese collective farming units were complicated issues. Perkins (1980, p148) argues that the low return to the investment on China's agriculture can be explained in part by China's factor endowment. He speculates: " In the absence of a severe shortage of land ....., current high rates of investment in Chinese agriculture would have achieved much more dramatic results." Moreover, no economic organization is born with success. It has taken capitalist institutions more than two centuries to evolve from their classic form to modern corporations. One should not be surprised that Chinese collective farms went through some costly periods of seeking proper operational forms.

Nevertheless, it is undeniable that there were many man-made policy mistakes which prevented Chinese rural teams from working on their frontiers of production efficiency. As we have reviewed, the performance of collective farms was significantly influenced by variations in the current policy regime. During the period of FYP-1 and the early 1960s when the speed of industrialization was moderate, the agricultural policy was less extractive, and workers' material incentives were encouraged, Chinese collective farms did fairly well in development of production and improvements in peasants' living standard. But during the Great Leap Forward and the Cultural Revolution, the Chinese government adopted policies that promoted urban and industrial development at the expense of the collective farmers, and encouraged egalitarianism. With such an external environment, Chinese collective farms could not be operated efficiently to meet the demands of the state, while simultaneously raising peasants' incomes significantly, except a minority of communes which had locational advantages in production and marketing. The lack of significant progress on peasants' living standard undermined the peasants' incentives to work hard and their support for the commune system.

Thus, it is safe to conclude that Chinese collective farms were not inherently inefficient rural institutions; it was the adoption of dysfunctional policy regimes that undermined the consolidation and development of Chinese collective farms.

Chapter 3

Literature Review

3.1 Internal incentives under perfect information

The origin of the theory of collective farms is intertwined with that of the theory of the labour-managed firm. In the self-management literature, a labour-managed firm is commonly defined as a productive enterprise in which the workforce of the firm exercises the ultimate decision-making rights in a politically egalitarian and democratic manner, and workers share the residual claim (Bonin and Putterman, 1987). Economists have studied the nature of labour managed firms from various angles, for instance, the firm's behavior in labour [10] and capital markets.[11] Among these broad issues, the internal incentives in a labour-managed firm have attracted growing attention, because it is well recognized that the enthusiasm and devotion of the participants to a firm is as important a determinant of the firm's productivity as resource allocation efficiency in team production.

Sen (1966) investigates the internal incentives in collective farms. In his study, the optimal choice of the individual peasant's effort is assumed to be influenced by an internal distribution scheme that consists of a proportion of the collective income distributed according to "need" and a proportion according to "work". Sen demonstrates that with no income distributed according to "need", collective farmers tend to perform an excessive amount of

[10]Ward (1958), Domar (1966) and Vanek (1970) discuss the situation in which the team chooses its short-run labour input to maximize net revenue per worker. They note that because a new employee with lower marginal productivity tends to reduce average net revenue for the existing members under such a maximand, an increase in demand for output leads the worker team to reduce its output and employment. The team's perverse short-run behavior can be corrected in the long run through the formation of new teams in a growing industry.

[11] Pejovich (1969), Furubotn and Pejovich (1973) and Jensen and Meckling (1979) contend that a producers' cooperative cannot operate efficiently because of its defective investment incentives. According to this argument, the investment incentive problem is located in the limited horizon of the team member with the enterprise and non-transferability of the collective asset shares. Sertel (1982) and Dow (1986) argue that the inadequate investment incentives can be remedied through the use of marketable memberships. In a competitive market environment, "capital can be collectively owned by a labour-managed firm without undermining investment incentives." (Dow, 1986).

labour, because the farmer earns an added claim on the common product even if his marginal contribution increases that product very little. On the other hand, there is too little work done in a system of distribution according to needs due to free-rider problems. Sen concludes that optimality requires a mixed system of distribution according to work and needs.

Following Sen, Bradley (1971, 1973), Cameron (1973a,b) Markusen (1975, 1976), Bonin (1977), Chinn (1979), Cremer (1982), Israelsen (1980), Putterman (1980, 1981) and Putterman and DiGiorgio (1985) study the various issues of the incentive system of collective farms. The common conclusion of these studies appears to be that the organizational design of a producers' cooperative implies compatibility between productive efficiency and income equality; therefore, it is not necessarily inherently inefficient.

In all these models, work is assumed to be costlessly and perfectly measurable. This assumption, however, is not plausible because identification of individual productivity in team production is rather difficult. The difficulty of monitoring labour input requires a departure from the perfect information context of the labour-managment literature in order to understand the intricacies of work incentives in a team setting.

3.2 Internal Incentives under Imperfect Information

3.2.1 Alchian-Demsetz on the Necessity of Central Monitoring

It is now widely recognized that imperfect metering of labour input results in internal incentive problems under conditions of team production. Whenever the rewards fail to accurately reflect individual contributions to the joint output, an individual motive exists for shirking. Alchian and Demsetz (1972) contend that free-riding on the part of agents can be avoided by bringing in a specialized monitor to check the work performance of agents. However, the central monitor also has a motive to shirk, for monitoring effort cannot be measured at zero cost. According to Alchian and Demsetz, moral hazard on the part of the monitor can be eliminated by making him a residual claimant. On this account, Alchian and Demsetz argue that a classical capitalistic firm will have an advantage over a worker team, because a worker team lacks an exclusive residual claimant, and hence no member has an adequate incentive to monitor the contribution of others.

One flaw in Alchian and Demsetz's argument is that labour monitoring by a central agent is only one way to deal with the moral hazard problem of team production. In reality, many types of monitoring system such as central monitoring, mutual monitoring and self-supervision are implemented in productive enterprises. A central monitoring system is a control system in which some agent specializes in monitoring. This agent can be an exclusive residual claimant, a hired subordinate supervisor, or a partial residual claimant. A mutual monitoring system is a monitoring system where monitors are production participants. A mutual monitoring system can rely on formal horizontal control of labour, or informal peer-group pressures. Self-supervision describes a situation in which individual workers' efforts are self-enforced. The desirability and feasibility of a particular monitoring system depends upon the information structure and organizational design of the firm.

To see the nature of the information structure of a production team, we have to distinguish the concepts of verifiability and observability, which are sometimes used interchangeably in the labour supervision literature. A worker's performance is said to be verifiable if by investing resources in the monitoring of actions, the team can measure effort with enough accuracy that it can be explicitly included in a binding contract, so that a third party (the contract enforcer) can determine whether the specified level of effort has been supplied, and can penalize the worker if not. Effort is observable if it can be determined simply by watching a worker's actions. Effort is verifiable only if it is observable, while the converse is not true. A contracting problem arises from the fact that labour performance cannot be verified at zero cost, but formal labour contracts cannot be based merely on what people have observed, that is, they require verifiability of effort, not just observability.

Nevertheless, it has been contended in the implicit contracts literature that both verified information and observed information are valuable for contract enforcement (Bull, 1987). Verifiability is pertinent for the implementation of explicit contracts, while observability is crucial to the enforcement of implicit contracts. So the two types of information are both potentially useful in improving the productive efficiency of a firm. Generally speaking, obtaining verified information is the major purpose of formal labour supervision, while intra-firm observed information introduces the possibility of informal monitoring among the participants in the team.

A productive enterprise usually has two types of human input: a central manager and subordinate agents. To ensure optimal work incentives both the manager and workers should be brought under control. However, the monitoring mechanisms for the manager and subordinate agents are not the same because they have different information characteristics.

3.2.2 Control of the Central Manager

The primary task of a central manager is to coordinate team production; a manager's performance involves considerable initiative. As a result, a manager's effort cannot be directly supervised, and thus cannot simply be purchased on a payment for service basis. As Alchian and Demsetz suggest, in a capitalist firm, the central manager's effort is self-enforced through his position as a sole residual claimant. Yet the exclusive residual claimancy of an owner-manager is not the only way to provide incentives for the central manager. [12]

Although the manager's performance cannot be directly supervised, his work attitude, initiative, fairness and so on can be observed by workers. Through annual team production, workers will have a fairly clear idea about how much effort the manager has devoted to the team. This type of intra-firm information is wasted in a capitalist firm because there is little that workers can do to punish an unsatisfactory manager even if the workers are informed (except perhaps going on strike if the workforce is unionized). However, the observed information will be fully utilized by the workers in a labour-managed team, because the workforce has the ultimate authority over the decisions of the firm, including the right to choose the team manager.

The special information characteristics of the central manager provide some simple and thrifty solutions for the moral hazard problems on the part of the central manager in a

[12] Many people disagree with the Alchian and Demsetz argument about residual claims and effort of the central manager (see Putterman (1984) and Dow (1987)).

worker team. For example, the team can pay the manager a competitive wage and let his work contract renewal be subject to worker's annual assessment and the democratic vote of the working membership. A lazy manager will be fired by the team, and his performance will be written into his job reference. In this way, managers who are concerned about their reputations may be deterred from shirking by the prospect of lost future earnings. In particular, suppose that the wage of a manager is higher than a worker's average shared income in the economy. Job loss can be made directly costly even without reputation effects. As Shapiro and Stiglitz (1984) and Bowles (1985) demonstrate, involuntary unemployment can be used to discipline workers, and the same approach can be applied to managers as well. As long as the sum of the discounted income loss and the non-monetary losses in self-accomplishment, self-esteem, privilege of avoidance of unpleasant manual labour etc. for a fired manager are large enough to wipe out the one period gain of shirking in his managerial duty, the manager will align his individual priorities to the team's objective.

The control of the central manager's performance in a democratic production team should not be more difficult than the use of takeovers or proxy fights by shareholders to minimize managerial shirking. In fact, a team can react to a manager's malfeasance more quickly than the shareholders, for the team's manager is subject to direct monitoring by workers, while shareholders have to detect managerial shirking by reading stock market prices, or through other indirect channels. The problem of controlling the central manager's performance in a production team is thus overstated by Alchian and Demsetz. In what follows, I will focus on the monitoring of workers in the rest of the thesis.

3.3 Control over workers

3.3.1 Information Characteristics and Monitoring Structure

Generally speaking, a worker's job is relatively concrete and specific compared with that of the central manager. So it is less costly to measure a worker's contribution than a manager's by such devices as working hours with expected intensity, pieces of job done with some qualitative requirements, task fulfillment and so forth. If metering of workers' performance is not prohibitively costly, close labour supervision offers one potential solution for agents' malfeasance in a situation of team production.

It has been argued that the efficient form of labour-supervision depends upon the technological characteristics of the production process (Putterman, 1984). For instance, it may be efficient for a firm to have some agents who specialize in monitoring when the workplace is relatively concentrated in space and jobs are standardized. In other cases, where production is conducted at widespread locations, or the input performance of workers cannot be accurately assessed by agents who do not directly engage in production, it would pay instead for a firm to have each worker devote some of his time to "monitoring" (Mirrlees, 1976).

A worker's effort is often observable to his co-workers. If the firm's size is not too large, it is not impossible for a worker to know how all the other workers have performed through joint production. In this case, informal mutual monitoring among workers may furnish another potential control mechanism over workers. Clearly, eliminating free-riding problems necessitates worker supervision, yet the best solution is not necessarily to have all monitoring carried out by a central residual claimant as Alchian and Demsetz suggest.

Despite the existence of many potential solutions for the internal incentive problems of team production, a firm's monitoring structure is shaped by the organizational form involved. It seems unlikely that an effective mutual monitoring system could be established in a capitalist firm. In such a firm, the principal holds the residual claim and the workers earn fixed wages; any additional effort extracted from workers will become a part of the residual of the principal. A formal mutual monitoring system is not feasible under such an income distribution structure. The conflict of interest between the principal and agents introduces a potential for negative collusion among workers, which might be aggravated if workers were responsible for checking each other's performance (Cable and FitzRoy, 1980a). Conceivably, if a formal mutual monitoring were implemented, the principal might lose control over the workforce, unless he also monitored the worker-monitors himself.

The problem of extracting monitoring effort from workers here is not that the rewards to monitoring are too diffusely shared, but that the income distribution structure of a capitalist firm provides no incentive for workers to monitor each other at all. As a result, the principal must engage in monitoring. If the size of the team is large, it would be desirable for the principal to hire some subordinate supervisors to monitor productive workers rather than implement a mutual monitoring system, even in the case where mutual monitoring appears to be technically superior. Separation of monitoring duty from productive workers and paying higher wages to subordinate labour supervisors are one way to prevent negative collusion among workers and to control subordinate supervisors in a capitalist firm. [13]

Informal mutual monitoring among workers seems also infeasible for a capitalistic firm. The lack of direct interest interdependence among workers limits the scope for peer-group pressure over workers. Also, the conflict of interest between the principal and agents discourages truthful revelation of workers' observed information.

The monitoring structure of a worker team is rather flexible. Due to their sharing of residual claimancy, workers directly bear the cost imposed by free-riders in the team and share the benefit of improved productive efficiency. The interest interdependence among workers may motivate workers to monitor their co-workers. The individual incentive for monitoring might be small in a static situation because the benefit of monitoring activity tends to be spread among the team members. However, the workers are unlikely to engage in negative collusion against their own interest.

The interest interdependence among workers also provides a fundamental material basis for efficient utilization of intrafirm observable information flows in a worker team. This type of information has limited scope for improving productive efficiency in a one-shot static situation because the team cannot formally penalize a shirker without showing verifiable evidence, while team members are unable to react informally to each other's actions in a one-shot game. But most production teams involve long-term relationships which allow

[13] Calvo and Wellisz (1978) show that under a hierarchical control system, it is incentive incompatible to pay supervisors and supervisees equal wages despite the fact that no more skill is required of the former than of the later.

workers to punish free-riders through the use of their observed but non-verified information.

A production team can therefore more fully exploit its informational resources by implementing a monitoring system which consists of two parts: formal labour supervision (either central or mutual or a mix) and informal dynamic mutual monitoring. The existence of mutual monitoring components is the primary characteristic of the control system of a worker team (Ben-Ner, 1988b).

In sum, we see that Alchian and Demsetz treat the special attributes of the monitoring system in a capitalist firm as a common feature of all team production, and then compare capitalist with labour-managed firms using this biased assumption. Needless to say, the conclusions derived may be misleading with respect to the true monitoring capacities of the labour-managed firm.

3.3.2 Static Mutual Monitoring

Compared with the centralized monitoring system that is commonly implemented in capitalist firms, a mutual monitoring system has numerous advantages. First of all, a mutual monitoring system may eliminate the information asymmetry between the principal and agents that exists under a central monitoring system. In team production, there are many activities that allow workers to work and observe one another's performance simultaneously. In these situations, a specialized monitor may not be able to observe labour input as accurately as productive workers themselves. [14] The difficulty of monitoring labour by a third party can be aggravated by the spatial and sequential charateristics of the production process. In this case, a specialized labour supervisor has to conduct his duty by travelling back and forth over widely separated parts of the field, and hence the chance of catching a shirker is low. Nonetheless, if the monitors are also productive participants, labour supervision can be automatically carried out wherever production takes place as long as workers are not

[14] Bradley and Gelb (1981) suggest that considerable scope for horizontal monitoring may exist regardless of technological constraints. They point out, "As a matter of fact, only 2 per cent of all employed Americans are estimated to work under assembly line-like conditions."

physically isolated from each other. The information asymmetry between the principal and agents will become more serious when the production process is subject to frequent random disturbances. In this case, a central monitor may not be able to accurately discount the random impact in assessing input performance, while by doing the same job and facing the same state of the environment every worker of the team has equal access to information on the performance of others. The information implications of a mutual monitoring system are particularly important to a farming team.

By reducing information asymmetry, a mutual monitoring system allows the team to reduce the mistakes that otherwise might be made in choosing payment schedules, deciding about worker promotion and dismissal etc. This type of mistake is counter-productive in labour supervision. Additionally, with a symmetric information structure and democratic decision making process, it would be relatively easy for the team to set up quantitative and qualitative job requirements and corresponding rewards which are mutually acceptable to the workers. Once workers believe that the rewards/penalties and job assignment procedures are fair, fewer resources will be lost in disputes and disagreement.

Furthermore, a mutual monitoring system may motivate workers as a group to internalize monitoring costs through self-discipline. By engaging in both production and monitoring, workers directly bear the cost of both labour effort and labour supervision, while they will share the benefits of improvement in productive efficiency. Workers therefore have incentives to find production techniques or methods of organization that reduce the net cost of monitoring. Conceivably, if the size of the team is not too large, with sufficient communication and coordination workers may be able to avoid the temptation to shirk through mutual self-discipline.

Bradley and Gelb (1981) formulate a mutual monitoring model. In this model, "mutual monitoring" is defined as "positive reinforcement by peers". Bradley and Gelb demonstrate that under peer pressure, a worker may provide greater effort than he would deliver if only motivated by his share of profits. A worker's mutual monitoring effort is induced by this perceived increment to his shared income, while the worker has to sacrifice his leisure by engaging in mutual monitoring. They assert that workers will participate in monitoring of fellow workers up to the point at which the marginal disutility of the monitoring effort equals the marginal utility of the increment to their individual profit shares.

Bradley and Gelb's study demonstrates both theoretically and empirically the existence of a mutual monitoring mechanism in a production team and its positive impact on the team's efficiency. In their study, mutual monitoring is treated as informal labour supervision to supplement rather than to replace a formal central control system. This approach to the incentive system of industrial producers' cooperatives is reasonable. Since the work processes are relatively concentrated in space, and the jobs are standardized in an industrial factory, a central monitoring system may be more efficient than a formal horizontal monitoring scheme.

A formal mutual monitoring model has not to our knowledge been formulated in the literature on collective farms, yet such a system has many desirable properties for a farming team. In the literature on labour supervision in producers' cooperatives, monitoring is often treated as an ordinary input; the utilization of monitoring inputs is regarded simply as a reduction of the team's net revenue available for distribution. What is the monitoring structure of a production team like? Who engages in monitoring? And who monitors the monitors (Alchian and Demsetz, 1972)? All these problems remain unsolved in the case of production cooperatives.

3.3.3 Choice of Monitoring Input and Payment Rule

Another important aspect of labour supervision that has not been dealt with adequately in the literature is the relationship between income distribution schemes and monitoring per se. A team cannot solve free rider problems simply by monitoring itself, for few people would care about monitoring if shirkers were not penalized. The penalty imposed on sub-standard work is an inseparable part of an effective monitoring system. However, the formulation of the rewards and penalties and the specification of monitoring have often been treated as separate issues in the literature on moral hazard in team production.

There are two parallel approaches to the moral hazard problem in firms. One focuses on the endogenous choice of payment schemes for a given information environment; while the other emphasizes the choice of the optimal degree of monitoring holding the incentive scheme fixed.[15] For example, Holmstrom (1982) shows that in the absence of observability of effort, free-rider problems can be removed without monitoring when there is a "third party" who is able to enforce group penalties or to finance bonuses. He argues that the third party must be a principal who owns the capital and receives the residual income of the firm. Eswaran and Kotwal (1984) note that Holmstrom ignores the fact that the imposition of group penalties by a principal creates a potential for moral hazard on the part of this third party, which could render the proposed equilibrium incentive-incompatible.

The opposite tack to Holmstrom's approach has been to treat methods of labour supervision as a choice variable. Such studies focus on the choice among alternative systems for monitoring labour effort, while dealing with the issue of income distribution in some ad hoc fashion. For instance, Bradley and Gelb (1981) introduce the rewards and penalties into the mutual monitoring system rather informally. They argue that mutual monitoring will create a self-disciplined environment in which the workers may not feel so disciplined as they might feel under a vertical control system. So mutual monitoring is assumed to influence the choice of effort level not by changing the monetary return to labour, but rather by altering the disutility of effort. The peer pressure approach to mutual monitoring is inadequate if one is concerned primarily with individual monetary incentives. If the information generated through monitoring did not have well-specified consequences for workers' income, it would be difficult to determine the effectiveness of a mutual monitoring system when workers were rather materialistic and did not care about "peer pressure" too much.

Lin (1988) implicitly introduces penalties into a model of the labour supervision system of Chinese production teams. According to Lin's model, under perfect supervision the work points a worker earns are equal to his effort e. Without supervision a worker is presumed to have worked at maximum intensity $e=1$ and to be assigned full work points. He

[15]Putterman and Skillman (1988) make this distinction.

assumes that a penalty equal to 1-e is imposed on a worker who is declared to have not worked at maximum intensity. Lin does not explain why a production team must impose penalties in this ad hoc way.

Economists have noticed the impact of the income distribution scheme on the efficacy of a team's labour supervision system. Bonin and Putterman (1987) and Putterman and Skillman (1988) formally demonstrate that the incentive effect of increased monitoring depends critically on the incentive scheme employed in a production team. In particular, they show that if all the collective income has been distributed according to need, rather than individual effort, monitoring has no effect on effort.

The authors of these two studies, however, do not deal simultaneously with the optimal choice of an income distribution scheme and the optimal choice of monitoring intensity. As the authors mention, an ad hoc approach to the payment scheme is assumed in their analysis as a simplification, to help focus on the relationship between monitoring and incentives. However, by holding the payment rule fixed, one can easily overlook the ways in which workers' incentives vary with the nature of the income distribution scheme. This omission has led to excessive emphasis on the role of monitoring per se in the investigation of labour supervision and incentives in a production team.

The interaction between the payment rule and monitoring intensity as determinants of labour effort has been modeled in the efficiency wage literature. Shapiro and Stiglitz (1984) and Bowles (1985) show that when labour effort cannot be observed costlessly and perfectly, involuntary unemployment is in fact a worker discipline device. By examining the firm's optimization program with respect to the wage package and monitoring intensity, they find that there exists a trade-off between the magnitude of the penalties imposed on a caught shirker and the degree of monitoring: a employer can curtail monitoring expenditures by choosing to pay high wages, while relying on the threat of involuntary unemployment to achieve a no-shirking equilibrium. Shapiro and Stiglitz and Bowles' studies provide insights into how to integrate the selection of a payment scheme with the choice of monitoring intensity in a team's optimization program.

3.3.4 Dynamic Mutual Monitoring

Formal labour supervision offers a static solution for free-rider problems in team production, whereas most production teams involve long-term relationships. It is well known that economic agents could behave quite differently in a one-shot game and in a long-term relationship. In a one-shot game, the concept of Nash equilibrium rules out strategies which take into account the reactions of other agents. A repeated game, however, allows agents to respond to one anothers' actions, so that each agent must consider the future reactions of his opponents in making his decision. The fear of retaliation may lead to outcomes that otherwise would not occur, and which may be Pareto-superior to Nash equilibria of the one-shot game (Fudenberg and Maskin, 1986).

In this spirit, Tyson (1979) proposes that the efficient level of effort can be obtained by a production team through infinite horizon trigger strategies. MacLeod (1984) formally models the dynamic incentive system of a worker team. MacLeod shows that one-period cheating gives the deviant a net utility gain due to high leisure and only slightly reduced income. However, once the defector is detected through his or her effect on output, all team members will revert to suboptimal effort levels in all periods following a deviation, and then the defector will be punished in the long run. With a sufficiently long horizon, and a low discount rate, it is never optimal to shirk. The first best level of team production can thus be attained despite the selfishness of team members, simply through monitoring total output and dynamic punishment strategies. MacLeod notes that because a long term relationship is a necessary condition for viable cooperative production, freedom of exit may make the existence of a worker team impossible, even if such an organization is socially desirable. In this regard, MacLeod recommends the imposition of exit barriers to create sufficient commitment to the cooperative on the part of workers.

Two objections to the dynamic incentive model of Tyson and MacLeod have been raised. First of all, the model assumes long-term relationships among a fixed group of agents. Although the firm may have an infinite life, each of its members participates for only finitely many periods. Workers who are about to retire will be tempted to renege on their effort contract, because they do not fear any punishment triggered by their deviation from the cooperative agreement that would happen in the next period (Conte, 1979). Secondly, the imposition of barriers to worker exit may be incompatible with the factor mobility required for efficient allocation of resources (Bonin and Putterman, 1987).

Jacques Cremer (1986) develops an overlapping generations model to remedy the first of these two objections. According to Cremer, while it is true that the oldest members of the firm will slack off due to their freedom from retaliation, the next oldest member of the firm has to behave himself properly, for he can still be threatened by the workers who will be living in the next period. These workers in turn will be threatened by the successors in the chain of members of the firm. Consider that a firm with an infinite life consists of N workers, and each period a new worker enters the firm while the oldest quits. Cremer shows that in this situation an equilibrium where N-1 workers choose cooperation and the oldest worker slacks off is attainable, even if the static equilibrium is inefficient. He also notes that in this ongoing firm, it is optimal to assign the young agents the hardest tasks, because they are least likely to cheat. Thus, the infinite horizon incentive model for a fixed group of agents is a plausible simplification of the incentive system in a more realistic production team where workers have finite lives.

To avoid the conflict between productive and allocative efficiency that might rise in the dynamic incentive system, special attention has to be paid to the notion of barriers to exit from the team. There are three kinds of exit barriers which need not be incompatible with efficiency in resource allocation. The first one is the natural barrier created by the factor endowments of an economy. Because of the existence of natural barriers, labour mobility varies across industries and nations. For instance, agriculture is essentially a declining sector; most labour on farms will be gradually transferred to non-farming sectors in the course of economic growth. The transfer of farm surplus labour is inevitable and desirable, so it should not be deterred on any account. This special feature of agriculture limits the scope of the dynamic incentive mechanism for a team mainly engaging in agricultural production when no labour exit barriers are imposed. Nevertheless, the dualistic economic structure of an underdeveloped economy such as China gives rise to various natural barriers that discourage labour migration from rural to urban areas, since the already overly-crowded cities provide very limited employment opportunities for rural migrants. The existence of these natural barriers helps provide a solution for the conflict between productive and allocative efficiency a farm team may encounter. The promotion of rural industrialization and economic diversification in rural producers' cooperatives will reduce labour turnover for the cooperatives, without deterring the transfer of rural surplus labour to the non-farming sector. It is likely that without the imposition of any additional barriers to rural migration, the most successful producers' cooperatives in rural areas will be those teams that diversify their economic activities rather than those mainly engaging in farming.

Another type of exit barrier which need not undermine allocative efficiency involves reductions in the value of a worker's reputation. A history of frequently quitting jobs serves as a signal about a worker's past performance . Since a worker's reputation influences future employment opportunities, he/she cannot afford to choose the option of shirking and quitting without hesitation.

The third type of natural exit barrier which is not necessarily in conflict with allocative efficiency is the firm's own internally generated barriers. The success of a team is obviously the fundamental material basis for workers' commitment to the team since the opportunity cost of leaving a good team for a worker is high. While workers' commitment to remain in the team is a necessary condition for viable cooperative production, the team's success reinforces this commitment. It is thus likely that unusually successful teams can achieve optimal effort levels through dynamic self-enforcing mechanisms, without the imposition of additional exit barriers.

Nevertheless, no firm is born successful; it takes time for a team to adjust to external market conditions and to establish behavioral rules among workers. As a result, in the early stage of a team, worker turnover could be quite high, other things being equal. Hence, without unusual commitment on the part of workers, the dynamic incentive mechanism is unlikely to completely replace formal labour supervision as a way of ensuring efficient effort levels. It is thus conceivable that at a certain phase of a team's life cycle, the team will implement a mix of formal and informal labour supervision schemes to ensure the proper incentives, without adopting any device of deterring labour mobility. The pure dynamic incentive model may not fully resolve the incentive problems facing a production team, but it may well be an important special feature of the incentive system in a successful team.

3.4 Labour Supervision in Team Farming

It has been argued that free riding pervades farming teams, for monitoring of labour effort in agricultural production is extremely difficult and very costly. Agriculture's sequential nature and spatial dimensions make it difficult to standardize job tasks, to control the quality of performance and to provide close supervision of farm workers at all stages of the production process. The high cost of supervising farm labour induced dramatic labour-saving technological change in western agriculture, which allows fairly large units of production to be operated as family farms (Bradley and Clark 1972, and Wittfogel 1971). Supervision problems have also resulted in labour-displacing investment in some dynamic areas in LDCs, despite low wages (Sen, 1981).

The special problems of supervising farm labour have been thought to be the ultimate economic factor behind the sensational demise of team farming after nearly a quarter century of practice in China. Lin (1988) formally models the monitoring and incentive problems facing Chinese collective farms. Lin's model assumes that individual workers make their effort decisions under the team's monitoring system, while the team management chooses the degree of supervision to maximize the average net income per worker. Lin shows that under the imperfect metering of labour effort, the incentives to work in a production team are positively correlated with the degree of monitoring in the production process. The optimal degree of monitoring in a team, nonetheless, is negatively related to the difficulty of monitoring. Since "it is thus costly to monitor each peasant's effort contribution in agricultural production", states Lin (1987), "the optimal degree of monitoring in a team mainly engaging in agricultural production must be very low." He asserts that the special problems of supervising farm labour are "the most important factor in explaining the failure of the work point system in Chinese agricultural production teams and in explaining the dominance of the household responsibility system over the production team system." (Lin, 1988)

Lin's study reinforces the view that team farming is not a viable production structure for agriculture. However, Lin overlooks many important aspects of labour supervision in team farming. Free riding is indeed a common potential problem in team production, and the nature of agriculture limits the ability of the team to solve its internal incentive problems through close labour supervision. However, metering of individual workers' effort contribution is only one way to deal with the moral hazard problems in team production. Team farming does not rule out the possibility that a team mobilizes workers' unsupervised initiatives by aligning workers' interests to the objectives of the team. If the workers were strongly motivated to work for the collectives, close supervision might not be required (Sen, 1967 and Riskin 1974).[16] What is crucial to an effective collective incentive mechanism is that the team must exercise control over its production process and output. Such a prerequisite could not be easily ensured in a centrally planned and underdeveloped economy.

Monitoring is an important aspect of a labour supervision system; yet it becomes pertinent to the incentives only if the information generated through monitoring can be translated into appropriate payment schemes. Such a necessary condition for a valid monitoring cost argument should not be taken for granted in the study of the internal incentives in Chinese collective farms, which had been responsible for the primary capital accumulation of the nation's industrialization and provision of "basic needs" for nearly 16 per cent of the world population, while having limited economic and technical options to generate enough economic surplus to achieve all goals.

Furthermore, it is undeniable that labour monitoring is more difficult in a large-scale farming team than in an industrial producers' cooperative. However, the difficulty of labour supervision caused by the nature of agriculture can be reduced by selecting an appropriate

[16] In a case study, Zweig (1985) notes that in some Chinese brigades where collectively distributed income accounted for a significantly large and growing proportion of peasant household income, the team members preferred time-rate schemes to the household responsibility system.

monitoring scheme. In other words, agriculture's sequential nature, spatial dimensions, and vulnerability to the random impact of nature make metering of labour effort rather costly under a vertical monitoring system; but these factors may not matter so much when a mutual monitoring scheme is implemented. It is unclear why collective farms must adhere to costly central monitoring rather than implementing a possibly feasible and better mutual monitoring scheme.

Due perhaps to his focus on the technological aspects of team farming, Lin does not pay sufficient attention to the special attributes of the incentive system of a production team. As a result, Lin's study is not able to go beyond the surface to tackle the fundamental factors behind the incentive problems of Chinese collective farms.

Lardy (1983) advances the view that there has been a long-term tendency in the Chinese government's pricing and marketing policies to constrain peasant income-earning opportunities and to inhibit the efficient use of resources in agriculture. These policies resulted in peasants' income stagnation and dampened their work incentives. Hence, the key obstacles to China's agricultural development lie not in the realm of rural institutions, but in government policies such as pricing, crop planning and marketing. Nolan (1983) asserts that the official advocacy of a high proportion of "basic ration" in total grain distribution in the past could have caused serious motivational problems in China's backward areas.

Putterman (1985) argues that despite the difficulty of labour supervision, democratic producers' cooperatives need not be inferior to family farms. The success of collective farming in Tanzania and China was inhibited by these nations' bureaucratic controls, extractive government policies and anti-incentive ideology. Putterman (1987) shows that the failure to choose an optimal mix of meeting basic needs and implementing a distribution scheme based on material incentives in China's collective teams was among many factors that caused the incentive problems of Chinese collective farms. Under the prevailing economic and ideological constraints, it was rather common in Chinese collective farms that the proportion of the collective income used to cover basic needs was too high, while the proportion of the income left to be distributed according to work turned out to be too low. This could result in serious moral hazard problems in these teams, which were in general located in poor areas. The low productivity of labour led to the use of a large share of distributable income for subsistence consumption, and left little scope for implementing material work incentives. In consequence, there may have existed a vicious cycle of poverty and moral hazard in these poor teams.

Parish (1985) suggests that given the difficulty of correcting market distortions and reorienting existing industrial and urban-biased development strategies, "to move toward quasi-family farming, then, may have been seen as a cheap alternative, providing a quick fix for many of the problems afflicting agriculture." Nevertheless, as the experiments with household farming in Poland show us, without proper attention to the macro-policy regime, family farming itself is unlikely to provide sustained growth in agriculture in the long run.

Due to the lack of a sound theoretical analysis of the incentive system of collective farms and its relation to the policy regime, these important notions have not attracted sufficient attention in the study of Chinese rural reform.

3.5 Empirical Work on the Incentive System

While the issues of labour supervision have been considered in the labour-management debate, there has been little rigorous empirical study on the subject. Some attempt has been made to examine incentive issues by estimating the productivity effects of participation, worker ownership and profit-sharing in free market economies. Among these studies are Jones and Backus (1977) and Jones (1982) for British producer cooperatives, Cable and Wilson (1989) for British engineering firms, Cable and FizRoy (1980a,b), Svejnar (1982) and FitzRoy and Kraft (1987) for German participatory firms, Conte and Svejnar (1981) for US participatory firms, Defourney, Estrin and Jones (1985) for French cooperatives, Jones and Svejnar (1985) for Italian producer cooperatives, Bradley and Gelb (1981) for Spanish Mondragon cooperatives and Estrin, Jones and Svejnar (1987) for producer cooperatives in various Western economies. The general approach adopted in these studies is to regress firms' output against such institutional variables as profit-sharing, worker ownership and worker participation as well as conventional production inputs, assuming that changes in working incentives will lead to variations in the intercept of the production function (interpreted here as the total factor productivity) of imperfectly measured factors such as labour. The broad findings of these studies are that profit-sharing and worker participation have positive effects on total factor productivity. This result appears to reject the pessimistic view that residual-sharing teams offer insufficient incentives for labour effort.

There have been a few empirical investigations of labour supervision and incentives in Chinese production teams. Lin (1988) presents such a study with a highly aggregated data set of agricultural production from 29 provinces in China over the years 1979-83. Because the data set does not contain any detailed information internal to the production teams, Lin is not able to conduct a rigorous test of his monitoring thesis. Instead, Lin approaches the issue of monitoring indirectly by estimating an induced institutional innovation model.

Lin postulates that the rate of adopting the household responsibility system (HRS) in an area is a function of the gains and costs of shifting to the new system in that area. Lin uses the average amount of machines and the number of draft animals per team in each province as proxies for the cost of the institutional change. According to Lin, machinery would make it more difficult to break up a production team, while draft animals would be suitable for individual household production. Arguing that the gain of the shift from the production team system to the HRS is the improvement in incentives since the difficulty of monitoring does not exist in the HRS, Lin use the average size of production teams and the ratio of the gross output value of crop cultivation to the gross value of animal husbandry in each province as proxies for the gains of the institutional change. Lin's claim is that the cost of monitoring is greater, the larger the production team, and that the more important crop cultivation was compared to animal husbandry, the more severe was the incentive problem. Lin argues that crop production was conducted by the team, while most animal husbandry was undertaken by individual households before the HRS was adopted. Hence, the team size and the ratio of the gross value of crop cultivation to the gross value of husbandry are expected to have a positive effect on the diffusion rate.

However, the relationship between the ratio of cropping to husbandry and the rate of adopting the HRS seems irrelevant to the monitoring issue. The ratios of cropping to animal husbandry across provinces are not an accurate measure of the relative weight of collective to individual household operation. As a matter of fact, in most areas such as Beijing, Shanghai, Tianjuan, Inner Mongolia, Xizhang, Qinghai and Xinjiang provinces where animal husbandry accounted for a large proportion of the gross value of agricultural production, animal husbandry was undertaken by collective units rather than by individual households before the rural reform. Hence, a positive correlation between the diffusion rate of the HRS and the ratio of cropping to husbandry can hardly be regarded as convincing evidence supporting Lin's monitoring thesis.

The use of team size as a proxy for monitoring cost is also not particularly convincing. One can always argue that the larger a production team, the harder it is to dismantle the team; therefore, the slower the rate of diffusion of the HRS. This dispute has to be settled empirically, yet Lin's empirical results are controversial. The result of an ordinary least squares estimation in Lin's study shows that team size is negatively correlated with the rate of diffusion of the HRS, but statistically insignificant. This outcome is obviously inconsistent with Lin's monitoring cost conjecture. Lin then argues that team size, together with the amount of machines and number of draft animals in a team, were endogenous variables. Decisions about those variables, he contends, were responsive for potential returns to scale. As a result, the model is re-estimated by a two stage least squares method (2SLS). The 2SLS result shows that the coefficient of team size has the predicted positive sign and is statistically significant.

Although the results finally support Lin's thesis, the treatment of team size as an endogenous variable raises some puzzles about Lin's model specification. If the observed team sizes were the teams' contemporary optimal choices, why did the teams want to change them by adopting the HRS? If Lin assumes that team size and other variables were the teams' choice when the option of shifting to the HRS was not feasible, i.e. those variables were predetermined before the rural reform, then the OLS estimates, which do not support Lin's monitoring argument, are consistent. Without an appropriate model of a team's decision-making with respect to resource allocation and institutional choice, these puzzles cannot be resolved. In sum, Lin does not provide convincing evidence for the widespread belief, which Lin himself holds, that the incentive problems of Chinese collective farming units were primarily caused by difficulties of labour monitoring.

Putterman (1990a) examines the incentive problems of Chinese production teams with a micro panel data set comprising 41 teams for the years 1974-79. In contrast to Lin's monitoring explanation of poor work incentives in the teams, Putterman argues that the lack of incentives may have resulted from the egalitarian nature of the income distribution practice in Chinese rural teams. According to Putterman, workers would work harder, the more of the team income they expected to be distributed to them as income, and the more of that income they expected to be distributed in the form of cash rather than grain, since grain distribution based on household demographics was largely egalitarian. The problem Putterman faces is how to measure work effort. By using the number of work points per worker as a proxy for work effort, Putterman estimates the effect of the incentive variables on the number of work points. Noting that the number of work points may be a poor proxy for work effort given a low level of monitoring, Putterman also calculates the effect of incentives on work quality by estimating the productivity effect of the incentive variables. The two equations are estimated by 2SLS. The results provide some mild support for Putterman's hypotheses.

John McMillan, John Whalley and Lijing Zhu (1989; hereafter MWZ) estimate the relative contribution of price increases and strengthened individual incentives due to the introduction of the HRS to the post-1978 growth of China's agricultural total productivity. With a standard growth-accounting technique, MWZ find that 78 percent of the measured productivity increase is attributable to the incentive effect of the new payment system and the remaining 22 percent to changes in price policy. Using the same method, MWZ also compute incentive indices, showing that the effort Chinese peasants in the communal system supplied was only 56 percent of what they did under the HRS because it is estimated that the peasants had expected to receive only about 30 percent of their marginal value product under the old payment system.

As MWZ mention, their procedure may exaggerate the incentive problems of Chinese rural teams and overstate the contribution of the HRS to the agricultural growth of China in the post-1978 period. Having been determined residually, the estimate of incentive effects might take all the credit for policy changes in such areas as crop diversification and opening of free markets, technological progress and favorable weather conditions (Kueh, 1986).

Despite this problem, the striking aspect of the MWZ study is that the results obtained by MWZ reject the widespread belief that there was no effective monitoring in Chinese production teams. According to Lin (1988), a team worker would expect to receive only 1/n of his marginal product if labour monitoring was negligible, and effort and rewards were entirely disconnected. Nevertheless, MWZ's estimated proportion of the marginal revenue a peasant expected to keep in the communal system is 30 percent, which appears to be too high to support the zero monitoring thesis, since the size of a typical Chinese team was not smaller than sixty.[17] The existence of effective labour monitoring in China's production teams is evident by the fact that peasants anticipated receiving a not insignificant fraction of their marginal product in return for additional effort.

Considering the fact that the high proportion of income distributed according to "basic needs" in China's rural teams could have undermined the efficiency of labour monitoring in linking rewards to individual effort, Putterman (1990b) notes that if the findings of MWZ are to be believed, much of the incentive failure in Chinese rural teams may not result from deficiencies in supervision. Putterman assesses the relative importance of monitoring problems and the egalitarian income distribution measures for the incentive problem in China's communal system by linking the MWZ method to a microanalytic model of team production. Applying MWZ's results and the parameter values obtained from other empirical studies, Putterman shows that the probability that a worker is observed perfectly in the teams fell in the range 0.28 to 0.96. If values at the high end of this range are relevant, the monitoring problems may have played only a trivial role in undermining work incentives.

[17]The average size of a production team across 29 provinces in China during the 1981-83 period was 63.8, according to *China Agricultural Statistic Yearbook, 1984*.

The Putterman study thus can be interpreted as a refutation of standard monitoring arguments such as those of Lin. However, there might be some possible deficiencies in Putterman's procedure. Putterman takes an ad hoc approach to the issue of payment rules and estimates the probability of monitoring as a residual. As a result, the study gives us little idea which factor explains the computed high value of the probability of monitoring: high observability of individual effort in team farming, or heavy investment in labour supervision. If it is the latter, the measured high value of the probability of monitoring does not necessarily contradict arguments that monitoring problems were central to the incentive problems in group farming. Moreover, being treated as a residual, the probability of monitoring in team farming may be overestimated. Unfortunately, we lack knowledge of the statistical properties of the estimates; therefore, it is difficult to assess the reliability of the results.

In sum, empirical work on incentives in team production has been subject to certain handicaps. There has been no rigorous theoretical model which deals simultaneously with a team's decision-making in the areas of income distribution and labour supervision. The crucial variables relating to incentive issues, such as effort and monitoring cost, are typically hard to measure. Data containing detailed information internal to teams are often not available. As a result, empirical investigations of incentive problems have so far been conducted in an indirect, partial and somewhat ad hoc fashion. The results are sometimes subject to contradictory interpretation. Greater research effort is needed to improve on this situation.

Chapter 4

A Static Mutual Monitoring Model

4.1 Introduction

To tackle the complicated issues of the incentives in Chinese collective farms, a model of mutual monitoring of work effort is constructed in two steps. We first focus on formal mutual monitoring and study the incentive issues in a static set-up. Then, we will extend the static mutual monitoring to a production team with an infinite time horizon to integrate formal labour supervision with informal mutual monitoring among workers. Thus, a simple model of static mutual monitoring is formulated in this chapter and estimated in chapter 5, and a dynamic mutual monitoring model will be developed in chapter 6 and examined empirically in chapter 7.

Chapter 4 is outlined as follows. The basic assumptions of the model are presented in section 2. The characteristics of individual workers' effort decisions in a noncooperative game are described in section 3. Section 4 discusses the constraints on choice of an optimal labour contract, existence and uniqueness of an optimal contract and the comparative static results. In the final section of the chapter, the results are summarized.

4.2 The Basic Assumptions

4.2.1 The Production Team

Let $N \equiv \{1 \, . \, . \, n\}$ be the set of workers in a production team, which is finite, and $n \geq 2$. All workers are identical. [11] The team implements a mutual monitoring system, in which

[11] The assumption of identical workers is justified by the fact that the teams have to standardize the contribution and rewards of potentially heterogenous workers, because the direct measurement of an individual worker's marginal contribution to the team is almost impossible. The Chinese remuneration system was designed using the guidance of Marx's labour theory of value, which treats time as a natural measure of labour contribution. The teams usually measured labour supply by the concept of "standard working hour", which is defined as the product of a natural unit of working time and an expected working intensity. In practice, the team defined an hour job that an adult male worker was able to handle as a standard

each worker engages in two tasks: productive labour, and monitoring the effort of other workers. [19] These activities are mutually exclusive: time allocated to monitoring comes at the expense of production. We normalize the length of the relevant time interval at unity. If the time devoted to production by a typical worker is $\alpha \in [0,1]$, then $1-\alpha$ is the fraction of the worker's total time devoted to monitoring activities. [20]

A worker's monitoring activities provide information about the effort of other workers engaged in production, as well as information about the diligence with which other workers are carrying out their own monitoring responsibilities. Because there is a positive probability that shirking in production will result in detection and punishment, monitoring induces team members to work, and is therefore indirectly productive. At the same time workers are deterred from shirking in their monitoring duties by the prospect that such shirking will be detected by some other worker who does not shirk in monitoring. The answer

[18](cont'd) working hour, and then discounted an hour job a woman or a teenage worker was able to manage as 0.8 and 0.6 of the standard working hour respectively. The team specified the total number of working days and assigned jobs associated with different intensity requirements to each type of worker. Daily work points of 10, 8, and 6 were then awarded to the workers in the corresponding categories. The required working intensity was enforced by imposing close labour supervision and penalties. The income structure under this type of remuneration was generally egalitarian. If no worker was accused of having done a substandard job, the observed income variation across workers reflected merely the differences in workers' age, sex, and perceived strength and skills.

[19] The work point system adopted by Chinese production teams was a mixed system of mutual monitoring and central monitoring: the team leaders organized monitoring activities such as job-checking and performance evaluation meetings, while the team members directly engaged in monitoring. To simplify the matter, I treat the team leader as one of n identical workers and assume away the special role of the leader in labour supervision. Admittedly, this assumption is not realistic. But it is consistent with the fact that there were no specialist monitors in Chinese production teams in the sense that the team leaders in China were required to participate in team production for nearly two thirds of the total number of working days (*ganbu-chan-jia-lao-dong*), while directing team production. More about the control of team managers was discussed in section 3.2 of chapter 3.

[20] The time spent on monitoring represents the team's cost of labour supervision in terms of foregone output. The expense for nonlabour monitoring inputs such as paper and pens for recording work points was usually negligible and therefore is ignored. The specification of worktime allocation to production and monitoring is made based on the fact that productive effort and formal labour supervision are mutually exclusive activities. When workers are arranged to check the task-fulfillment of their co-workers or to hold a performance evaluation meeting, they are unable to engage in direct production at the same time. Hence, monitoring activities take away a proportion of worktime from production.

to the question "who monitors the monitor?" (Alchian and Demsetz, 1972) is thus "another monitor."

In the simplest case, each worker monitors the effort of only one other worker. To formalize the organizational structure of the team, let $i \rightarrow j$ mean that i directly monitors the effort of j, with $i, j \in N$ and $i \neq j$. Let $i \twoheadrightarrow j$ mean that there is some sequence of distinct team members $i_1, i_2, \ldots i_k$ such that $i \rightarrow i_1 \rightarrow \ldots \rightarrow i_k \rightarrow j$. Then we assume

A1: (a) For each $i \in N$, there is some unique $j \in N$ such that

$i \rightarrow j$.

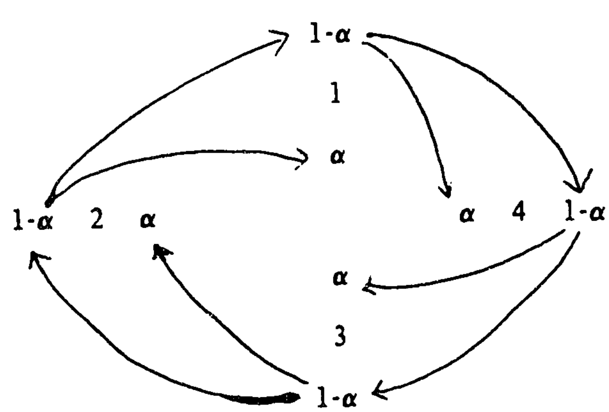
(b) For each $j \in N$, there is some unique $i \in N$ such that

$i \rightarrow j$.

(c) For every $i, j \in N$, $i \twoheadrightarrow j$.

Assumption (c) ensures that every worker is indirectly monitored by every other worker. This rules out situations where the team is divided into several independent pairs, for example, with each partner in a pair monitoring the other. The team members will be labelled so that $i+1 \rightarrow i$ for each $i \in N$, adopting the convention $n+1 \equiv 1$. This gives rise to a circular monitoring structure, as shown in Figure 1 for the case $n = 4$.

Figure 1.



In Figure 1, α and 1-α denote productive and monitoring activities, respectively. Worker i+1 monitors the performance of worker i in both production and monitoring. Through such mutual job evaluations, the team compiles a performance record for each of its members. [21]

4.2.2 The Monitoring Technology

Assume that a worker is punished by the team for shirking in production and/or monitoring only if the monitor involved can produce verifiable evidence that shirking has taken place. Let $w_i \in [0,1]$ and $x_i \in [0,1]$ be worker i's effort levels in production and monitoring, respectively. $w_i = x_i = 1$ stands for fully effective work, and $w_i = x_i = 0$ for complete idleness. We assume that the probability of verifiable evidence of shirking being obtained is a function of these effort levels and the fraction of time spent on monitoring.

In particular, the probability that worker i is punished for shirking in production will be written as

$$(1) \qquad P_{i1}(w_i,x_{i+1},\alpha) \equiv x_{i+1}(1-w_i)\Phi_1(\alpha)$$

The probability that worker i will be punished for shirking in the monitoring activity is

$$(2) \qquad P_{i2}(x_i,x_{i+1},\alpha) \equiv x_{i+1}(1-x_i)\Phi_2(\alpha)$$

The true efforts thus influence the probabilities of punishment in a multiplicative fashion. The probability that worker i is caught shirking rises when worker i+1 devotes more effort to monitoring, and/or when worker i slacks off in either production or monitoring. The technical efficiency of monitoring is determined by $\Phi_1(\alpha)$ and $\Phi_2(\alpha)$ for the two activities. These coefficients depend upon the allocation of time between production and monitoring as well as

[21]The one-to-one mutual monitoring structure describes the situation that at a particular point in time, a worker is able to check only one worker's job (not two), and that a worker's job is checked only once; double-checking is minimized by the team.

difficulty of collecting documentary evidence about each form of shirking.

In each case, the probability of punishment for shirking is a function of the time fraction spent on monitoring, for given levels of effort. The functions $\Phi_r(\alpha)$, $r = 1, 2$, have the following properties.

A2: (a) $\Phi_r(1) = 0$,

(b) $\Phi_r'(\alpha) < 0$,

(c) $\Phi_r''(\alpha) > 0$,

(d) $\lim_{\alpha \to 0} \Phi_r(\alpha) = k, \quad 0 < k < 1.$

Property (a) means that if no time is spent on monitoring, the probability of punishment is zero. Properties (b) and (c) state that as the time spent on monitoring decreases, the probability of punishment decreases at a decreasing rate, for any effort level less than one. Property (d) indicates that the probability of punishment for a shirker is bounded away from one as the time fraction spent on monitoring tends to one. As long as production occurs, a shirker can never be punished with certainty, no matter how much time has been devoted to monitoring.

4.2.3 The Payment Scheme

We assume a Cobb-Douglas agricultural production function, given by

$$(3) \qquad Q = \Theta(\alpha \Sigma_{j=1}^{n} w_j)^{\beta}$$

where Q is gross output in value terms. Θ is the technical coefficient of labour input.[22] $\alpha\Sigma_{j=1}^{n} w_j$ stands for total input of productive labour. It is assumed that the workers contribute equal time α to team production, despite their potentially different levels of

[22] In Chapter 5, I adopt the interpretation $\Theta = \lambda_0 K^{\lambda_2} T^{\lambda_3} F^{\lambda_4}$, where K, T, and F are capital, land, and non-labour current inputs, respectively. The parameter λ_0 captures the price at which output is sold. λ_i, $i = 2,3,4$ are output elasticities.

productive effort, w_i. The parameter, β, where $0 < \beta < 1$, is the elasticity of output with respect to labour input. The team's net revenue available for distribution is

$$(4) \qquad Y = Q - D,$$

where D is the sum of current farm expenditures, taxes, collective savings and welfare funds.

All net revenue is distributed among team members, but only a fraction $\gamma \epsilon [0,1]$ is distributed according to performance. The remainder is set aside for distribution according to "basic needs". The parameter γ is determined exogenously by ethical factors and the ideological preferences of the state. This places an upper limit on the size of the penalties that can be imposed for shirking, since an individual worker can be deprived at most of his/her claim on income that is distributed according to performance. [23]

To formalize this idea, write worker i's income as

$$(5) \qquad y_i = \frac{1}{n}(Q - D) - T_{i1}C_1 - T_{i2}C_2,$$

The workers share equally in net revenue because they are identical. The penalty $C_r \geq 0$ is paid whenever a worker is caught shirking in activity $r = 1,2$. The binary variables T_{ik} are defined as $T_{ir} = 1$ if worker i is caught shirking in activity r, and $T_{ir} = 0$ otherwise.

In this model, the penalties are introduced as an important choice variable of the team. This approach differs from Lin's model (1988). According to Lin, under perfect supervision the work points a worker earns are equal to his effort, e. Without supervision, a

[23]The net income of a Chinese team was distributed to individual households according to the work points earned by each household and the grain ration to which each household was entitled. The value of this ration distribution was offset against each household's work point earnings, and the balance was paid in cash. For those households whose work points were insufficient to cover their basic grain distribution, the team allowed them to go into debt for the amount of the short-fall, usually for a unlimited time period. The ration portion of household income was therefore perceived as an income floor and imposed a limit to the reward scale the team could implement to discriminate with respect to labour input performance. This is the intuitive story behind the specified income distribution scheme in this model.

worker is presumed to have worked at maximum intensity, e=1 and to be assigned full work points. He in fact assumes that a penalty equal to 1-e is imposed on a worker who is declared to have not worked at maximum intensity. This assumption is, however, not plausible. It is a common knowledge that the purpose of imposing a penalty is not merely to cover the loss but to deter workers from violating collective agreements. When a penalty takes its minimum possible value, all the burden of worker-discipline will inevitably lie on the side of monitoring. To have a balanced view of the roles that penalties and monitoring play in an effective labour supervision system, we introduce the penalties into the model explicitly and allow the size of penalties to be determined in the context of team optimization.

For tractability, it is assumed that the fines C_1 and C_2 are not transferred to any other team members, and represent pure social losses from the team's point of view.[24] The model differs from Holmstrom (1982), where individual effort is assumed to be unobservable, and the removal of free-rider problems requires a third party who is able to enforce group penalties or to finance bonuses. Under the present income distribution structure, the penalties imposed on shirkers can be enforced by non-shirking workers; there is no need for the presence of a "third party". However, to gain clarity and simplicity in the following mathematical proofs, we assume that prevailing legal institutions prevent non-shirking team members from appropriating the income withheld from shirkers ex post; e.g. penalties are paid to the state. This assumption will not alter the qualitative results of the study, because distributing the income withheld from shirkers back to the other team members could provide additional incentive for mutual monitoring, and therefore, strengthen rather than weaken the general results of the model.

The size of the penalties is limited by the proportion γ of income distributed according to performance. In other words, the team cannot penalize a shirker by cutting his/her income for subsistence. This imposes a credibility constraint on the penalties, i.e.

(6) $$\frac{\gamma}{n}[\Theta(\alpha\Sigma_{j\neq i}w_j)^{\beta} - D] \geq C_1 + C_2 \geq 0, \qquad \text{all } i \in N.$$

[24] A similar assumption is also made by Calvo and Wellisz (1978).

The constraint on penalties requires that it must be possible to apply both penalties simultaneously if worker i shirks both in production and in monitoring, while all others provide the levels of productive effort w_j, $j \neq i$ required by the team. The team must define the penalties according to an ex ante credibility constraint, because C_1 and C_2 have to be announced before production takes place. If the penalties C_1 and C_2 satisfy equation(6), and each worker i believes that all other workers will contribute the required labour inputs w_j, then each worker will believe that the penalties C_1 and C_2 can be enforced. This conjecture about the behavior of all other workers will be correct whenever the effort levels $w = (w_1,\ldots,w_n)$ form a Nash equilibrium. On the other hand, it will be assumed that if (6) fails to hold, the team members disregard the penalties, and the effective penalties are $C_1 = C_2 = 0$ [25].

4.2.4 The Team's Labour Contract

The team's internal incentive system can now be summarized by defining a two-part labour contract. The first part of the contract specifies the time dimensions of labour input. The second part of the contract specifies the levels of effort expected from a worker, and the penalties on a worker who is declared to have cheated on the effort contract. I limit alternatives to symmetric contracts where equal effort is expected of all workers. Thus, a contract between the team and an individual worker is given by

$$(7) \qquad W \equiv \{\alpha;\ w,\ x,\ C_1,\ C_2\}$$

where $w = w_i$, all $i \in N$, is the common level of required work effort and $x = x_i$, all $i \in N$, is the common level of monitoring effort. In the contract W, a worker is entitled to an equal

[25] MacLeod (1987) argues that the Nash equilibrium concept is most appropriate when modeling the incentive system of a cooperative firm, while the notion of dominant strategy equilibrium better models a non-cooperative firm. According to MacLeod, a labour contract ensuring that no individual worker prefers to shirk given that all other workers are expected to abide by the contract, is a self-enforcing contract for a production team.

share of the net revenue, contingent on the provision of labour measured by time to the team. Since working time can be counted easily by its natural unit, no resources need to be spent on enforcing workers' work time fractions. However, a worker's effort cannot be verified at zero cost. It is therefore necessary for the team to set up an enforcement mechanism involving the penalties C_1 and C_2, so that individual workers are not tempted to renege on the agreed effort levels (w,x).

The nature of the two-part labour contract introduces a Stackelberg type of relation between the team as a group and individual workers. Individual workers make their own effort decisions, taking α, C_1 and C_2 as given, after the team has collectively chosen efficient values for these parameters. In order to determine the optimal labour contract, however, it is essential to know how workers' effort decisions vary with the contractual parameters (α, C_1, C_2).

4.3 Characterization of Nash Equilibria

Assume that all workers dislike putting forth effort but enjoy consuming goods, and consumption equals income. A worker's expected utility can be written as

$$(8) \qquad U_i(w,x) = \frac{1}{n}(Q - D) - C_1 x_{i+1}(1-w_i)\Phi_1(\alpha) - C_2 x_{i+1}(1-x_i)\Phi_2(\alpha) - v_1 \alpha w_i - v_2(1-\alpha)x_i$$

where $w = (w_1,\ldots,w_n)$, and $x = (x_1,\ldots,x_n)$. Additive separability in income and effort is standard in models of moral hazard applied to team production. For tractability, we assume that workers are risk neutral with respect to income and that the disutility of effort is linear. The parameters $v_1 > 0$ and $v_2 > 0$ are the per unit disutility of work and monitoring, respectively. [26]

[26]While the disutility of labour is usually assumed to be nonlinear in the principal-agent literature, a simple linear form of the disutility of effort is standard in models of efficiency wages (Shapiro and Stiglitz, 1984). A constant marginal disutility of effort seems not a particularly unreasonable assumption for the peasants in a low income country. With such a simple specification, we are able to obtain a

In a one-shot game, each worker selects effort levels to maximize his/her expected utility, given his conjecture concerning the behavior of other workers. The derivatives of worker i's expected utility with respect to w_i and x_i are

$$(9) \qquad U_{iw}(w, x) = \frac{\beta}{n}\Theta\alpha^{\beta}(\Sigma_j w_j)^{\beta-1} + x_{i+1}\Phi_1(\alpha)C_1 - \alpha v_1$$

$$(10) \qquad U_{ix}(w, x) = C_2 x_{i+1}\Phi_2(\alpha) - (1-\alpha)v_2$$

The notation U_{iw} indicates the derivative of U_i with respect to w_i, and similarly for U_{ix}.

[illegible] pair (w^*, x^*) represent a symmetric Nash equilibrium where

[illegible] and $x_i^* = x^*$, all $i \in N$.

The corresponding effort vectors are $w^* = (w^*,\ldots,w^*)$ and $x^* = (x^*,\ldots,x^*)$.

Then the following equilibrium conditions must hold for all $i \in N$:

$$U_{iw}(w^*, x^*) \geq 0 \qquad \text{if } w^* = 1$$
$$U_{iw}(w^*, x^*) = 0 \qquad \text{if } 0 < w^* < 1$$
$$U_{iw}(w^*, x^*) \leq 0 \qquad \text{if } w^* = 0;$$

$$U_{ix}(w^*, x^*) \geq 0 \qquad \text{if } x^* = 1$$
$$U_{ix}(w^*, x^*) = 0 \qquad \text{if } 0 < x^* < 1$$
$$U_{ix}(w^*, x^*) \leq 0 \qquad \text{if } x^* = 0;$$

$$(11) \qquad \frac{\gamma}{n}[\Theta(\alpha(n-1)w^*)^{\beta} - D] \geq C_1 + C_2 \geq 0.$$

(11) is not a requirement for Nash equilibrium, yet it must hold if $C_1 > 0$ and/or $C_2 > 0$ are

[26](cont'd) corner solution for the optimal level of effort in this chapter, and thus simplify the empirical estimation in chapter 5.

to be enforceable when some worker unilaterally shirks in both production and monitoring.

We assume throughout that

A3: $v_1 \geq \frac{\beta}{n}\Theta^{1/\beta}D^{1-1/\beta}$.

Given the assumption in A3, if $\alpha > 0$, and $C_1 = 0$, then it can be shown that $U_{iw}(w^*, x^*) < 0$ for all $w_{min} < w^* \leq 1$. $w_{min} > 0$ is defined as the level of productive effort such that

$$\Theta[(n-1)w_{min}]^{\beta} \equiv D > 0. \quad (12)$$

We assume that some $w_{min} < 1$ exists, so that it is possible for the team to impose a positive penalty when a shirker offers zero effort, and all the other workers exert maximum effort $w = 1$ and $\alpha = 1$. Note that any common effort level w such that $0 < w \leq w_{min}$ must make $C_1 > 0$ and/or $C_2 > 0$ infeasible. Assumption 3 guarantees that there is a free-riding problem in team production whenever $w \geq w_{min}$, because an effort level $w \geq w_{min}$ cannot be sustained as a Nash equilibrium in the absence of monitoring and penalties. Without A3, all workers could voluntarily contribute enough productive effort to cover the fixed cost D even in the absence of monitoring. This would occur because the direct incentive provided by a worker's shared residual claim suffices to overcome the disutility of work. Assumption A3 rules out this possibility and will hold whenever the team is sufficiently large. Note that whenever $C_1 > 0$ and/or $C_2 > 0$ are feasible for $\alpha \in (0,1)$ and $w \in [w_{min}, 1]$, the relationship among the parameters Θ, D, n, and β,

A4: $\Theta(n-1)^{\beta} > D$

is assumed.

All equilibria are taken to be symmetric in what follows. We can therefore denote a particular equilibrium by $(\bar{w}, \bar{x})$, where $\bar{w}$ is the common effort level in production, and $\bar{x}$ is the common effort devoted to monitoring. The associated utility of worker i will be written as $\bar{U}_i(\bar{w},\bar{x})$. Team effort equilibria are described by the following sequence of propositions.

Theorem 1: Given A3 and $\alpha > 0$, $\bar{x} = 0$ if and only if $\bar{w} = \hat{w}$, where $U_{iw}(\hat{w},0) = 0$ and $0 < \hat{w} < w_{min}$. Shirking in production and monitoring ($\bar{w}=\hat{w}$, $\bar{x}=0$) is always an equilibrium. A typical worker in this equilibrium obtains the payoff

$$(13) \qquad \bar{U}_i(\hat{w}, 0) = \frac{1}{n}[\Theta(\alpha n\hat{w})^{\beta} - D] - \hat{w}\alpha v_1.$$

Proof: If $\bar{x} = 0$, then $U_{iw}(\bar{w}, 0) < 0$ for all $\bar{w} \in (w_{min}, 1]$, given A3. Since U_{iw} is strictly decreasing in $\bar{w}$, equilibrium requires $\bar{w} = \hat{w}$ where $U_{iw}(\hat{w},0) \equiv 0$, and $0 < \hat{w} < w_{min}$. On the other hand, if $\bar{w} = \hat{w} < w_{min}$, credible penalties cannot be imposed, and we must have $C_1 = C_2 = 0$. Therefore, $U_{ix}(\hat{w}, \bar{x}) = -(1-\alpha)v_2 < 0$. This implies $\bar{x} = 0$. Hence, ($\bar{w}=\hat{w}$, $\bar{x}=0$) is an equilibrium. The payoff expression $\bar{U}_i(\hat{w},0)$ is derived by substituting $w_i = \hat{w}$ and $x_i = 0$ in the expected utility function for all i. Q. E. D.

Theorem 2: $w_{min} < \bar{w} \leq 1$, and $0 < \bar{x} \leq 1$ form an equilibrium with credible penalties if and only if the following conditions hold for all $i \in N$

$$\text{(a)} \qquad \beta\frac{\Theta}{n}\alpha^{\beta}(n\bar{w})^{\beta-1} + \bar{x}\Phi_1(\alpha)C_1 - \alpha v_1 \geq 0 \text{ if } \bar{w} = 1; \quad = 0 \text{ if } w_{min} < \bar{w} < 1,$$

$$\text{(b)} \qquad \bar{x}C_2\Phi_2(\alpha) - (1-\alpha)v_2 \geq 0 \quad \text{if } \bar{x} = 1; \quad = 0 \quad \text{if } ^{\frown} < \bar{x} < 1,$$

$$\text{(c)} \qquad \frac{\Upsilon}{n}[\Theta(\alpha(n-1)\bar{w})^{\beta} - D] - C_1 - C_2 \geq 0,$$

A representative worker in such an equilibrium obtains the payoff

$$(14)\qquad \bar{U}_i(\bar{w},\bar{x}) = \frac{1}{n}[\Theta(\alpha n\bar{w})^{\beta} - D] - C_1\bar{x}(1-\bar{w})\Phi_1(\alpha) - C_2\bar{x}(1-\bar{x})\Phi_2(\alpha) - v_1\alpha\bar{w} - v_2(1-\alpha)\bar{x}.$$

Proof: Follows from the definition of an equilibrium and the credibility constraint for the penalties. A direct computation gives the utility of worker i. Q. E. D.

We will assume throughout that there exist credible equilibria with positive effort in production and monitoring which Pareto dominate the shirking equilibrium. [27] Hence, the shirking equilibrium ($\bar{w}=\hat{w}$, $\bar{x}=0$) can be ignored in searching for an optimal labour contract.

4.4 Team Optimization Program

4.4.1 Implications of An Optimal Labour Contract

The team collectively chooses the parameters of the labour contract $\{\alpha, C_1, C_2, \bar{w}, \bar{x}\}$ to maximize the expected utility of a representative worker subject to the constraint that the effort levels $(\bar{w},\bar{x})$ form a Nash equilibrium with credible penalties. Formally, the team optimization problem is defined as

$$\max_{\alpha,\bar{w},\bar{x},C_1,C_2} \quad \frac{1}{n}[\Theta(\alpha n\bar{w})^{\beta} - D] - C_1\bar{x}(1-\bar{w})\Phi_1(\alpha) - C_2\bar{x}(1-\bar{x})\Phi_2(\alpha) - \bar{w}\alpha v_1 - \bar{x}(1-\alpha)v_2$$

Subject to: $(IC_1)\ \bar{x}\Phi_1(\alpha)C_1 \geq \alpha v_1 - \frac{\beta}{n}\Theta\alpha^{\beta}(n\bar{w})^{\beta-1}$

$(IC_2)\ \bar{x}\Phi_2(\alpha)C_2 \geq v_2(1-\alpha)$

$(Cred.)\ [\frac{\gamma}{n}[\Theta(\alpha(n-1)\bar{w})^{\beta} - D] - C_1 - C_2 \geq 0.$

The first two conditions are the incentive compatibility constraints (ICs) for production and monitoring, which ensure $\bar{w} > 0$ and $\bar{x} > 0$ are incentive compatible. [28] The third constraint that guarantees $C_1 > 0$ and $C_2 > 0$ are credible penalties. In the remainder of this section, I

[27] In chapter 6, we prove this result with an assumption of $\bar{w} = x$.
[28] Technically, equality must hold in IC_1 (IC_2) if $\bar{w}$ (x) < 1. However, see Lemma 3 below.

assume that a feasible solution exists, and characterize its parameters. In section 4.3.2 below, I describe conditions under which some solution must exist.

Lemma 1: At a team optimum, α must have an interior solution such that $0 < \alpha < 1$.

Proof: If $\alpha = 0$, then the credibility constraint on the penalties is violated because $D > 0$. On the other hand, if $\alpha = 1$, then $\Phi_1(1) = \Phi_2(1) = 0$; we have $\bar{U}_{iw}(\bar{w}, \bar{x}) < 0$, given $\bar{w} \geq w_{min}$ and A3. Q. E. D.

Lemma 2: At a team optimum, $C_1 > 0$ and $C_2 > 0$.

Proof: If $C_1 = 0$, $\bar{U}_{iw}(\bar{w}, \bar{x}) < 0$, given $\bar{w} > w_{min}$ and A3. Likewise, with $C_2 = 0$, $\bar{U}_{ix}(\bar{w}, \bar{x}) < 0$, given $\alpha < 1$ by Lemma 1. Q. E. D.

Lemma 3: The incentive compatibility constraints can be assumed to hold with equality at a team optimum.

Proof: Suppose the inequality holds in IC_1. Since the objective function is non-increasing in C_1, a small reduction in C_1 does not reduce the objective function and is feasible, holding all other variables constant. So there is no loss of generality in assuming the first IC holds with equality if $C_1 > 0$ at an equilibrium. Similarly, we can assume the second IC holds with equality if $C_2 > 0$ at an equilibrium. Q. E. D.

Substituting the ICs into the objective function and using the credibility constraint, the team's optimization problem can therefore be written as

$$\max_{\alpha,\bar{w},\bar{x}} \quad \frac{\Theta}{n}(\alpha n\bar{w})^{\beta} + (1-\bar{w})\frac{\beta\Theta}{n}\alpha^{\beta}(n\bar{w})^{\beta-1} - \alpha v_1 - (1-\alpha)v_2 - D/n$$

$$\text{Subject to: } \frac{\gamma}{n}[\Theta(\alpha(n-1)\bar{w})^{\beta} - D] - (\bar{x}\Phi_1(\alpha))^{-1}[\alpha v_1 - \frac{\beta}{n}\Theta\alpha^{\beta}(n\bar{w})^{\beta-1}]$$

$$- (\bar{x}\Phi_2(\alpha))^{-1}(1-\alpha)v_2 \geq 0$$

Lemma 4: $\bar{x} = 1$ can be assumed at a team optimum.

Proof: The new objective function is independent of $\bar{x}$, and the credibility constraint is least binding when $\bar{x} = 1$. Q. E. D.

Now the choice variables of the team optimization problem can be reduced to α and $\bar{w}$. The team's optimization problem is then rewritten as

$$\max_{\alpha,\bar{w}} \quad \frac{\Theta}{n}(\alpha n\bar{w})^{\beta} + (1-\bar{w})\frac{\beta\Theta}{n}\alpha^{\beta}(n\bar{w})^{\beta-1} - \alpha v_1 - (1-\alpha)v_2 - D/n$$

Subject to: $\frac{\gamma}{n}[\Theta(\alpha(n-1)\bar{w})^{\beta} - D] - \Phi_1(\alpha)^{-1}[\alpha v_1 - \frac{\beta}{n}\Theta\alpha^{\beta}(n\bar{w})^{\beta-1}] - \Phi_2(\alpha)^{-1}(1-\alpha)v_2 \geq 0$

It will be assumed that the team's fixed cost D cannot be covered by the output of a single worker, even with maximum labour time and productive effort; that is,

A5: $D \geq \Theta$.

Lemma 5: For given α with $0 < \alpha < 1$, the team always chooses the largest feasible $\bar{w}$.

Proof: The derivative of the objective function with respect to $\bar{w}$ is $\frac{\beta\Theta}{n}\alpha^{\beta}n^{\beta-1}\bar{w}^{\beta-2}[\bar{w}(n-\beta) - (1-\beta)]$. It is greater than zero if $\bar{w} > \frac{1-\beta}{n-\beta}$. But $\bar{w} > w_{min}$ must hold at a team optimum with credible penalties. Given $0 < \alpha < 1$ and $D \geq \Theta$,

$$w_{min} = \frac{1}{n}\left(\frac{D}{\Theta}\right)^{1/\beta} > \frac{1-\beta}{n-\beta}.$$

Hence, the objective function is increasing in $\bar{w} \in [w_{min},1]$ for $0 < \alpha < 1$. Q. E. D.

It is assumed that

A6: $\beta\Theta n^{\beta-1} - v_1 + v_2 > 0.$

This assumption ensures that utility is increasing as a function of productive working time given $\bar{w} = 1$ and $\bar{x} = 1$ and all else equal.

Lemma 6: The credibility constraint on the penalties is binding at a team optimum, given A5 and A6.

Proof: Suppose the credibility constraint holds with strict inequality, and either $\bar{w} < 1$ or $\bar{w} = 1$ at a team optimum. If $\bar{w} < 1$, then by Lemma 5 the objective function can be increased by increasing $\bar{w}$ until the credibility constraint holds with equality. This contradicts optimality. If $\bar{w} = 1$ at the optimum, then the team can always increase its utility by increasing α and keeping $\bar{w} = 1$, due to A6. Again, this contradicts optimality. Q. E. D.

Lemma 7: There is no feasible contract with $\alpha > \alpha^*$, where α^* is the largest value of α such that $\bar{w} = 1$ is feasible.

Proof: Write the credibility constraint as $Z(\bar{w},\alpha) \geq 0$. By definition $Z(1,\alpha^*) = 0$ and $Z(1,\alpha) < 0$ for all $\alpha > \alpha^*$. We have

$$\frac{\partial Z(\bar{w},\alpha)}{\partial \bar{w}} = \frac{\beta\Theta}{n}\alpha^{\beta}\bar{w}^{\beta-2}[\gamma(n-1)^{\beta}\bar{w} - \Phi_1(\alpha)^{-1}(1-\beta)n^{\beta-1}] > 0.$$

The expression in brackets is strictly increasing in $\bar{w}$. Therefore, the maximum of $Z(\bar{w},\alpha)$ with respect to $\bar{w}$ on the interval $[w_{min},1]$ must occur at one of the endpoints. However, from the definition of w_{min} and the fact that $C_1 > 0$, $C_2 > 0$, it is easy to show that $Z(w_{min},\alpha) < 0$ for all $\alpha \in [0,1]$. Hence $Z(1,\alpha) < 0$ implies $Z(\bar{w},\alpha) < 0$ for all $\bar{w} \in [w_{min},1]$. Thus there is no feasible $\bar{w} \in [w_{min},1]$ when $\alpha > \alpha^*$. Q. E. D.

Theorem 3: At a team optimum, $\bar{w} = 1$, $\bar{x} = 1$, $0 < \alpha < 1$, $C_1 > 0$, and $C_2 > 0$ must hold, and the ICs and the credibility constraint hold with equality. A typical worker in the equilibrium obtains the payoff

$$(15) \qquad \bar{U}_i(1,1) = \frac{1}{n}[\Theta(\alpha n)^{\beta} - D] - \alpha v_1 - (1-\alpha)v_2.$$

Proof: Follows from Lemma 1 to 7. Q. E. D.

4.4.2 Existence of a Feasible Solution

From Theorem 3, we can summarize the team optimization problem as

$$\max_{\alpha} \quad \frac{\Theta}{n}(\alpha n)^{\beta} - \frac{D}{n} - \alpha v_1 - (1-\alpha)v_2$$

$$\text{Subject to: } \frac{\gamma}{n}[\Theta(\alpha(n-1))^{\beta} - D] - \Phi_1(\alpha)^{-1}[\alpha v_1 - \frac{\beta}{n}\Theta\alpha^{\beta}n^{\beta-1}] - \Phi_2(\alpha)^{-1}(1-\alpha)v_2 = 0.$$

Now the only choice variable of the team optimization program is α. Write the objective function as $T(\alpha)$, and the left-hand side of the credibility constraint as $Q(\alpha)$. $Q(\alpha) \geq 0$ defines the feasible set which has the following properties.

Fact A: Assume A3, A4, A5, and A6 hold. If Θ, γ and n are large, and v_1, v_2 and D are small but positive, then the feasible set $C \equiv \{\alpha: 0 < \alpha < 1, Q(\alpha) \geq 0\}$ is non-empty and compact.

Proof: 1.Nonemptiness: Choose some arbitrary fixed α_0 such that $0 < \alpha_0 < 1$, implying $\Phi_r(\alpha_0) > 0$, $r = 1,2$. Take the largest values of Θ, γ and n and the smallest values of v_1, v_2 and D which are permitted by A3-A6. We then have $D = \Theta$ by A5, $v_1 = \frac{\beta\Theta}{n}$ by A3 and A5, n is large but finite because A4 and $v_1 > 0$, $\beta > 0$ and $\Theta > 0$, and v_2 is small but positive, because A6 holds for any $v_2 \geq 0$ when A3 and A5 hold. It is easy to show that

$$Q(\alpha_0) = \gamma\Theta[(\alpha_0(n-1))^{\beta} - 1] - \Phi_1(\alpha_0)^{-1}\alpha_0\beta\Theta[1 - \frac{1}{\alpha_0 n}^{1-\beta}] - \Phi_2(\alpha_0)^{-1}(1-\alpha_0)v_2 n \geq 0.$$

2.Closedness: $Q(\alpha)$ is continuous on the open interval $\alpha \in (0,1)$. Now consider the limiting cases when α approaches its boundary:

(a) $\lim_{\alpha \to 0} Q(\alpha) = -\gamma D/n - k^{-1}v_2 < 0,$

and

(b) $\lim_{\alpha \to 1} Q(\alpha) = -\Phi_1(1)^{-1}[v_1 - \beta\Theta n^{\beta-2}] = -\infty < 0,$

in (b) $v_1 > \beta\Theta n^{\beta-2}$ follows A3 and A5 and $\Phi_1(1) = 0$. All points near the endpoints of the interval [0,1] have $Q(\alpha) < 0$. The boundary of the feasible set must therefore lie entirely in the interior of [0,1]. Since $Q(\alpha)$ is continuous, the feasible set C is a closed set.

3.Boundedness: Because the interval [0,1] is bounded, C is also bounded, and hence is compact.

Q. E. D.

Theorem 4: There exists some feasible solution for the team's optimum if the conditions in Fact A hold.

Proof: The objective function $T(\alpha)$ is continuous on the interval [0,1]. Also, according to Fact A, given the condition of Theorem 4, the feasible set $C \equiv \{\alpha: 0 < \alpha < 1, Q(\alpha) \geq 0\}$ is non-empty and compact. From the Weierstrass theorem, there exists a solution for the present team optimization problem.

Q. E. D.

The intuitive interpretation of Theorem 4 is that to implement an effective monitoring scheme, the team must possess sufficient economic rent. If the profitability of team production is restricted by the external policies of the state, an effective monitoring system may be infeasible. This could occur, for example, if taxation results in a large value of D; if agricultural pricing policies limit the value of Θ; or if γ, the proportion of income based on performance, is too low for ideological reasons.[29]

[29]The concept of economic rent adopted here is defined as the income available for discriminating with respect to labour input performance. In the context of Chinese

4.4.3 Comparative Statics

Since the credibility constraint must hold with equality at the optimum by Theorem 3, we can set up the Lagrangean as,

(16) $$L = T(\alpha) + \mu Q(\alpha).$$

We then obtain the first order conditions

(17) $$\frac{\partial L}{\partial \alpha} = \frac{dT(\alpha)}{d\alpha} + \mu\frac{dQ(\alpha)}{d\alpha} = 0$$

(18) $$\frac{\partial L}{\partial \mu} = Q(\alpha) = 0$$

At the team equilibrium, we cannot have $\frac{dQ}{d\alpha} \geq 0$. Since $\frac{dT(\alpha)}{d\alpha} > 0$ by A6, and $\mu > 0$ as a result of no inequality in the constraint by Theorem 3, $\frac{\partial L}{\partial \alpha} = 0$ holds only if $\frac{dQ(\alpha)}{d\alpha} < 0$. If not, the team could increase α, and then raise the objective function without violating the credibility constraint. The optimal solution of α is given by equation (18), along with the fact that $\frac{dQ}{d\alpha} < 0$ at the optimum. Equation (18) will be estimated with some modification in chapter 5.

Theorem 5: If the conditions of Theorem 4 are met, the team optimization problem has a unique solution of the form

(19) $$\alpha^* = \alpha^*(\Theta, n, D, \gamma, v_1, v_2)$$

(20) $$C_1^* = [\alpha^* v_1 - \frac{\beta}{n}\Theta\alpha^{\beta} n^{\beta - 1}]\Phi_1(\alpha^*)^{-1}$$

(21) $$C_2^* = [v_2(1 - \alpha^*)]\Phi_2(\alpha^*)^{-1}.$$

Proof: First, (18) has continuous partial derivatives with respect to α and all the parameters if a feasible solution exists. Second, the derivative of $Q(\alpha)$ with respect to α is less than zero

[19](cont'd) agricultural teams, "economic rent" is referred to as the distributable income net of the basic food consumption.

at the optimum. We know that there is some unique maximum value of α such that $Q(\alpha) = 0$ holds. From the implicit-function theorem, (18) defines the solution for α as a function of the parameters in some neighbourhood of this optimal α, and the solution has continuous partial derivatives with respect to all parameters.

The optimal penalties are obtained by substituting α^* into the no-shirking conditions which hold with equality. Since the optimal solution for α is unique, the optimal penalties are also unique. Any deviation from the penalties defined by (20) and (21) will violate either the credibility constraint or the incentive compatibility constraints. Q. E. D.

Now we turn to the relationships between the optimal time allocation and the parameters of the model. The results of a simple comparative static study are given below.

(1) $$\frac{\partial\alpha^*}{\partial\Theta} = [-\frac{1}{n}\alpha^{\beta}(\gamma(n-1)^{\beta} + \Phi_1(\alpha)^{-1}n^{\beta-1}\beta)][\frac{dQ}{d\alpha}(\alpha^*)]^{-1} > 0$$

(2) $$\frac{\partial\alpha^*}{\partial D} = \gamma[\frac{dQ}{d\alpha}(\alpha^*)]^{-1} < 0.$$

(3) $$\frac{\partial\alpha^*}{\partial\gamma} = [\frac{dQ}{d\alpha}(\alpha^*)]^{-1}[-(\frac{1}{n}\Theta((\alpha(n-1))^{\beta} - D/n] > 0$$

(4) $$\frac{\partial\alpha^*}{\partial n} = [\frac{dQ}{d\alpha}(\alpha^*)]^{-1}[\frac{\Theta}{n}\alpha^{\beta}(\gamma(n-1)^{\beta}(\frac{1}{n} - \frac{\beta}{n-1}) - \Phi_1(\alpha)^{-1}\beta(\beta-2)n^{\beta-2})] < 0,$$
$$\text{if } 0 < \beta < \frac{n-1}{n}$$

(5) $$\frac{\partial\alpha^*}{\partial v_1} = (\alpha/\Phi_1(\alpha))[\frac{dQ}{d\alpha}(\alpha^*)]^{-1} < 0$$

(6) $$\frac{\partial\alpha^*}{\partial v_2} = [(1-\alpha)/\Phi_2(\alpha)][\frac{dQ}{d\alpha}(\alpha^*)]^{-1} < 0$$

(7) $$\frac{\partial\alpha^*}{\partial\omega} = -[(\alpha v_1 - \frac{\beta}{n}\Theta\alpha^{\beta}n^{\beta-1})\Phi_1^{-2}(\alpha)\frac{\partial\Phi_1}{\partial\omega} + (1-\alpha)v_2\Phi_2^{-2}(\alpha)\frac{\partial\Phi_2}{\partial\omega}][\frac{dQ}{d\alpha}(\alpha^*)]^{-1} > 0$$

where in (7) ω is a shift parameter that increases the effectiveness of monitoring, so that $\frac{\partial\Phi_1}{\partial\omega} > 0$, and $\frac{\partial\Phi_2}{\partial\omega} > 0$.

The impact of changing each of the parameters on the team's time allocation can be easily interpreted. A rise in Θ implies either an increase in the price of the product or an improvement in the technical efficiency of production, or both. By increasing the team's distributable income, a rise in Θ aggravates the direct cost to individual workers of shirking in production, and thus reduces the motive for shirking. It also allows the team to discipline the shirkers with higher enforceable penalties. As a result, the team is able to reallocate some of the time spent on monitoring to production without violating the non-shirking conditions. A fall in Θ forces the team to increase its monitoring time in order to avoid shirking. A poor team therefore has to invest more labour input in monitoring than a rich team, ceteris paribus.

A change in production cost, D, has a negative impact on the time allocated to production. This happens because a higher production cost implies a lower net income available for distribution. The penalties must therefore be smaller to remain credible. This will force the team to invest more resources in monitoring in order to prevent workers from shirking.

An increase in the proportion of income distributed according to labour performance has a positive impact on the team's productive time and a negative effect on the time spent on monitoring. A large ratio of distribution according to need limits the magnitude of the penalties that the team is able to impose on a shirker. To compensate for this decline in the credible penalties for shirking, the team must monitor more intensively.

An increase in the size of the team tends to dilute the negative effect of shirking in production by relaxing the credibility constraint and allowing larger penalties on individual shirkers, but also reduces the direct cost to a worker from shirking in production through worker's residual claim. However, the impact of the latter dominates that of the former under fairly general conditions. Hence, as n rises, the team has to raise the share of time allocated to monitoring, and thus reduce productive work time.[30]

[30] Putterman comments that many authors assume the same effect for reasons of greater difficulty of monitoring, but this model rules that out by the assumption of Figure 1 structure.

An increase in the disutility of productive work or monitoring has a negative effect on the fraction of time allocated to production. The shirking motives in both production and monitoring are strengthened as productive work or monitoring becomes more unpleasant. To prevent workers from shirking, the team must reduce its productive labour input in order to monitor more intensively. An improvement in the efficiency of the monitoring technology allows the team to reduce the required monitoring input, and therefore increases productivity.

4.5 Conclusion

In this chapter, we have constructed a static mutual monitoring model in which workers engage in both production and monitoring. The workers' performance in each activity is controlled by formal labour supervision which is conducted by workers themselves. We also examined the nature of the workers' behavior in a non-cooperative Nash equilibrium. It is found that workers will contribute the optimal level of effort if and only if other workers are also expected to adhere to the team's labour contract.

The major contribution made in this chapter to the literature is that we integrate the selection of a payment scheme with the choice of monitoring intensity in a team's optimization program. With this general approach, we are able to highlight the economic, technological and political constraints facing Chinese production teams in the field of labour supervision. It is demonstrated that the possibility of rent is a necessary condition for an effective monitoring system, because otherwise it is not possible to impose substantial penalties on shirkers. If the government's policies toward pricing, marketing, or production planning diminish the technical efficiency of a team's production, or an egalitarian ideology severely restricts the magnitude of penalties a team can impose, the operation of a monitoring system will become rather costly or even impossible. The team's managerial ability, the technological characteristics of the production process, the size of the team, the disutility of work, and tastes for engaging in mutual monitoring also play important parts in an effective mutual monitoring system. However, the impact of these factors on the efficacy of a mutual monitoring system can be offset by larger penalties for shirking, if the economic environment and government policies enable a team to generate large rents through its production activities.

Chapter 5

Labour Supervision in a Chinese Commune (I)

5.1 Introduction

In this chapter, the static incentive model is examined using the detailed micro-level data from Dahe People's Commune in Hebei Province in North China. The data were collected by Steven Butler in 1979-80 and by Louis Putterman in 1986. Dahe commune is located 12 miles west of Shijazhuang, the capital of Hebei province. Dahe's principal products before the 1980's rural reform were wheat, corn, and cotton. Owing to its advantage in location, the standard of living in Dahe commune was above the average for rural China as a whole. Butler (1985) describes it as "broadly representative of moderately prosperous grain-producing regions of North China."

The data set covers the years 1970 to 1985. During the 1970's Dahe commune consisted of 16 production brigades, which were further subdivided into an average of about 100 production teams. The production teams, each containing 51 households and about 80 able-bodied adult workers, had been the basic units of farm production, distribution and accounting until the early 1980's. To examine the economic and political constraints on Chinese collective farms' incentives, we will in the following study focus on the period 1970-76, when China was undergoing the Cultural Revolution.

Beginning in the mid-1960's, the production teams in Dahe adopted a form of the Dazhai work point system. [31] The Dazhai work point system was a form of time-based renumeration associated with close labour supervision. The team assigned workers a definite number of working days and specified the job tasks to be fulfilled during a length of time according to the needs of production and the approval of the general meeting of the team members. The team members earned work points according to the number of days worked and a retrospective evaluation of task fulfilment by their teammates. The income distribution structure under this type of renumeration appeared to be egalitarian. If no worker was

[31] The Dahe work point system was introduced in detail in chapter 2.

accused of having done a substandard job, the observed income variation across workers reflected merely differences in age, sex and perceived strength and skills. Nonetheless, how effective such a supervisory system could be in preventing workers from shirking depended upon the team's ability to detect and punish the workers for substandard work. According to Butler's field study (1983), many local cadres reported that the Dazhai work point system worked reasonably well at first, yet labour supervision became more and more difficult due to the policy distortions during the Cultural Revolution era.

The production teams had little control over their production process during the period. Under the national guideline of "taking grain as a key link", the production teams were directed to promote cereal production through the use of increasing amounts of expensive modern inputs. The production cost then grew at an annual rate of 14.4 % from 1970 to 1976, while the state purchasing prices for farm products remained frozen. As a result, the ratio of team current non-labour production expenses to gross income rose from 0.35 in 1970 to 0.52 in 1976. During 1969 to 1976, yields more than doubled at Dahe; yet the value of a labour day fell from 0.51 *yuan* in 1970 to 0.41 *yuan* in 1976, a reduction of 20%. [32] The production teams thus could not improve their members' standard of living through farm production.

Apart from the state's irrational production strategy and unfavorable terms of trade, the heavy financial obligation of the teams to local development and public welfare further reduced the material incentives the teams could provide for their members. At Dahe commune from 1970 to 1976, the average accumulation, welfare and reserve funds of the production teams accounted for 21.3 % of net, after-tax income. The average percentage of net income actually distributed to team members was only 78.7 %. The annual distributed income per worker was around 200 *yuan*.

In principle, the entire revenue of each team net of non-labour costs, collective savings, and welfare funds was distributed according to work points earned, yet a significant proportion of the income distributed was actually invariant with respect to labour contribution

[32] *yuan* is one unit of the Chinese currency *reminbi* (RMB). The official exchange rate was roughly 1.4 RMB to 1 U. S. dollar in 1980.

to the teams. To guarantee basic food needs, some portion of the income in the form of staple food-stuff produced was rationed out according to household demographics. Although grain rationed was offset against each household's work point earnings, the households whose work points were insufficient to cover their basic grain distribution were allowed to go into debt for the amount of the short-fall at zero interest and over indefinite periods of time. The ration portion of distributed income represented an income floor which limited the penalty that the team could impose. During 1970-1976, the annual average grain ration per capita at Dahe was about 212 kgs in volume and 59.4 *yuan* in value. The needs distribution was further enlarged by the fact that some portion of the team population such as the old, weak, sick and disabled and children did not actually participate in team production, but their basic food needs were also guaranteed by the teams. In Dahe commune, an able-bodied adult worker supported an average of 2.5 persons. As a result, grain rations accounted for 60% to 80% of the distributed grain, and an average of 58.5% of the total distributable income. The income actually distributed according to workers' performance accounted for only 41.5% of the distributable income. The high proportion of needs distribution exacerbated the difficulties of labour supervision in the production teams.

The remainder of the chapter is organized as follows. The next section presents the derivation of the estimating equation and hypotheses regarding static mutual monitoring in Chinese production teams. The third section discusses the estimation results. The final section summarizes the findings of the chapter.

5.2 The Estimated Model

In section 4.4.3, we developed a team's optimal time allocation equation,

$$(1) \qquad \frac{\gamma}{n}[\theta(\alpha(n-1))^{\beta} - D] - \Phi_1(\alpha)^{-1}[\alpha v_1 - \frac{\beta}{n}\theta\alpha^{\beta}n^{\beta-1}] - \Phi_2(\alpha)^{-1}(1-\alpha)v_2 = 0.$$

To estimate this equation, some modifications have to be made. Assuming that the monitoring technologies adopted are the same whether a worker is monitoring his/her co-worker's productive effort or the co-worker's monitoring activities, i.e. $\Phi_1(\alpha) = \Phi_2(\alpha) = \Phi(\alpha)$, and that the disutilities of production and monitoring are identical, i.e. $v_1 = v_2 = v$, equation (1) can then be written as

$$(2) \qquad \frac{\gamma}{n}[\Theta(\alpha(n-1))^{\beta} - D] - \Phi(\alpha)^{-1}[v - \frac{\beta}{n}\Theta\alpha^{\beta}n^{\beta-1}] = 0.$$

The production function is specified as

$$(3) \qquad Q = \lambda_0 L^{\lambda_1} K^{\lambda_2} T^{\lambda_3} F^{\lambda_4}$$

where Q is gross output in value terms. $L = \alpha\Sigma_{j=1}^{n} w_j$ is labour input; given the optimal level of productive effort, $\bar{w} = 1$, $L = \alpha n$.[33] K, T and F are capital stock, sown area and current farm expenditures, respectively. λ_0 is the product of the price of the output and the technical efficiency of production, and λ_i, i = 1,2,3,4 define output elasticities. Thus the parameter Θ from Chapter 4 is given by $\Theta = \lambda_0 K^{\lambda_2} T^{\lambda_3} F^{\lambda_4}$, and $\beta = \lambda_1$. Assuming that the production function displays constant returns to scale, i.e. $\Sigma_{j=1}^{4}\lambda_j = 1$, the production function at the optimal work effort $\bar{w} = 1$ can be written as

$$(4) \qquad q = \lambda_0 \alpha^{1-\Sigma\lambda} k^{\lambda_2} t^{\lambda_3} f^{\lambda_4}.$$

Here q, k, t, and f represent gross value of output, capital stock, sown area and current farm expenditures per worker, respectively, and $\Sigma\lambda = \lambda_2 + \lambda_3 + \lambda_4$.[34]

It is assumed that the monitoring technology is described by an exponential function which has all the properties assumed in section 4.1.2, that is,

[33] By Theorem 3 in chapter 4, an optimal labour contract must have $\bar{w} = 1$.

[34] The assumption of constant returns to scale will be tested by estimating the following equation,

$$\gamma[q((n-1)/n)^{\lambda_1} - d] - [1 - \exp(\omega((qxn/(\lambda_0 n^{\lambda_1} K^{\lambda_2} T^{\lambda_3} F^{\lambda_4}))^{1/\lambda_1} - 1))]^{-1} \{v - qx\lambda_1/n\} = u.$$

This is the unrestricted version of equation (6) below.

(5) $\Phi(\alpha,\omega) = 1 - \exp[-\omega(1-\alpha)].$

ω is the technical coefficient of monitoring. Replacing β by $1 - \Sigma\lambda$, the team's optimal time allocation equation developed in Section 4.4.3 is now written as

$$(6) \qquad \gamma_i[\lambda_0(\alpha_i(n_i-1)/n_i)^{1-\Sigma\lambda}k_i^{\lambda_2}t_i^{\lambda_3}f_i^{\lambda_4} - d_i] - [1-\exp(-\omega(1-\alpha_i))]^{-1} [v - (1-\Sigma\lambda)/n_i\lambda_0\alpha_i^{1-\Sigma\lambda}k_i^{\lambda_2}t_i^{\lambda_3}f_i^{\lambda_4}] = u_i.$$

Here i is the team index, u_i is a stochastic term, γ_i is the proportion of the income distributed according to performance, and $d_i \equiv D/n$, represents cost and saving deduction per worker.

Equation (6) cannot be directly estimated because data on the time fraction spent in production, α, are unavailable to us. To proceed with the estimation, substitute gross income per worker q for the time fraction on production α, using the identity under the assumption of constant returns to scale,

$$(7) \qquad \alpha = \{q/(\lambda_0 k^{\lambda_2}t^{\lambda_3}f^{\lambda_4})\}^{1/(1-\Sigma\lambda)}.$$

This gives us

$$(8) \qquad \gamma_i[q_i\{(n_i-1)/n_i\}^{1-\Sigma\lambda} - d_i] - [1 - \exp(\omega[(q_i/(\lambda_0 k_i^{\lambda_2}t_i^{\lambda_3}f_i^{\lambda_4})]^{1/(1-\Sigma\lambda)} - 1))]^{-1} \{v - q_i(1-\Sigma\lambda)/n_i\} = u_i.$$

In this equation, the λ_k, v, and ω are the parameters to be estimated. q_i is the endogenous variable, and γ_i, d_i, k_i, t_i, f_i and n_i are exogenous variables. The exogeneity assumptions are justified as follows. Labour in rural China is not mobile among different accounting units owing to the collective ownership of land and capital. Team size n_i was thus essentially determined by demographic factors. Capital, K_i, and land, T_i, are fixed in the short run. Current variable expenditure F_i is also treated as predetermined, because the teams' employment of modern variable inputs was enforced primarily through heavy top-down administrative pressure. The exogeneity of the variables d_i and γ_i reflects the fact that the members of Chinese collective farms had little or no control over this aspect of their income schemes.

The theoretical model of Chapter 4 provides several testable hypotheses. First, it is a conventional belief that individual workers have incentives to free ride on their co-workers' labour contribution in a one-shot game due to their shared residual claimancy. In other words, a shirker in production bears only 1/n of the income loss and saves a full unit of disutility. Since direct income sharing incentives are not large enough to outweigh the gain in leisure from shirking, workers will slack off in team production. This belief was formalized in chapter 4 through the assumption A3; nonetheless, A3 is a sufficient but not necessary condition for the existence of free-riding problems in team production. So the hypothesis is formulated based on the derivatives of worker i's expected utility with respect to w_i at $\bar{w} = 1$, $\alpha = \alpha^*$ and $C_1 = 0$ instead of A3. That is,

$$H_{01}: \alpha v - \{(1-\Sigma\lambda)/n\}\lambda_0\alpha^{1-\Sigma\lambda} k^{\lambda_2} l^{\lambda_3} f^{\lambda_4} = 0$$
$$H_{a1}: \alpha v - \{(1-\Sigma\lambda)/n\}\lambda_0\alpha^{1-\Sigma\lambda} k^{\lambda_2} l^{\lambda_3} f^{\lambda_4} > 0$$

where H_{a1} is consistent with free-riding and H_{01} is the null of no free-riding.

Second, in contrast to the conjecture that labour supervision in Chinese production teams was prohibitively costly and therefore almost zero (Lin, 1988),[35] it is hypothesized that an effective labour supervision system was actually implemented in Chinese collective farms. In other words, both the resources invested in labour supervision and the probability of catching and penalizing a shirker in the team were greater than zero. That is,

$$H_{02}: 1 - \alpha = 0$$
$$H_{a2}: 1 - \alpha > 0$$

[35]See the discussion in chapter 3.

where H_{a2} implies non-zero monitoring investment and H_{02} is the null of no labour supervision. These hypotheses imply, respectively,

$$H_{03}: 1 - \exp(-\omega(1-\alpha)) = 0$$
$$H_{a3}: 1 - \exp(-\omega(1-\alpha)) > 0.$$

Third, the Chinese production teams made the optimal decision on resource allocation in labour supervision under the economic and political constraints they faced. The team's optimum is characterized by a binding credibility constraint and the negativity of the derivative of the credibility expression Q(α) with respect to α, as shown in Section 4.3.2. Since the credibility constraint is assumed binding in the estimation, we will test

$$H_{04}: \frac{dQ(\alpha)}{d\alpha}\Big|\alpha^{*} = 0$$
$$H_{a4}: \frac{dQ(\alpha)}{d\alpha}\Big|\alpha^{*} < 0.$$

Fourth, the so called "internal incentive problems" of Chinese production teams are not the simple outcome of team farming per se, but the consequences of the dysfunctional economic and political policies adopted during the Cultural Revolution era. Following the comparative static study conducted in Section 4.4.3, we conjecture that the team's monitoring input (1-α) is negatively related to the price of output and/ or production efficiency, the proportion of the income distributed based on work, and the efficiency of monitoring technology, and positively related to the cost and saving deduction of the team's gross revenue, the team size, and the disutility of production and monitoring, i.e $\frac{\partial\alpha}{\partial\lambda_0} > 0$, $\frac{\partial\alpha}{\partial d} < 0$, $\frac{\partial\alpha}{\partial\gamma} > 0$, $\frac{\partial\alpha}{\partial n} < 0$, $\frac{\partial\alpha}{\partial v} < 0$, and $\frac{\partial\alpha}{\partial\omega} > 0$. The empirical expressions of the derivative of Q(α) with respect to α and the derivatives of α with respect to λ_0, d, γ, n, v and ω are presented in Appendix 5.1.

Fifth, it is hypothesized that the distribution of income by needs was determined by ethical and ideological factors, rather than the efficiency-improving considerations suggested by Sen (1966).[36] In other words, the proportion of the income distributed according to work,

[36]See chapter 3 for details.

γ_i, is an exogenous variable and a test for exogeneity is executed.

The variables involved in the estimation are defined as follows. The team's total gross income Q is defined as the sum of crop, orchard, animal husbandry, transportation, fishery, and sideline incomes plus income earned by the team for labour services supplied to commune and brigade enterprises and to other outside units. The teams' production is specified in such broad terms because it was the aggregate revenue rather than the revenue from crop production alone that determined team members' shared income, and also because some inputs cannot be disaggregated.

Team size, n, is measured by the number of able-bodied adult workers in the team. Land, T, is represented by sown area in *mu*. One *mu* equals 0.1647 acres. The proxy for capital, K, is gross value of capital assets in *yuan*. Capital assets consist of tractors, diesel engines, electric motors for pumping water, sowing, harvesting, and threshing machines, winnowers, grinders, sprayers, blowers, pumps, buildings, wells, pig sties, mules, horses, donkeys, oxens, trucks, human-drawn carts, and other vehicles and carts. Both land and capital are stock rather than flow measures. Because land and capital markets did not exist in China, adequate measures for the service flows from these inputs are not available.

Current variable expenditures, F, include such items as seed expenditures, fertilizer expenditures, agricultural chemical expenditures, repair expenses, draft animal expenses, machine plowing expenses, irrigation expenses, other agricultural expenditures, forestry and orchard, animal husbandry, fisheries, and "sideline" expenses, and management expenses. F is also measured in *yuan*.

Variable D is taken as the sum of current farm expenditures, taxes, accumulation, welfare and reserve funds. It is also measured in *yuan*. Subtracting D from the gross total income gives us the team's total distributable income. The income distributed according to needs is represented by grain valued at the state purchase prices and rationed out according to household population. One minus the ratio of the income distributed based on household demographics to the total distributable income is the proportion of the income distributed according to labour contribution, γ.[37] Dividing Q, K, T, F, and D by n respectively generates variables q, k, t, f, and d.

5.3 Estimation and Results

The estimation was performed for each of seven cross-sectional data sets. Due to missing-data problems, the number of production teams for which all required variables were available varies over the periods. More precisely, there were 38 teams for 1970, 37 for 1971, 38 for 1972, 47 for 1973, 58 for 1974, 78 for 1975, and 75 for 1976. The data were treated as separate cross-sectional studies for two reasons. First, only a small number of production teams had all required variables for a continuous set of 7 years. To conduct a time-series and cross-sectional study, too many cross-sectional observations had to be excluded. Second, some variables such as q, k, f, and d were evaluated at current prices, but an ideal price deflator for these variables was not available to handle nominal variation across time periods.

Equation (8) was estimated by the nonlinear two-stage least squares (NL2S) method using the Gauss-Newton algorithm in TSP version 4.1. The estimation procedure is outlined in Appendix 5.2. The NL2S estimator is consistent and is asymptotically normally distributed under certain assumptions. [38]

Heteroskedasticity is a potential problem in cross-sectional studies. In the presence of heteroskedasticity, the NL2S estimators will be inefficient and their estimated standard errors will be biased. The heteroskedasticity test suggested by Koenker (1981) was used. The test is

[37]The definitions of the variables involved in terms of the series numbers in the published Dahe commune data set are listed as follows: Q = v18, n = v55 + 0.5 x v56, K = v128, T=v60 + v69 + v72 + v78 + v81 + v84, F = v35, D = v35 + v37 + v44, and γ = 1 - {(v45-v46) x v154 x 0.01/v45}. The notation v18 stands for variable 18, and similarly for other variables. (See Putterman (1987).) Ideally, we should use the number of standardized able-bodied adult workers as a proxy for the size of the team, n, for the reason discussed in footnote 18 in chapter 4. However, data on the age, sex and physical strength composition of the workers were not available. The variable n used in this estimation presents a partial adjustment of labour standardization assumed in chapter 4 due to the data constraint.

[38]Since the asymptotic efficiency of the NL2S estimator depends upon the choice of the instruments rather than the normality of disturbances, the test for normality is omitted (Amemiya A, 1974).

defined as the number of observations times the R^2 from the regression of the standardized residual squares on all explanatory variables. The test results are reported in Table 5.1. The hypothesis of homoskedastic variances of disturbances cannot be rejected at the 5% level of significance in any of the regressions except that for 1973. To correct the inconsistency of the NL2S estimates of standard errors caused by the heteroskedasticity problem, the heteroskedasticity consistent variance-covariance matrix estimates of the NL2S are reported for the year 1973, while those reported for the rest of the regressions are the NL2S estimates.

The Hausman test was conducted to examine the exogeneity of the proportion of the income distributed on performance, γ. The test procedure is outlined in Appendix 5.2, and the test results are reported in Table 5.6. The hypothesis that γ is an exogenous variable cannot be rejected at the 5% level of significance in any of the cases except 1970. The test statistic for 1970 is negative, because the variance-covariance matrix of the vector of contrast is negative definite. In this case, the Hausman test fails. [39] Comparing the estimated results for 1970 in Table 5.1 and 5.6, we note that the results are very close. These results provide support for the hypothesis that the production teams have little control over this aspect of their income distribution schemes.

Now we turn to the main results. To assess the performance of the present model specification, we estimated the production function in a conventional log-linear form using the OLS technique. The assumption of constant returns to scale is imposed, because this assumption cannot be rejected at the 5% level of significance in each of the years. The results are reported in Table 5.3. The outcome of the OLS estimates is not particularly impressive in terms of the adjusted R^2 especially for the early periods. The estimated output elasticity of capital in two cases and land in one instance has the "wrong" sign. Eight of the 14 estimates on capital and land, including those with the "wrong" sign, and one of the 7 estimates on current farm expenditures are statistically insignificant. Of those with significant test statistics, the average output elasticity with respect to labour, capital, land, and current farm expenditures is 0.35, 0.13, 0.25, and 0.53, respectively. By comparison, the factor shares for

[39]The failure of the Hausman test may be explained by the poor performance of a large sample test in a finite sample excercise.

Chinese agriculture in the 1950s suggested by Anthony M. Tang (1980) are 0.50 for labour, 0.10 for capital, 0.25 for land, and 0.15 for current inputs. It is argued that this set of weights is consistent with those for Southeast Asian countries during the compatible stage of economic development. Using Tang's result as a benchmark, the estimates of the conventional log-linear production function appear to be low for labour and high for current inputs.

The estimation bias for the conventional production function is further evident by the calculated average values of the marginal product of current expenditures reported at the bottom of Table 5.3. The marginal product of expenditures is defined as the gross revenue per 1 *yuan* increase in expenditure on current farm inputs. The competitive allocational rule for the factor implies the value of this marginal product should equal unity. However, it was observed that during 1969 to 76, yields more than doubled at Dahe; yet the peasants' cash income declined because the ratio of current non-labour production expenses to gross income rose by 49 percent. A decline in the team's income net of current non-labour farm cost implies that the marginal return to expenditures on current inputs was unable to cover the marginal factor cost, i.e. the marginal product of current farm expenditures was less than one. However, the estimated average values of the marginal product of current expenditures from the conventional production functions regression shown in Table 5.3 ranges from a low of 0.77 to a high of 1.53, with an average value over the 6 estimates of 1.27 yuan. This result fails to truly reveal the resource mis-allocation existing in Dahe.

The unsatisfactory estimates for factor inputs may be explained by the inadequate data on capital and land stock and the potential measurement error in labour input. In the conventional production function estimation, number of workers was used as a proxy for labour inputs, implicitly assuming that $\alpha = 1$ and $\bar{w} = 1$. However, a proportion of labour input was in fact taken away from direct production as an agency cost paid for team production. When the variables are subject to the measurement error, the estimates will be biased and inconsistent. The empirical results could be improved by estimating the production function in an incentive context, since the mutual monitoring model introduced a mechanism to correct the potential measurement error in labour input.

The estimated results of the time allocation equation are presented in Table 5.1.[40] Comparing the estimates of the production function in Table 5.1 with those in Table 5.3, we note that the non-conventional approach has brought about some mild improvement in the testing significance. All the 7 estimated output elasticities of current farm expenditures are significant at the 5 % level or better. The estimates of land and capital in Table 1 are not worse than the OLS results. The average output elasticities of those with significant test statistics are 0.76 for labour, 0.09 for capital, 0.19 for land, and 0.14 for current farm expenditures. While the labour coefficients are high, the estimates of the other inputs appear to be compatible with Tang's result. The high values of labour coefficients in most cases are caused by the poor estimates for land and capital. With the assumption of CRS, the labour coefficients were calculated residually. Therefore, we note that the output elasticities for labour are high when the corresponding estimates for land and capital are not significantly different from zero or have a negative sign. Nevertheless, in the cases of 1970 and 1975 where all the estimated coefficients of the production function are positive and statistically significant, the labour coefficient has a reasonable value of 0.64 and 0.62, respectively.

The consistency of the estimates of the non-conventional approach can also be checked by computing the average values of the marginal products of factor inputs. The results are shown at the bottom of Table 5.1. The calculated average of the marginal products of current farm expenditures show a low of 0.17 to a high of 0.79, with an average value over the 7 estimates of 0.40 *yuan*, indicating that the average value of the marginal product of current inputs covered less than half of the marginal factor cost. This result is consistent with the observed excessive use of current farm inputs in Dahe during the period.

The marginal product of capital is the productive value per 1 *yuan* increase in the gross value of capital stock. The estimated average values range from a low of 0.04 to a high

[40]The model was also estimated without the restriction of constant returns to scale. The results are reported in Table 5.2. The estimated scale parameters range from 0.97 to 1.06, except for 0.87 in 1971. The hypothesis of constant returns to scale cannot be rejected at the 5% level of significance in any run. The rest of the discussion will be focused on the restricted model (8), because the estimation efficiency can be improved by imposing legitimate restrictions on the parameters.

of 0.28, with an average over the 3 estimates of 0.14. These rates are not much different from the rates of return to capital commonly found in the other less developed countries. The marginal product of land is the value of output attributable to an additional 1 *mu* of sown area. Its average values range from a low of 11.38 to a high of 16.07 *yuan*, with an average over the 3 estimates of 13.76 *yuan*. For comparison, the estimated marginal product of land for all of China from data at the provincial level in the period 1980-83 was 15.5 *yuan* (Putterman, 1987). Since capital and land markets did not exist in China, it is not clear whether the teams allocated their capital and land resources efficiently. The marginal product of labour is the productive value in *yuan* of an additional man-year at the full working intensity $\bar{w} = 1$. The estimated marginal product of labour is 236.9 *yuan* in 1970 and 317.6 *yuan* in 1975. The estimated marginal product of labour in the other cross-sectional regressions might be biased upwards due to the poor estimates for capital and land, and hence should be read with caution.

The estimates of disutility and the technical coefficient of monitoring appear quite satisfactory. Summarizing over all runs, these coefficents have the predicted signs and are significant at the 1% level. The results also look fairly stable across the periods. The estimated v lies between 3.88 and 7.86, and ω varies from 0.11 to 0.28. This suggests that the workers' tastes for working and monitoring their co-workers, as well as the monitoring technology, are rather invariant across production teams and over time, despite the variation displayed by the coefficients for production technology. The positive value of disutility supports the view that workers dislike working and monitoring their co-workers. However, compared with the average utility from annual consumption of 202.7, the disutility from work and monitoring appears fairly low. This may reflect the minor weight given by the peasants to leisure in relation to income in a poor nation.

We now turn to the tests of hypotheses. The first four hypotheses in Section 5.2 are formulated as nonlinear functions of the estimated parameters. The test statistics for these hypotheses are calculated at the mean values of the variables involved. The test statistic in each case has an asymptotic chi-squared distribution with 1 degree of freedom. The test statistics reported in Table 5.4 and Table 5.5 are the values of the standard normal variables. (See appendix 5.2 for details.) [41]

Let us first discuss the results reported in Table 5.4. The workers' motive to shirk in production is confirmed by the positive value of all the test statistics for the free-riding problem except the one for 1975; 5 of the tests are significant at the 5% level or better. The free-rider test for 1975 is statistically insignificant. The hypothesis that labour supervision is zero in the team is rejected at the 5% level of significance in all but one case. [42] Except for 1975, the estimated time fraction spent on production, α, is always less than unity, lying between 80 and 98 percent of the team's total working time. This result indicates that the production teams allocated 2 to 20 percent of their working time to labour supervision in order to prevent workers from shirking. [43] [44] With the exception of 1975, the estimated

[41]The calculation of the test statistics for most of these hypotheses involves the mean value of the dependent variable, q. The distribution of the test statistic is derived based on the assumption that the mean value of per capita income is a constant number. This is admittedly not a plausible assumption. Nonetheless, it is worth noting that in consumer demand analysis the estimates of elasticities are often presented as functions of the dependent variables such as budget shares. It is rather common to obtain the approximate test statistics of elasticities under the assumption that the budget shares are constant and equal to their sample means (Binswanger, 1974 and Pindyck, 1979).

[42]The null hypothesis of $\alpha = 1$ places the test on a boundary of the parameter space. Generally speaking, the test statistic formulated from the estimation of the unrestricted model in this case has a mixture of chi-squared distribution, which is different from the $\chi^2(1)$ distribution under the null hypothesis (Godfrey, 1987). The test of the hypothesis that $\alpha = 1$, however, used the conventional $\chi^2(1)$ distribution as a rough approximation to the unknown distribution of the test statistic. Specifying the exact distribution of the test statistic on a boundary of the parameter space is beyond the scope of this thesis.

[43]The test of the hypothesis that $\alpha = 1$ is also carried out for each individual estimated $\hat{\alpha}_i$, (see Table 5.4). The rejection rate of the hypothesis that $\alpha = 1$ against $\alpha < 1$ varies from 53% to 79% of the total $\hat{\alpha}_i$s for each cross-section, with an average of 64.1%. It confirms the reliability of the test results for $\alpha = 1$ based on the mean values of the variables involved. Nevertheless, we are not able to ensure that every individual $\hat{\alpha}_i$ falls in the range from zero to one. About 8% to 36% of the calculated $\hat{\alpha}_i$ exceed one, with an average of 23.8%. This problem can be handled by a nonlinear programming approach which would constrain the estimate to be less than one. The lack of knowledge of the statistical properties of the resultant estimates makes this approach problematic.

[44]It has been found in consumer demand analysis that regularity conditiona are often satisfied at a rather low proportion of the sample points (Howe, Pollak and Wales, 1979). The high failure rate of regularity condition tests opens an avenue to the improvement of model specification and the refinement of the data in demand analysis.

probability of catching and penalizing a shirker who provides zero effort consistently has the correct sign and is significant at the 10% level or better. The estimated chance of a shirker being caught and penalized in a team is 1 to 4 percent. These results suggest that labour monitoring in team farming is not technically impossible, but is nonetheless imperfect and costly.

Concerning the negativity of the first derivative of the credibility expression $Q(\alpha)$ with respect to α, the estimated test values in all runs except for 1975 have the expected signs; in five years, the test statistic is significant at the 10% level or better. The test with a "wrong" sign is insignificant. This result provides some empirical evidence for the economic rationality of production teams in China.

The test for the comparative static results are reported in Table 5.5. Except for the estimates of $\frac{\partial\alpha}{\partial d}$ and $\frac{\partial\alpha}{\partial\gamma}$ in 1975, all the estimates of the partial derivatives have the correct signs. Four of $\frac{\partial\alpha}{\partial\lambda_0}$ and $\frac{\partial\alpha}{\partial n}$, and five of $\frac{\partial\alpha}{\partial d}$, $\frac{\partial\alpha}{\partial\gamma}$, and $\frac{\partial\alpha}{\partial v}$, and seven of $\frac{\partial\alpha}{\partial\omega}$ are significant at the 10% level or better. The two cases of 'wrong' signs are not statistically significant.

With respect to magnitudes, the average elasticities of the time fraction spent on production α with respect to various variables are calculated and shown in Table 4. [45] The results are interpreted as follows. Other things being equal, a 1% increase in the price of output or production efficiency leads to a 0.57% to 3.52% of increase in working time spent on production (an average of 2.02%); a 1% rise in fixed cost per worker forces the team to increase monitoring time by 0.13% to 2.04% (an average of 0.86%); and a 1% increase in proportion of income distribution based on work contribution reduces monitoring time by 0.10% to 1.18% (an average of 0.59%). The strong negative effects on labour productivity of the extractive price policy, excessive state-imposed fixed costs, and anti-incentive ideology in China's production teams is broadly evident in these results. Moreover, a 1% increase in the disutility of production and monitoring requires the team to raise monitoring time by 0.55% to 2.41% (an average of 1.29%); a 1% rise in the technical efficiency of monitoring, ω, allows the

[45] Only those with significant test statistics are reported.

team to reduce monitoring time by 1.00% to 3.52% (an average of 2.12%); and a 1% increase in team size reduces time spent on production by 0.58% to 2.98% (an average of 1.36%). Clearly, the disutility of work, tastes for engaging in mutual monitoring, the team's managerial ability, the technological characteristics of the production process and the size of the team also play important roles in an effective mutual monitoring s...

5.4 Conclusion

In this chapter, a rich panel data set from a commune in China during the Cultural Revolution period was used to examine the static mutual monitoring model. An implicit function determining the team's optimal allocation of labour time between production and monitoring was estimated by the nonlinear two-stage least squares method. The estimated results appear quite satisfactory, providing empirical support for the highly abstract theoretical model of static monitoring in Chinese rural teams developed in chapter 4.

By estimating a production function in an incentive context, we calculated the labour cost that the teams spent on labour supervision. It was found that 2 to 20 percent of potential production time was taxed away by monitoring requirements. By accounting for such monitoring costs, we corrected a common measurement error where number of workers is used as a proxy for labour input in the estimation of a production function. Allocative efficiency and its implications for team management were also examined. The results show that due to the state's irrational production strategy and price distortions, the financial return to land was fairly low, and the return to current farm inputs covered less than half of their cost. This resource misallocation itself could create incentive problems for the team even in a world of perfect information. It surely aggravated the difficulties facing teams in the area of labour supervision.

Several hypotheses concerning work incentives were tested. The results support the conventional belief that Chinese collective farmers faced a free-rider peoblem. It was also shown that labour supervision in group farming is not technically impossible, but is nonetheless imperfect and costly. It was further demonstrated that Chinese peasants in the various teams were rational economic agents. Their practices were sensible responses to prevailing internal incentive problems, given the constraints they faced. However, under the extractive policy regime, state bureaucratic controls and anti-incentive ideology, labour monitoring requirements taxed away a significant proportion of the team's productive resources.

Based on the findings of this chapter, it is plausible to argue that much of the difficulty of labour supervision in group farming did not primarily stem from the intrinsic technological characteristics of farm production. There were indeed agency costs associated with team farming. Yet the agency costs incurred by Chinese teams would have been smaller if the Chinese government had consistently adopted sound agricultural policies, provided correct price signals to the teams and encouraged material incentives for individual peasants. If the dynamic benefits of cooperative production for China's rural development discussed in chapter 2 could offset its monitoring cost, then collective farms need not to be inferior to private household farms.

Apart from the contribution made to the incentive literature, this model raises a number of important issues in the field of applied econometrics. These issues are: (1) how to specify the exact distribution of the test statistics on a boundary of the parameter space? (2) how to test the accuracy of functional form in a highly nonlinear estimating equation? (3) what will happen to the empirical results if we switch from the current econometric approach to a nonlinear programming approach? All these issues deserve to be investigated in future research. Nonetheless, most of the results reported here are plausible in sign and magnitude, and seem unlikely to change substantially even with such refinements in the estimation procedure used.

Table 5.1
THE NL2S ESTIMATION RESULTS
(Constant Returns to Scale)

	1970	1971	1972	1973	1974	1975	1976
λ_0	149.6 (6.49)a	344.7 (3.87)a	201.6 (1.41)c	159.6 (2.33)a	247.9 (5.45)a	122.4 (5.96)a	177.2 (2.82)a
λ_1	0.64 (22.83)a	0.82 (18.64)a	0.84 (14.00)a	0.74 (6.17)a	0.83 (17.57)a	0.62 (15.90)a	0.80 (12.12)a
λ_2	0.04 (1.96)b	-0.14 (-1.66)c	-0.01 (-0.14)	0.02 (0.35)	-0.01 (-0.30)	0.14 (4.97)a	0.02 (1.79)b
λ_3	0.24 (7.14)a	0.18 (3.35)a	-0.04 (-0.64)	0.06 (0.75)	0.04 (0.91)	0.16 (6.33)a	0.01 (0.68)
λ_4	0.08 (1.71)b	0.14 (2.39)a	0.21 (2.53)a	0.18 (4.57)a	0.14 (2.85)a	0.08 (2.59)a	0.17 (2.43)a
v	3.87 (6.37)a	7.86 (6.20)a	7.60 (4.84)a	6.12 (4.96)a	5.77 (11.05)a	3.77 (10.54)a	5.64 (9.51)a
ω	0.13 (2.33)a	0.21 (3.27)a	0.25 (3.98)a	0.28 (3.60)a	0.22 (2.92)a	0.11 (4.30)a	0.19 (4.46)a
KH	(2.67)	(11.47)	(3.92)	(17.97)b	(5.22)	(6.62)	(9.01)
MPK	0.10	---	---	---	---	0.28	0.04
MPT	13.84	11.38	---	---	---	16.07	--
MPF	0.26	0.45	0.79	0.45	0.36	0.17	0.34
MPL	236.9	314.8	380.6	326.7	419.3	317.6	393.4
APL	256.4	263.8	286.8	266.0	306.3	277.1	248.8

Notes: (1) This set of the NL2S estimation results is obtained with the use of γ, n, k, t, f, and d as the instrumental variables.
(2) T statistics are reported in paretheses below the coefficient estimates.
(3) c denotes that the test is significant at the 10% level.
(4) b denotes that ....at the 5% level.
(5) a denotes that ....at the 1% level.
(6) KH is the Koenker heteroskedasticity test value.
(7) MP and AP stands for marginal and average product, respectively.
(8) λ_1 is calculated as $1 - \Sigma\lambda_r$, r=2,3,4.

Table 5.2
THE NL2S ESTIMATION RESULTS
(Unrestricted Model)

	1970	1971	1972	1973	1974	1975	1976
λ_0	157.8	602.4	111.7	207.5	293.9	71.3	193.2
	(6.37)a	(1.04)	(1.73)b	(1.68)b	(5.40)a	(1.65)b	(1.90)b
λ_1	0.55	0.71	0.83	0.77	0.84	0.70	0.70
	(6.75)a	(4.70)a	(9.70)a	(5.72)a	(4.57)a	(11.26)a	(12.49)a
λ_2	0.06	-0.17	-0.04	-0.01	-0.02	0.18	0.05
	(2.24)b	(-1.97)b	(-0.86)	(-0.16)	(-0.38)	(5.82)a	(3.60)a
λ_3	0.29	0.14	-0.03	0.02	0.04	0.07	0.07
	(8.81)a	(2.88)a	(-0.56)	(0.23)	(0.44)	(1.49)c	(2.62)a
λ_4	0.07	0.19	0.21	0.20	0.13	0.11	0.14
	(1.70)b	(2.92)a	(2.89)a	(7.56)a	(3.40)a	(1.34)c	(2.25)b
v	3.45	7.93	7.42	6.66	5.86	4.49	4.53
	(7.21)a	(10.85)a	(6.96)a	(3.90)a	(3.27)a	(5.49)a	(6.41)a
ω	0.14	0.27	0.25	0.37	0.24	0.15	0.19
	(2.05)b	(2.41)a	(2.73)a	(1.82)b	(2.51)a	(3.10)a	(2.81)a
RTS	0.97	0.87	1.06	0.97	0.99	1.05	0.96
$\chi^2(1)$	(0.62)	(0.59)	(0.64)	(0.17)	(0.07)	(0.64)	(0.40)
α(mean)	0.88	0.81	0.89	0.87	0.98	1.04	0.93

Notes: (1) c denotes that the test is significant at the 10% level.
(2) b denotes that ....at the 5% level.
(3) a denotes that .... at the 1% level.
(4) RTS is the estimated scale parameter and the χ^2 test is given under each estimated value.

Table 5.3
THE OLS ESTIMATION RESULTS
(Constant Returns to Scale)

	1970	1971	1972	1973	1974	1975	1976
$\text{Log}\lambda_0$	4.44 (9.33)a	3.96 (10.98)a	2.36 (3.41)a	3.44 (9.99)a	3.20 (8.13)a	1.98 (6.76)a	2.11 (8.24)a
λ_1	0.59 (22.69)a	0.38 (17.27)a	0.28 (10.77)a	0.32 (21.33)a	0.40 (22.22)a	0.20 (16.67)a	0.27 (30.00)a
λ_2	0.11 (1.51)c	0.05 (0.84)	0.24 (2.95)a	-0.03 (-0.48)	-0.01 (-0.23)	0.06 (1.33)	0.05 (1.47)c
λ_3	0.16 (1.92)b	0.33 (3.63)a	0.002 (0.02)	0.25 (3.62)a	0.04 (0.52)	0.04 (0.58)	-0.02 (-0.32)
λ_4	0.14 (1.12)	0.24 (2.98)a	0.48 (4.19)a	0.46 (7.60)a	0.57 (6.05)a	0.70 (10.26)a	0.70 (12.73)a
KH	(3.81)	(6.59)	(13.00)	(3.97)	(1.60)	(6.80)	(1.76)
R^2	0.23	0.52	0.44	0.72	0.59	0.80	0.80
MPK	0.27	---	0.49	---	---	---	0.09
MPT	9.26	20.90	---	19.30	---	---	--
MPF	---	0.77	1.31	1.16	1.45	1.53	1.42
MPL	218.4	145.9	126.8	141.3	202.1	102.4	132.8

Notes: (1) This table presents the OLS estimation results of the production function in a conventional log-linear form.
(2) T statistics are reported in parentheses below the coefficient estimates.
(3) c denotes that the test is significant at the 10% level.
(4) b denotes that ....at the 5% level.
(5) a denotes that ....at the 1% level.
(6) KH is the Koenker heteroskedasticity test value.
(7) MP stands for marginal product. Only those with significant test statistics are calculated.
(8) λ_1 is calculated as $1 - \Sigma\lambda_r$, r=2,3,4.

Table 5.4
RESULTS OF THE HYPOTHESIS-TESTING

	1970	1971	1972	1973	1974	1975	1976
H_1	0.45	2.56	1.44	1.19	0.38	-0.09	0.88
(z)	(2.84)a	(3.45)a	(2.50)a	(1.89)b	(1.21)	(-0.31)	(3.29)a
α(mean)	0.85	0.80	0.89	0.86	0.98	1.00	0.93
H_2	0.15	0.20	0.11	0.14	0.02	-0.0004	0.07
(z)	(9.76)a	(7.01)a	(11.17)a	(11.34)a	(1.84)b	(-0.03)	(10.10)a
No.1	20	29	27	33	33	44	49
%	53	78	70	70	57	56	65
No.2	8	5	3	5	8	6	5
%	21	14	9	11	14	8	7
No.3	10	3	8	9	17	28	21
%	26	8	21	19	29	36	28
H_3	0.02	0.04	0.03	0.04	0.01	-0.0001	0.01
(z)	(5.85)a	(3.49)a	(4.68)a	(3.23)a	(1.45)c	(-0.03)	(4.29)a
H_4	-92.4	-216.4	-338.4	-121.9	-4361.2	4028300	-816.52
(z)	(-1.40)c	(-3.68)a	(-3.78)a	(-1.32)c	(-0.64)	(0.02)	(-5.04)a

Notes: (1) H_1 stands for the test of free-riding in production.
(2) H_2 is the test for the hypothesis that $1-\alpha$ is equal to zero with the use of the mean values of the variables involved.
(3) No.1 is the number of teams having $\alpha < 1$ at the 5% level of significance.
(4) No.2 is the number of teams having $\alpha = 1$ at the 5% level of significance.
(5) No.3 is the number of teams having $\alpha > 1$ at the 5% level of significance.
(6) H_3 is the estimated value of the probability that a shirker is caught shirking and penalized in the team.
(7) H_4 is the test for the negativity of the first derivative of the credibility expression $Q(\alpha)$ with respect to α.

Table 5.5
COMPARATIVE STATIC TESTS AND ELASTICITIES

	1970	1971	1972	1973	1974	1975	1976
P_{λ_0}	0.02	0.004	0.01	0.02	0.001	0.0002	0.003
(z)	(1.43)c	(6.11)a	(1.60)c	(1.00)	(0.81)	(0.03)	(2.16)b
P_d	-0.01	-0.002	-0.002	-0.004	-0.0001	0.00	-0.0004
(z)	(-1.40)c	(-3.68)a	(-3.78)a	(-1.32)c	(-0.64)a	(0.02)	(-5.04)b
P_γ	2.09	0.87	0.64	1.60	0.64	-0.0001	0.23
(z)	(1.40)	(3.68)a	(3.77)a	(1.31)c	(0.64)	(-0.02)	(5.05)a
P_n	-0.03	-0.01	-0.01	-0.02	-0.004	-0.0003	-0.006
(z)	(-1.52)c	(-7.60)a	(-2.63)a	(-1.28)	(-0.93)	(-0.03)	(-5.48)a
P_v	-0.53	-0.11	-0.10	-0.22	-0.05	-0.005	-0.09
(z)	(-1.55)c	(-6.30)a	(-2.63)a	(-1.32)c	(-0.93)	(-0.03)	(-5.23)a
P_ω	22.99	8.86	7.27	9.80	5.67	9.08	7.46
(z)	(2.23)b	(5.35)a	(3.77)a	(1.91)b	(2.62)a	(2.10)b	(7.28)a
E_{λ_0}	3.52	1.72	2.27	---	---	---	0.57
E_d	-2.04	-0.48	-0.52	-1.12	---	---	-0.13
E_γ	1.18	0.51	0.32	0.82	---	---	0.10
E_n	-2.98	-1.06	-0.81	---	---	---	-0.58
E_v	-2.41	-1.08	-0.85	-1.57	---	---	-0.55
E_ω	3.52	2.33	2.04	3.19	1.27	1.00	1.52
sample size	38	37	38	47	58	78	75

Notes: (1)P_{λ_0}, P_d, P_γ, P_n, P_v, P_ω are the estimates of the partial derivatives of α with respect to the corresponding parameters.

(2)E_{λ_0}, E_d, E_γ, E_n, E_v, and E_ω are the elasticities of the time fraction spent on production with respect to λ_0, d, γ, n, v, and ω respectively.

Table 5.6
THE NL2S ESTIMATES AND HAUSMAN TESTS

	1970	1971	1972	1973	1974	1975	1976
λ_0	145.1 (9.00)a	334.7 (2.97)a	316.9 (0.84)	145.5 (4.06)a	252.2 (3.92)a	113.2 (5.87)a	161.2 (3.36)a
λ_2	0.04 (1.75)b	-0.13 (-2.18)b	-0.07 (-0.46)	0.03 (0.75)	-0.01 (-0.32)	0.14 (5.27)a	0.02 (2.22)b
λ_3	0.24 (7.61)a	0.16 (2.45)a	-0.02 (-0.26)	0.06 (1.13)	0.04 (0.69)	0.16 (6.95)a	0.01 (0.45)
λ_4	0.08 (2.19)b	0.14 (1.37)c	0.17 (1.54)c	0.19 (8.92)a	0.14 (1.61)c	0.08 (3.36)a	0.19 (3.25)a
v	4.26 (6.93)a	7.49 (8.44)a	8.42 (3.79)a	5.50 (4.20)a	6.40 (11.22)a	3.72 (8.37)a	5.57 (9.93)a
ω	0.20 (1.96)b	0.18 (3.58)a	0.26 (3.43)a	0.25 (2.86)a	0.28 (2.75)a	0.12 (4.18)a	0.18 (4.76)a
H	(-0.50)	(0.03)	(0.30)	(0.09)	(7.15)	(0.80)	(3.67)

Notes: (1) This set of the NL2S estimation results is obtained with the use of g, n, k, t, f, d as the instrumental variables.
(2) H is the Hausman test.

Chapter 6

A Dynamic Mutual Monitoring Model

6.1 Introduction

In this chapter we will extend the static mutual monitoring system developed in chapter 4 to a production team with a long time horizon. The dynamic mutual monitoring system is based on the conjecture that labour effort in a production team has two attributes: it is perfectly observable by every team member, but not verifiable at zero cost.

It is assumed that the workers engage in formal mutual monitoring as we described in Chapter 4. Meanwhile, the workers conduct informal mutual monitoring, meaning that the workers accumulate information about the other workers' labour performance through the production process. Informal mutual monitoring is a joint activity with production. Through team production, every worker knows how all the others have performed in production and monitoring. Information gathered informally cannot be used to impose formal penalties, because the imposition of formal punishments must be based on verifiable evidence. Nevertheless, due to the existence of long-term relationships, the workers are able to punish shirkers using their non-verified information. With the use of additional information at zero cost, the dynamic incentive system will supplement the static monitoring system and improve the efficacy of labour supervision in team production.

The agenda of chapter 6 is outlined as follows. Section 2 describes the characteristics of individual workers' effort decisions in a single period game. For easy reference, we provide a brief introduction of the static monitoring system constructed in chapter 4. Some modifications in the basic assumptions are made in order to keep the model tractable. A description of subgame perfect equilibria is contained in section 3. The nature of an optimal labour contract in a dynamic setting is discussed in section 4. Among the key issues addressed in this section are the implications, existence and uniqueness of an optimal contract and the comparative static results. The final section of the chapter summarizes the results.

6.2 The Single-Period Game

Define the single period game to be $\Gamma = (N, W, U)$. $N \equiv \{1,....,n\}$ is the set of n identical workers, which is finite, and $n \geq 2$. Denote the strategy of the ith worker by w_i. $w_i \in [0,1]$ is the effort level of worker i. To keep the model tractable, I simplify the framework used in chapter 4 by assuming that a worker provides the same level of effort in both production and monitoring. A vector of strategies, one for each worker, is denoted $w = (w_1,....,w_n) \in W$. Defining $w_{-i} \equiv (w_1,...,w_{i-1},w_{i+1},...,w_n)$, where w_{-i} consists of the strategy choice of all workers except the ith, one can write $w = (w_i,w_{-i})$.

Focusing on symmetric situations, we write w as the common level of effort and z as the effort level of a worker who deviates from w. The production function of a team when n-1 workers supply w, and one worker deviates to z is then defined as

$$(1) \qquad Q = \Theta\alpha^{\beta}[w(n-1)+z]^{\beta}$$

where Q is gross output in value terms. Θ is the technical coefficient of labour input. $\alpha \in [0,1]$ is the time fraction spent on production, and $(1-\alpha)$ is the time fraction spent on formal monitoring. $\alpha[w(n-1)+z]$ stands for total input of productive labour. The parameter β, where $0 < \beta < 1$, is the elasticity of output with respect to labour input. For a given α, the team's net revenue available for distribution is

$$(2) \qquad Y(\alpha,w,z) = Q - D \geq 0,$$

where D is the team's fixed cost. It is assumed that when $Q < D$, the team will receive transfer payments from the government to cover its fixed cost. Hence, the distributable income is non-negative. [44] A worker is assumed to hold an equal share of the net revenue if he/she has not been formally accused of shirking. Worker i's income is thus defined as

[44] In reality, Chinese teams received the state's poverty-relief funds when they were unable to provide a subsistence level of food consumption for peasants.

$$(3) \qquad y_i = \frac{1}{n}Y(\alpha,w,z) - T_iC.$$

T_i is a binary variable; $T_i = 1$ if worker i is caught shirking and $T_i = 0$ otherwise. The probability that worker i is punished for shirking in production and monitoring is

$$(4) \qquad P_i(w,z) \equiv w(1-z)\Phi(\alpha),$$

where $\Phi(\alpha)$ is the technical efficiency of monitoring. Equation (4) assumes that all other workers maintain the effort w, while worker i offers z. C is the penalty imposed on a shirker. The size of C is limited by some ethical and ideological factors; the team has to allocate the proportion $(1-\gamma)$ of the distributable income to ensure workers' basic needs as explained in chapter 4 and 5. This imposes a credibility constraint on the penalty C, i.e.

(5) CRED: (a) if $Y(\alpha,w,0) < 0$, then $C = 0$

(b) otherwise, $\frac{\gamma}{n}Y(\alpha,w,0) \geq C \geq 0$,

where γ is the proportion of the income distributed according to performance. $Y(\alpha,w,0)$ is the team net income when a single worker deviates from w to $z = 0$, while all others maintain w. CRED states that a penalty cannot be imposed if it is possible for a worker to make team income negative through deviation. In general, a penalty is feasible only when team income is positive for all possible deviations. Since $Y(\alpha,w,z)$ is increasing in z, if a penalty is feasible at $z = 0$, while all the other workers provide the level of effort w, then it will be feasible for all $z \in [0,1]$.

Given the income distribution structure defined above, the expected utility for worker i when choosing effort z is

$$(6) \qquad U_i(z,w) = \begin{cases} \frac{1}{n}Y(\alpha,w,z) - \Phi(\alpha)w(1-z)C - zv & \text{if } Y(\alpha,w,z) \geq 0 \\ -zv & \text{otherwise.} \end{cases}$$

The disutility of production and monitoring are assumed to be identical, i.e $v_1 = v_2 = v$. Denoting the optimal deviation from w at given α and C as

(7) $$\hat{z} = \hat{z}(\alpha, w, C) \equiv \operatorname{argmax} U_i(z,w),$$

we then define the optimal payoff for the worker who deviates from w as

(8) $$\hat{U}_i(w) \equiv U_i(\hat{z},w) \equiv \max_{z \epsilon [0,1]} U_i(z,w).$$

Note that

(9) $$\tilde{U}_i(w) \equiv U_i(w,w) ,$$

where $\tilde{U}_i(w)$ is the expected utility achieved through compliance with the team effort contract, i. e. when $z = w$. $U = (U_1,....,U_n)$ is the payoff vector.

It is well known that production teams generally face free-riding problems in a single-period game. The following assumptions are made to specify this problem. First, define the minimum time fraction spent on production as α_{min}, using the identity

(10) $$\Theta(\alpha_{min}(n-1))^{\beta} \equiv D,$$

where $0 < \alpha_{min} < 1$. The first inequality follows from $\Theta > 0$ and $D > 0$. The second holds if the team is capable of imposing a positive penalty when a shirker offers zero effort, and all the other workers exert maximum effort $w = 1$. Corresponding to $\alpha \epsilon [\alpha_{min},1]$, the effort level w^{α}_{min} is defined by

(11) $$\Theta(\alpha(n-1)w^{\alpha}_{min})^{\beta} \equiv D.$$

Again $w^{\alpha}_{min} \in (0,1)$ for $\alpha \in [\alpha_{min},1]$ whenever $\Theta > 0$, $D > 0$ and $C > 0$. Note that part (b) of CRED in (5) holds only if $\alpha \in [\alpha_{min},1]$ and $w \in [w^{\alpha}_{min},1]$. Moreover, whenever $C > 0$ is feasible for $\alpha \in (\alpha_{min},1]$ and $w \in (w^{\alpha}_{min},1]$, the relationship among the parameters Θ, D, n, and β,

A1: $\Theta(n-1)^{\beta} > D$

is assumed.

It is assumed throughout that the leisure cost of a worker does not exceed the fixed cost per worker, $d \equiv D/n$, weighted by the output elasticity of labour, that is

A2: $\beta D/n \geq v$.

A2 indicates that increased effort is always socially desirable. A2 can be justified by the fact that peasants with a low income may put a minor weight on leisure in relation to cash income. Given A2, $\bar{U}_{iw}(w) \geq 0$ for all $\alpha \in [\alpha_{min},1]$, $w \in [w^{\alpha}_{min},1]$, and $C = 0$, where $\bar{U}_{iw}$ is the derivative of $\bar{U}_i$ with respect to a common effort level w. Since the marginal product of effort exceeds its disutility for the team as a whole, a team which could contract directly on effort would always set $w = 1$.

Nonetheless, the socially desirable effort level usually cannot be sustained in the absence of monitoring and penalties in a one-shot game. It is assumed that

A3: $v \geq \frac{\beta}{n}\Theta^{1/\beta}D^{1-1/\beta}$.

Given A3, it can be shown that $\alpha = 1$, and/or $C = 0$ implies $U_{iz}(z,w) < 0$ for all $w^{\alpha}_{min} \leq w \leq 1$, where U_{iz} is the derivative of U_i with respect to z when the other workers' effort level is taken as given. $U_{iz}(z,w) < 0$ indicates that increased effort is perceived by individual workers to be undesirable when they make their own effort decisions in isolation. Note that A2 and A3 together imply that $D \geq \Theta$.

The Nash equilibria in a single-period game are summarized in the following theorems.

Theorem 1: Given A3, shirking in production and monitoring $\tilde{w} = 0$ is a Nash equilibrium of Γ. A typical worker in this equilibrium obtains the payoff

$$(12) \qquad \bar{U}_i(\tilde{w}) = 0.$$

Proof: If all the other workers provide an effort level $w \in [0, w^{\alpha}_{min})$, then $C = 0$ due to part (a) of CRED in (5). Worker i cannot do better by offering $w \in [w^{\alpha}_{min}, 1]$ given A3. Moreover, an effort level $w \in [0, w^{\alpha}_{min}]$ always generates zero net income due to the state transfer payments, while the effort cost is minimized at $\tilde{w} = 0$ over the range $[0, w^{\alpha}_{min})$. Hence, when $w_{-i} \in [0, w^{\alpha}_{min})$ is anticipated, the best reply of worker i is to provide zero effort, i.e. $w_i = 0$, and so does every worker. The payoff follows a direct calculation.

Q. E. D.

Theorem 2: $\bar{w} \in [w^{\alpha}_{min}, 1]$ forms an equilibrium of Γ if and only if the following conditions hold for all $i \in N$:

$$(13)\ (IC): \beta\frac{\Theta}{n}\alpha^{\beta}(n\bar{w})^{\beta-1} + \bar{w}\Phi(\alpha)C - v \geq 0 \quad \text{if } \bar{w} = 1$$
$$= 0 \quad \text{if } w^{\alpha}_{min} < \bar{w} < 1,$$

and

$$(14)\ (CRED): \frac{\gamma}{n}Y(\alpha, w, 0) \geq C \geq 0.$$

A representative worker in such an equilibrium obtains the payoff

$$(15) \qquad \bar{U}_i(\bar{w}) = \frac{\Theta}{n}(\alpha n\bar{w})^{\beta} - d - \bar{w}(1-\bar{w})\Phi(\alpha)C - \bar{w}v.$$

Proof: Follows from the definition of Nash equilibrium. (14), i.e. part (b) of CRED, is not a requirement for Nash equilibrium, yet it must hold if the penalty is to be enforceable when some worker unilaterally shirks. A direct computation gives the utility of worker i. Q. E. D.

Theorem 3: An equilibrium with a credible penalty C and $0 < \alpha < 1$ which has $\bar{w} \epsilon [w^{\alpha}_{min}, 1]$ Pareto dominates the shirking equilibrium $\tilde{w} = 0$.

Proof: $\bar{U}_i(\bar{w}) > \bar{U}_i(\tilde{w})$ if and only if $\bar{U}_i(\bar{w})$ is greater than zero. It is easy to show that $\bar{U}_i(\bar{w})$ is always positive when the static incentive compatibility condition in (13) and the credibility constraint in (14) hold. Assuming $\alpha < 1$, (13) and (14) imply

$$(16) \qquad \gamma[\frac{\Theta}{n}(\alpha\bar{w}(n-1))^{\beta} - d] - [1/(\bar{w}\Phi(\alpha))][v - \frac{\beta\Theta}{n}\alpha^{\beta}(n\bar{w})^{\beta-1}] \geq 0.$$

Denoting (16) by $G(\bar{w}) \geq 0$, we can write

$$(17) \qquad \bar{U}_i(\bar{w}) - G(\bar{w}) = \frac{\Theta}{n}(\alpha\bar{w})^{\beta}[n^{\beta} - (n-1)^{\beta} - \beta n^{\beta-1}] + [1/(\bar{w}\Phi(\alpha)) - 1][v - \frac{\beta\Theta}{n}\alpha^{\beta}(n\bar{w})^{\beta-1}] + (1-\gamma)[\frac{\Theta}{n}(\alpha\bar{w}(n-1))^{\beta} - d].$$

Note that (17) holds for both $\bar{w} < 1$ and $\bar{w} = 1$. When $\bar{w} < 1$, IC in (13) holds with equality, so that substituting (13) into the objective function and CRED (14) gives us (17). When $\bar{w} = 1$, $\bar{U}(1) = \frac{\Theta}{n}\alpha^{\beta}n^{\beta} - d - v$. (17) is obtained by substituting the identity

$$(18) \qquad \frac{1}{\Phi(\alpha)}[v - \frac{\beta\Theta}{n}\alpha^{\beta}n^{\beta-1}] \equiv [v - \frac{\beta\Theta}{n}\alpha^{\beta}n^{\beta-1}] + [\frac{1}{\Phi(\alpha)} - 1][v - \frac{\beta\Theta}{n}\alpha^{\beta}n^{\beta-1}]$$

into G(1). We can show that the expression in (17) is strictly positive. The first term of (17) is positive, because $[n^{\beta} - (n-1)^{\beta} - \beta n^{\beta-1}] > 0$ due to strict concavity of n^{β} (see the proof below). The second and third terms are both nonnegative, using A3 and CRED. Hence, $\bar{U}_i(\bar{w}) > \bar{U}_i(\tilde{w})$. Q. E. D.

The proof of $n^{\beta} - (n-1)^{\beta} - \beta n^{\beta-1} > 0$ is given below:

Proof: Draw a tangent line to n^{β} at n_0. Due to strict concavity, the tangent line at n_0 is strictly above the curve n^{β} for all $n \neq n_0$. The tangent line at n_0 can be defined as $y = \beta n_0^{\beta-1} n + (1-\beta) n_0^{\beta}$. Now set $n = n_0 - 1$. Since the tangent line is strictly above the curve for all $n \neq n_0$, this must be true for $n = n_0 - 1$. Therefore,

$$\beta n_0^{\beta-1}(n_0-1) + (1-\beta) n_0^{\beta} > (n_0 - 1)^{\beta}.$$

Rearranging this, we have

$$n_0^{\beta} - \beta n_0^{\beta-1} - (n_0 - 1)^{\beta} > 0.$$

But since n_0 was arbitrary, we conclude

$$n^{\beta} - \beta n^{\beta-1} - (n - 1)^{\beta} > 0 \text{ for all } n.$$

Q. E. D.

Theorem 3 shows that it pays for the team to implement a formal mutual monitoring system when such a system is feasible, and no institutional alternatives are available to the workers.

Individual workers are always tempted to deviate from the optimal level of effort when the static incentive compatibility condition or the credibility constraint is violated. It is obvious that the outcome for an individual worker who chooses an optimal deviation from the level of effort $\bar{w}$ is always at least as good as the one with $\bar{w}$, i.e. $\hat{U}_i(\bar{w}) \geq \bar{U}_i(\bar{w})$. This follows from the maximization of $U_i(z,\bar{w})$ with respect to z. Since $\hat{z} = \bar{w}$ is feasible when (13) and (14) hold, $\hat{U}_i(\bar{w})$ always dominates $\bar{U}_i(\bar{w})$.

6.3 Characterization of Subgame Perfect Equilibria

Define a repeated game of team production to be $T=(N, \Sigma, G)$. N is the set of workers in the team. Denote the action of the ith worker in period t by w_{it}; $w_{it} \in [0,1]$ is the effort level of worker i at period t. A vector of actions, one for each player at period t is denoted $w_t = (w_{1t}, \ldots\ldots, w_{nt})$. A history is defined as $h_t = (w_0, w_1, \ldots\ldots, w_{t-1}) \in H_t$ specifying all actions prior to period t. A strategy for i in period t is a map from his/her information set to a chosen action i.e.

$$\sigma_{it}: H_t \rightarrow [0,1] \text{ for each worker,}$$

where $\sigma_{it}(h_t) = w_{it}$. A strategy sequence for worker i is a collection of functions,

$$\sigma_i = (\sigma_{i0}, \sigma_{i1}, \ldots, \sigma_{it}, \ldots) \in \Sigma_{i0} \times \Sigma_{i1} \times \ldots \times \Sigma_{it} \times \ldots = \Sigma_i.$$

and a strategy combination is a vector of strategies

$$\sigma = (\sigma_1, \ldots, \sigma_n) \in \Sigma_1 \times \ldots \times \Sigma_n = \Sigma.$$

σ can be written as $\sigma = (\sigma_i, \sigma_{-i})$, where $\sigma_{-i} = (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_n)$ are the strategies of the other players. The expected lifetime utility for worker i at time t as a function of all workers' strategies will be

$$(19) \qquad G_{it}(\sigma | h_t) = \Sigma_{s=t}^{\infty} \delta^{s-t} U_i(w_s),$$

where δ is the discount factor, which reflects the probability that a worker will remain in the team in future and his/her intertemporal preference. Let $G = (G_1, \ldots, G_n)$ be the lifetime payoff vector at time $t = 0$.

A set of strategies is called a subgame perfect Nash equilibrium if the strategies of the players are best replies to one another at all times and for all possible histories.

Definition 1: A strategy combination $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*)$ is a subgame perfect Nash equilibrium (SPE) if and only if

$$G_{it}(\sigma^* | h_t) \geq G_{it}(\sigma_i, \sigma_{-i}^* | h_t) \text{ for all } \sigma_i \in \Sigma_i, \text{ and all } h_t \in H_t, \text{ all } t \geq 0, \text{ and all } i \in N.$$

Theorem 4: $\bar{\sigma}_{it}(h_t) = \bar{w}$ for all i, h_t, and t, is a subgame perfect Nash equilibrium in

$T=(N,\Sigma,G)$ if (13) and (14) hold.

Proof: Given the conditions in (13) and (14), $\bar{w}$ is a Nash equilibrium in a one-shot game. It is well known that a strategy combination which plays one-shot Nash actions in every period for any history is a subgame perfect equilibrium of the supergame (Friedman, 1986).Q. E. D.

Theorem 5: Shirking in production and monitoring $\tilde{\sigma}_{it}(h_t)= \tilde{w} = 0$ for all i, h_t and t, is a subgame perfect Nash equilibrium in T.

Proof: Follows the same argument in Theorem 4. Q. E. D.

In the repeated game these need not be the sole subgame perfect Nash equilibria. It is well known that with the use of trigger strategies workers can achieve efficient cooperative outcomes that they are unable to obtain in the one-period game. [47]

Consider symmetric trigger strategies of the following form: for each i,

$$\hat{\sigma}_{it}(h_t) = \bar{w} \quad \text{if } t=0, \text{ or if } w_{jr} = \bar{w} \text{ for all } j \in N \text{ and } r < t \text{ if } t \geq 1;$$
$$= \tilde{w} = 0 \quad \text{if } w_{jr} \neq \bar{w} \text{ for some } j \in N \text{ and } r < t.$$

Under $\hat{\sigma}$, workers will switch to the inefficient Nash equilibrium with zero effort to punish the shirker after any deviation from the efficient cooperative agreement. The threat to move to the inefficient Nash equilibrium is always credible because the repetition of $\tilde{w}$ is also a SPE of T.

Consider the lifetime expected utility functions in the repeated game when other workers follow the trigger strategies defined above. If worker i follows the trigger strategy, then according to (15) the lifetime utility for worker i at time t=0 is

[47]This chapter considers only trigger strategies rather than more complex supergame strategies. The results can be interpreted as a lower bound on the utility a team can generate for its members. If some more complex strategies were allowed, it is possible that the team might do better.

$$(20) \qquad G_i(\hat{\sigma}) = \Sigma_{t=0}^{\infty}\bar{U}_i(\bar{w})\delta^t = \frac{1}{(1-\delta)}\bar{U}_i(\bar{w})$$

$$= \frac{1}{1-\delta}\left[\frac{\Theta}{n}(n\alpha\bar{w})^{\beta} - d - \bar{w}v - (1-\bar{w})\bar{w}\Phi(\alpha)C\right].$$

If worker i decides to deviate from $\hat{\sigma}$ in period t, given that the other workers use the trigger strategies, his/her largest possible payoff from period t on will be

$$(21) \qquad G_i(\sigma_i,\hat{\sigma}_{-i}) = \hat{U}_i(\bar{w}) + \frac{\delta}{1-\delta}\tilde{U}_i(\bar{w})$$
$$= \frac{\Theta}{n}[(n-1)\bar{w} + \hat{z}]^{\beta}\alpha^{\beta} - d - \hat{z}v - \bar{w}(1-\hat{z})\Phi(\alpha)C$$

where $\hat{z}$ is the optimal one-shot deviation. Note that from the discussion of a single-period game, the payoffs are ranked as

$$\hat{U}_i(\bar{w}) \geq \bar{U}_i(\bar{w}) > \tilde{U}_i(\bar{w}) = 0.$$

Theorem 6: The trigger strategy $\hat{\sigma}$ is a subgame perfect Nash equilibrium point of T if and only if

$$(22) \qquad \bar{U}_i(\bar{w}) \geq (1-\delta)\ \hat{U}_i(\bar{w})$$

Proof. Let us first show that $\hat{\sigma}$ is an equilibrium point. The incentive compatibility condition (IC) in (22) implies that $G_{it}(\hat{\sigma}|h_t) \geq G_{it}(\sigma_i,\hat{\sigma}_{-i}|h_t)$ for all σ_i and all i. The trigger strategy of worker i is the best reply to the trigger strategies of others; therefore, the trigger strategy combination $\hat{\sigma}$ is an Nash equilibrium point. On the other hand, it can be shown that if $\hat{\sigma}$ is a SPE, then (22) must hold. If it did not, there would be subgames where it is optimal to deviate, i.e. $G_{it}(\hat{\sigma}|h_t) < G_{it}(\sigma_i,\hat{\sigma}_{-i}|h_t)$. Hence, (22) is also necessary for $\hat{\sigma}$ to be a SPE.

Now it is shown that the equilibrium point is subgame perfect. By Definition (1), an equilibrium point is subgame perfect if it is a Nash equilibrium strategy combination for any possible subgame of the original game. Let us define an arbitary subgame by selecting a value for time t and some history h_t, and partition all subgames into two sets. The first consists of those subgames in which $h_t = (\bar{w}, \ldots, \bar{w})$ for all $t \geq 0$. The second set is the complement of the first. For a subgame in the first set, by using the trigger strategy, worker i will have utility $\bar{U}_i(\bar{w})$ in each period; that is, $\hat{\sigma} = (\bar{w}, \tilde{w})$ induces an equilibrium point on the subgame. For a subgame in the second set, all worker j will choose $\tilde{w}$ in all periods. This is also an equilibrium point for the subgame; therefore, $\hat{\sigma} = (\bar{w}, \tilde{w})$ is subgame perfect (Friedman, 1986).

Q. E. D.

6.4 Team Optimization Program

6.4.1 Implications of An Optimal Labour Contract

The team collectively chooses the parameters of an optimal labour contract $\{ \alpha, w, C \}$ given individual workers' effort decisions in a repeated game. Dropping the subscript i in the discussion of the team optimization program, we rewrite the expected utility of a representative worker who chooses effort z when others supply w as $U(\alpha,w,C,z)$, and define $\hat{U}(\alpha,w,C) \equiv \max_{z \in [0,1]} U(\alpha,w,C,z)$. We also write $\bar{U}(\alpha,w,C) \equiv U(\alpha,w,C,w)$, where $\bar{U}(\alpha,w,C)$ is the expected utility achieved through compliance with the contract, i.e. $z = w$.

From the discussion of SPEs discussed in Section 6.3, we note that a feasible contract must satisfy the following conditions:

IC: $\bar{U}(\alpha,w,C) \geq (1-\delta)\, \hat{U}(\alpha,w,C)$

CRED: (a) if $Y(\alpha,w,0) \leq 0$, then $C = 0$;

(b) otherwise, $\frac{\gamma}{n} Y(\alpha,w,0) \geq C \geq 0$.

Requiring that a worker cannot be better off through deviation from an equilibrium, IC must hold for any SPE. Note that if $\delta = 0$, then we are back to the static model of chapter 4 because the IC implies that no deviation from the contractual effort can make an individual worker better off, even in a one-shot situation. Part (a) of CRED applies when any of the cases: $\alpha < \alpha_{min}$, $w < w^{\alpha}_{min}$ or $\Theta(n-1)^{\beta} \leq D$ is true. Part (b) of CRED applies only if $\alpha \in [\alpha_{min},1]$, $w \in [w^{\alpha}_{min},1]$ and $\Theta(n-1)^{\beta} > D$. Now let us consider the characteristics of feasible labour contracts.

Theorem 7: Every contract (α,w,C) satisfying IC and part (a) of CRED has $w = \tilde{w} = 0$ and $\bar{U}(\alpha,w,0) = 0$. For any $\alpha \in [0,1]$, the contract $(\alpha,0,0)$ is feasible.

Proof: Set $z = 0$, part (a) of CRED implies $U(\alpha,w,0,0) = 0$ for any α, $w \in [0,1]$. For any $z > 0$ such that $Y(\alpha,w,z) \leq 0$, we have $U(\alpha,w,0,z) = -zv < 0$. For any $z > 0$ such that $Y(\alpha,w,z) > 0$, we have $U(\alpha,w,0,z) = \frac{1}{n}Y(\alpha,w,z) - zv$. But we now show that this is nonincreasing in z:

$$U_z(\alpha,w,0,z) = \frac{\beta\Theta}{n}\alpha^{\beta}[w(n-1)+z]^{\beta-1} - v, \tag{23}$$

but $Y(\alpha,w,z) \geq 0$ implies $\Theta\alpha^{\beta}[w(n-1)+z]^{\beta} \geq D$, or $[w(n-1)+z]^{\beta-1} \leq [D/\Theta\alpha^{\beta}]^{1-1/\beta}$. Hence,

$$U_z(\alpha,w,0,z) \leq \frac{\beta\Theta}{n}\alpha^{\beta}[D/\Theta\alpha^{\beta}]^{1-1/\beta} - v \leq \frac{\beta}{n}\Theta^{1/\beta}D^{1-1/\beta}(\alpha-1) \leq 0, \tag{24}$$

where the middle inequality follows from A3. Since $U(\alpha,w,0,z)$ is nonincreasing in z for $Y(\alpha,w,z) \geq 0$, and equals $-zv < 0$ when $z > 0$ and $Y(\alpha,w,z) < 0$, we have shown that $U(\alpha,w,0,z)$ obtains a unique maximum of zero at $z = 0$. However, the IC condition requires $\bar{U}(\alpha,w,C) \geq (1-\delta)\,\hat{U}(\alpha,w,0) = 0$. Since $U(\alpha,w,C,w) < 0$ for any $w > 0$, we conclude that $w = \tilde{w} = 0$. In this case, $\bar{U}(\alpha,w,0) = 0$, establishing the first part of Theorem 7. The second part follows by observing that for any $\alpha \in [0, 1]$, $w = z = 0$ implies $Y(\alpha,w,z) < 0$, so part (a) of CRED is satisfied; and IC holds because $\bar{U}(\alpha,0,0) = (1-\delta)\,\hat{U}(\alpha,0,0) = 0$ by the preceding argument. Q. E. D.

Corollary: Any feasible contract (α, w, C) yielding $\bar{U}(\alpha, w, C) > 0$ must satisfy IC and part (b) of CRED. Any such a contract requires $w = \bar{w} \in [w^{\alpha}_{min}, 1]$ for $\alpha \in [\alpha_{min}, 1]$.

From Theorem 7 and its corollary, we observe that a feasible contract always exists because $(\alpha, 0, 0)$ is always feasible. A feasible contract with $\bar{U}(\alpha, w, C) > 0$ may or may not exist; but if it exists, then the team optimum must satisfy IC and part (b) of CRED.[41] If there is a feasible contract with $\bar{U}(\alpha, w, C) > 0$, the team does not choose a contract of type (a).

In what follows, we assume that there exists a feasible contract (α, w, C) with $\bar{U}(\alpha, w, C) > 0$. The team optimum then solves the problem

$$\max_{\alpha, w, C} \quad \bar{U}(\alpha, w, C)$$

$$\text{Subject to: IC:} \quad \bar{U}(\alpha, w, C) \geq (1-\delta)\hat{U}(\alpha, w, C)$$

$$\text{CRED:} \quad \frac{\gamma}{n} Y(\alpha, w, 0) \geq C \geq 0.$$

In CRED, we use the fact that an optimal contract will satisfy part (b) of the original CRED constraint. It follows that $\alpha \in [\alpha_{min}, 1]$, $w \in [w^{\alpha}_{min}, 1]$ and $C \geq 0$.

We first consider the choice of optimal work effort w, and then discuss the features of optimal α and C. By A2, we know that increased effort is always socially desirable; therefore, $w = 1$ is preferred to $w < 1$ whenever $w = 1$ is feasible. But under what conditions is the team able to choose $w = 1$ when $w < 1$ appears to be feasible? This question will be answered by Theorem 8-10. Theorem 8 identifies the feasibility conditions of $w = 1$ when $\alpha < 1$ and $\delta < w$. Theorem 9 considers the same issue in the case where $\alpha = 1$ and $\delta < w$. Theorem 10 handles the case where $\delta \geq w$.

[41] In Section 6.2, we showed that a labour contract which satisfies the static incentive compatibility condition and part (b) of CRED must yield positive utility. At the end of this section, we will show that this is also true in the presence of a dynamic monitoring mechanism.

Theorem 8: Assume A1-A3 hold. Arbitrarily fix α such that $\alpha_{min} < \alpha < 1$. Also fix $\delta \in (0,1)$ and $\lambda \equiv D/\Theta(n-1)^{\beta} \in (0,1)$, and $\eta \equiv \frac{vn}{\beta D} \in [1/[(n-1)\lambda^{1/\beta},1]$. There exists some n^* such that whenever $n > n^*$ if there is some feasible contract $(\alpha,\hat{w},\hat{C})$ with $\delta < \hat{w} < 1$, then (α,w',C') with $w' = 1$ is also feasible for some $C' \geq 0$. The value of n^* depends on δ, β, Θ, v and D, but not (α,w,C).

Note that the restriction on the values of λ is given by A1, and the restriction on possible values of η follows from A2 and A3. By keeping λ and η fixed, we assume that A1-A3 continue to hold as n becomes large. This is certainly true for A2 and A3, and therefore a fixed η. A fixed λ can be justified by the fact that both total revenue and total cost are increasing as the size of a team grows. What is important in A1 is that a team must possess positive economic rent before any efficient contract can be implemented.

To prove theorem 8, we introduce the following notations and lemmas.

(25) $$C_m(w) \equiv \gamma[\frac{\Theta}{n}(\alpha(n-1)w)^{\beta} - \frac{D}{n}]$$

(26) $$C_i(w) \equiv [\Phi(\alpha)w(w-\delta)]^{-1}[(1-\delta)(\frac{\Theta}{n}(\alpha(n-1)w)^{\beta} - \frac{D}{n}) - (\frac{\Theta}{n}(\alpha nw)^{\beta} - \frac{D}{n} - wv)]$$

and

(27) $$C_0(w) \equiv [\Phi(\alpha)w]^{-1}[v - \frac{\beta\Theta}{n}\alpha^{\beta}(w(n-1))^{\beta-1}],$$

where $C_m(w)$ is the maximum penalty a team is able to impose, $C_i(w)$ is the minimum penalty which satisfies dynamic IC when $\hat{z} = 0$, and $C_0(w)$ is the minimum penalty with which the static IC condition holds.

Lemma 1: $C_0(w) \geq 0$ for all $w \in [w^{\alpha}_{min},1]$ and $\alpha \in [\alpha_{min},1)$, and $C_0(1) > 0$.

Proof: Follows the fact that $\Phi(\alpha) > 0$ with $\alpha > 0$, $w > 0$ and $v - \frac{\beta\Theta}{n}\alpha^{\beta}(w(n-1))^{\beta-1} \geq 0$

for all $w \in [w^{\alpha}_{min},1]$ and $\alpha \in [\alpha_{min},1]$ given A3. Q. E. D.

Lemma 2: For fixed v, β, λ and η, there exists some n_1 such that $n > n_1$ implies $C_0'(w) \leq 0$ for all $w \in [w^{\alpha}_{min},1]$ and $\alpha < 1$.

Proof: Rewrite $C_0(w)$ in terms of λ and η, i.e.

$$(28) \qquad C_0(w) \equiv [1/\Phi(\alpha)w]^{-1}v[1-\frac{1}{\lambda\eta(n-1)}\alpha^{\beta}w^{\beta-1}].$$

The derivative of $C_0(w)$ with respect to w is

$$(29) \qquad C_0'(w) = -[\Phi(\alpha)w^2]^{-1}v[1 - \frac{(2-\beta)}{\eta\lambda(n-1)}\alpha^{\beta}w^{\beta-1}]$$

$C_0'(w) \leq 0$ if $[1 - \frac{(2-\beta)}{\lambda\eta(n-1)}\alpha^{\beta}w^{\beta-1}] > 0$, because $\Phi(\alpha) > 0$ at $\alpha < 1$, $w > 0$ and $v > 0$. But if this inequality holds for $\alpha = 1$ and $w^1_{min} \equiv (D/\Theta)^{1/\beta}(n-1)^{-1}$, it must hold for all $\alpha < 1$ and $w \in [w^{\alpha}_{min},1]$. Evaluating the inequality at $\alpha = 1$ and $w = w^1_{min}$, we note that the inequality holds if $1 > (2-\beta)/[\eta\lambda^{1/\beta}(n-1)]$. It therefore suffices to choose

$$(30) \qquad n_1 \geq 1 + (2-\beta)/[\lambda^{1/\beta}\eta].$$

Note that n_1 is independent of δ as well as (α,w,C). Q. E. D.

Lemma 3: $\hat{z} > 0$ if and only if $C > C_0(w)$.

Proof: The derivative of $U(\alpha,w,C,z)$ with respect to z is

$$(31) \qquad U_z(\alpha,w,C,z) = \frac{\beta\Theta}{n}\alpha^{\beta}[(n-1)\bar{w}+z]^{\beta-1} - v + \bar{w}\Phi(\alpha)C.$$

Note that $U_z(\alpha,w,C,z)$ is strictly decreasing in z. We can show that if $C > C_0(w)$, then $U_z(\alpha,w,C,0) > 0$. It pays for a worker to increase effort; an optimal is reached with $\hat{z} \in (0,1)$, where $U_z(\alpha,w,C,\hat{z}) = 0$, or $\hat{z} = 1$, where $U_z(\alpha,w,C,\hat{z}) \geq 0$. If $C \leq C_0(w)$, then $U_z(\alpha,w,C,0) \leq 0$, and an optimum implies $\hat{z} = 0$. Q. E. D.

Define $R(\alpha,w,C) \equiv \tilde{U}(\alpha,w,C) - (1-\delta)\hat{U}(\alpha,w,C)$. IC holds if and only if $R(\alpha,w,C) \geq 0$.

Lemma 4: For fixed $0 < \beta < 1$, $0 < \delta < 1$, and feasible (α,w,C), there exists n_2 such that $n > n_2$ implies $\frac{\partial R(\alpha,w,C)}{\partial w} > 0$ whenever $\hat{z}(\alpha,w,C) > 0$ and $\frac{\partial R(\alpha,w,C)}{\partial C} \geq 0$.

Proof: The derivative of $R(\alpha,w,C)$ with respect to w is

$$(32) \qquad \begin{aligned} \frac{\partial R(\alpha,w,C)}{\partial w} &= \frac{\beta\Theta}{n}\alpha^{\beta}(wn)^{\beta-1} - \Phi(\alpha)(1-2w)C - v \\ &\quad - (1-\delta)[\frac{\beta\Theta}{n}\alpha^{\beta}[w(n-1)+\hat{z}]^{\beta-1}(n-1) - \Phi(\alpha)(1-\hat{z})C \\ &= \frac{\beta\Theta}{n}\alpha^{\beta}[w^{\beta-1}n^{\beta} - (1-\delta)(w(n-1) + \hat{z})^{\beta-1}(n-1)] \\ &\quad + \Phi(\alpha)wC - v - \Phi(\alpha)C(1-w) + (1-\delta)\Phi(\alpha)C(1-\hat{z}). \end{aligned}$$

Note that the term $\frac{\partial R(\alpha,w,C)}{\partial z}\frac{\partial \hat{z}}{\partial w}$ is ignored by using the envelope theorem. But $\frac{\partial U(\alpha,w,C,z)}{\partial z} = 0$ if $\hat{z} \in (0,1)$, or $\frac{\partial U(\alpha,w,C,z)}{\partial z} \geq 0$ if $\hat{z} = 1$, that is

$$(33) \qquad \frac{\beta\Theta}{n}\alpha^{\beta}[w(n-1)+\hat{z}]^{\beta-1} + w\Phi(\alpha)C - v \geq 0.$$

Using this result, we have

$$(34) \qquad \begin{aligned} \frac{\partial R(\alpha,w,C)}{\partial w} &\geq \frac{\beta\Theta}{n}\alpha^{\beta}[w^{\beta-1}n^{\beta} - (w(n-1) + \hat{z})^{\beta-1}((1-\delta)(n-1) + 1)] \\ &\quad - \Phi(\alpha)C[(1-w)-(1-\delta)(1-\hat{z})]. \end{aligned}$$

But $\frac{\partial R}{\partial C} = -\Phi(\alpha)w[(1-w)-(1-\delta)(1-\hat{z})]$. If $\frac{\partial R}{\partial C} \geq 0$, then $(1-w) - (1-\delta)(1-\hat{z}) \leq 0$, because

$\Phi(\alpha) > 0$ with $\alpha < 1$, and $w > 0$. Hence, the last term of the right hand side of (35) is equal or greater than zero for $C \geq 0$ and $\Phi(\alpha) \leq 1$. It thus suffices to show that

$$w^{\beta-1}n^{\beta} > [w(n-1) + \hat{z}]^{\beta-1}[(1-\delta)(n-1) + 1]$$

or

$$(wn)^{\beta} > [w(n-1) + \hat{z}]^{\beta-1}[w(1-\delta)(n-1) + w] \equiv \text{RHS}.$$

But

$$\text{RHS} \leq (w(n-1))^{\beta-1}[w(1-\delta)(n-1) + w] = w^{\beta}(n-1)^{\beta-1}[(1-\delta)(n-1) + 1].$$

So it is sufficient to show that $n^{\beta} > (n-1)^{\beta-1}[(1-\delta)(n-1) + 1]$ or $(\frac{n}{n-1})^{\beta} > 1 - \delta + \frac{1}{n-1}$; but this holds for a sufficiently large n when $\delta > 0$. Choose n_2 such that $n > n_2$ implies this inequality holds; note that n_2 depends on β, δ, but not (α, w, C). Q. E. D.

Lemma 5: For $\alpha < 1$ and $w > \delta$, if $\hat{z} = 0$, then IC holds if and only if $C \geq C_i(w)$.

Proof: Follows from the definition of IC when $\hat{z} = 0$, $\alpha < 1$ and $w > \delta$. Q. E. D.

Lemma 6: CRED holds if and only if $C_m(w) \geq C$.

Proof: Follows from the definition of CRED. Q. E. D.

Lemma 7: Given $\gamma > 0$, $\beta > 0$, $\Theta > 0$ and $n > 1$, $C'_m(w) > 0$ for all $\alpha \in (0,1]$ and $w \in (0,1]$.

Proof: Follows from $C'_m(w) = \frac{\gamma}{n}\beta\Theta\alpha^{\beta}w^{\beta-1}(n-1)^{\beta} > 0$. Q. E. D.

Lemma 8: For fixed β, δ, λ and η, there exists some n_3 such that $n \geq n_3$ implies if $C_0(\hat{w}) \geq C_i(\hat{w})$, then $C_0(w) \geq C_i(w)$ for all $w > \hat{w}$.

Proof: Write $C_0(w)$ and $C_i(w)$ in terms of η and λ, we have

$$(35) \qquad C_0(w) \equiv \frac{v}{\Phi(\alpha)w}[1 - \frac{1}{\eta\lambda(n-1)}\alpha^{\beta}w^{\beta-1}],$$

and

$$(36) \qquad C_i(w) \equiv \frac{d}{\Phi(\alpha)w(w-\delta)}[\lambda^{-1}(\alpha w)^{\beta}(1-\delta-(\frac{n}{n-1})^{\beta}) + \delta + \eta\beta w].$$

$C_0(w) \geq C_i(w)$ holds if and only if

$$(37) \qquad (w-\delta)v[1-\frac{1}{\eta\lambda(n-1)}\alpha^{\beta}w^{\beta}\ ] \geq d[\lambda^{-1}(\alpha w)^{\beta}(1-\delta-(\frac{n}{n-1})^{\beta}) + \delta + \eta\beta w].$$

or

$$(w-\delta)\beta[\eta-\frac{1}{\lambda(n-1)}\alpha^{\beta}w^{\beta-1}] \geq [\lambda^{-1}(\alpha w)^{\beta}(1-\delta-(\frac{n}{n-1})^{\beta}) + \delta + \eta\beta w].$$

or

$$\lambda^{-1}(\alpha w)^{\beta}[\delta + (\frac{n}{n-1})^{\beta} - 1 - \frac{\beta}{n-1} + \frac{\delta\beta}{w(n-1)}] \geq \delta(1+\beta\eta).$$

Suppose the inequality holds for some $\hat{w}$. Then it suffices if the LHS is a strictly increasing function of w. The derivative of the LHS with respect to w is

$$(38) \qquad \beta\lambda^{-1}\iota^{\beta}w^{\beta-1}[\delta + (\frac{n}{n-1})^{\beta} - 1 - \frac{(1-\beta)\delta}{(n-1)w} - \frac{\beta}{(n-1)}].$$

It suffices if $\delta(n-1) > \beta + (1-\beta)\frac{\delta}{w}$. But a feasible w must be greater than $w^{\alpha}_{min} = (\frac{D}{\Theta})^{1/\beta}[(n-1)\alpha]^{-1} = \lambda^{1/\beta}\alpha^{-1} \geq \lambda^{1/\beta}$. Hence, it is sufficient to set $n_3 > 1 + \frac{\beta}{\delta} + (1-\beta)\lambda^{-1/\beta}$. Note that n_3 depends on β, δ and λ, but is independent of η as well as (α,w,C).

Q. E. D.

We now show the proof of Theorem 8.

Proof: Let $n^* = \max(n_1, n_2, n_3)$ where n_i, $i = 1,2,3$ are defined by Lemma 2, Lemma 4 and Lemma 8. Assume $n > n^*$. Suppose $(\alpha,\hat{w},\hat{C})$ is feasible where $\hat{w} < 1$. By Lemma 6, $\hat{C} \le C_m(\hat{w})$. With regard to $\hat{C}$ and $C_0(\hat{w})$, there are two possibilities: either (a): $\hat{C} > C_0(\hat{w})$ or (b): $\hat{C} \le C_0(\hat{w})$.

Consider case (a): $\hat{C} > C_0(\hat{w})$ and $\hat{z}(\alpha,\hat{w},\hat{C}) > 0$ by Lemma 3. Since $C_0'(w) \le 0$ for all $w \in [w^{\alpha}_{min}, 1]$ by Lemma 2, all contracts of the form (α, w', C) with $w' \ge \hat{w}$ have $\hat{z} > 0$. With regard to IC, we note that $\frac{\partial R(\alpha,w,C)}{\partial C} = -\Phi(\alpha)w[1-w-(1-\delta)(1-\hat{z})]$, which depends on C only through $\hat{z}$. For fixed α and w, $\frac{\partial^2 R}{\partial C^2} = -\Phi(\alpha)w(1-\delta)\frac{\partial \hat{z}}{\partial C} \le 0$. If $(\alpha,\hat{w},\hat{C})$ is chosen such that $\frac{\partial R}{\partial C} \ge 0$, and $\hat{z} > 0$, then $(\alpha,1,\hat{C})$ is feasible because $\frac{\partial R}{\partial w} > 0$ whenever $\hat{z} > 0$ and $\frac{\partial R}{\partial C} \ge 0$ by Lemma 4, and $C'_m(w) \ge 0$ by Lemma 7. If $(\alpha,\hat{w},\hat{C})$ is such that $\frac{\partial R}{\partial C} < 0$ and $\hat{z} > 0$, then one can always hold $\hat{w}$ fixed and reduce $\hat{z}$ through curtailing C until $\frac{\partial R}{\partial C} \ge 0$. Note that $\hat{z} = 0$ when $C = 0$ for any α, $\hat{w}$, and $\frac{\partial R}{\partial C} > 0$. So there is a $C' > 0$ and a $\hat{z} > 0$ such that $\frac{\partial R}{\partial C} = 0$ by continuity. Because initially $\frac{\partial R}{\partial C} < 0$ at $(\alpha,\hat{w},\hat{C})$, we can reduce C until $\frac{\partial R}{\partial C} = 0$ without violating IC, because $R(\alpha,\hat{w},\hat{C}) \ge 0$ by assumption and R is increasing as C is reduced from $\hat{C}$ to C'. The remainder of the proof is as above.

Now consider case (b): $\hat{C} \le C_0(\hat{w})$ and $\hat{z} = 0$. Feasibility implies $\hat{C} \ge C_i(\hat{w})$ by Lemma 5. Allowing the penalty to vary with w, we define $\tilde{C}(w) \equiv \max\,[0, C_i(w)]$. $(\alpha,\tilde{C}(\hat{w}),\hat{w})$, where $\tilde{C}(\hat{w}) \le \hat{C}$ is also feasible and also has $\hat{z} = 0$. Since $C'_0(w) \le 0$ by Lemma 2, and $C_0(\hat{w}) \ge C_i(\hat{w})$ implies $C_0(w) \ge C_i(w)$ for all $w \ge \hat{w}$ by Lemma 8, we have $\tilde{C}(w) \le C_0(w) \le C_m(w)$ for all $w \ge \hat{w}$. Moreover, $C_0(1) > 0$ by Lemma 1. Now consider the contract $[\alpha,1,\tilde{C}(1)]$ where $\tilde{C}(1) \equiv \max[0, C_i(1)]$. This contract has $\hat{z} = 0$ since $\tilde{C}(1) \le C_0(1)$. IC holds because $\tilde{C}(1) \ge C_i(1)$ and $\hat{z} = 0$. Since $C_m(1) \ge \tilde{C}(1)$, CRED also holds, and so the contract is feasible. Q. E. D.

Theorem 9: Assume A1-A3 hold. Fix $\delta \in (0,1)$. If the contract $(1,\hat{w},\hat{C})$ is feasible, then so is $(1,1,0)$, except possibly for a region in the parameter space whose measure is zero.

Proof: If $\hat{C} > 0$, then the contract $(1,\hat{w},0)$ is also feasible since C does not enter into IC when $\alpha = 1$, and $C = 0$ is permitted by CRED. Hence, in what follows we assume $\hat{C} = 0$. Define $\delta^* = [v-\frac{\Theta}{n}(n^\beta-(n-1)^\beta)][\frac{\Theta}{n}(n-1)^\beta - d]^{-1}$. For a fixed $\delta \geq \delta^*$, IC holds for $w = 1$, $\alpha = 1$ and $C = 0$, i.e. $R(1,1,0) \geq 0$, and CRED is also satisfied. Hence, if $(1,\hat{w},0)$ is feasible, then $(1,1,0)$ is also feasible.

Now consider the case where $\delta < \delta^*$. In this case, $R(1,1,0) < 0$ and a contract with $\alpha = 1$ and $w = 1$ is not feasible. What could happen in this case depends upon the nature of $R(1,w,0) \equiv R(w)$, where

$$R(w) = \frac{\Theta}{n}(nw)^\beta - d - wv - (1-\delta)[\frac{\Theta}{n}((n-1)w)^\beta - d]\,. \tag{39}$$

The derivative of R(w) with respect to w is

$$\frac{\partial R(w)}{\partial w} = w^{\beta-1}\frac{\beta\Theta}{n}[n^\beta - (1-\delta)(n-1)^\beta] - v. \tag{40}$$

Denote w^* as a w that maximizes R(w), $\hat{\delta} \equiv \eta\lambda + 1 - (\frac{n}{n-1})^\beta$, and $\bar{\delta} \equiv \eta\lambda^{1/\beta} + 1 - (\frac{n}{n-1})^\beta$. We note that if $\delta^* > \delta \geq \hat{\delta}$, then $w^* = 1$. Because of the concavity of R(w), R(w) is strictly increasing in $w \in [w^1_{min},1]$. Hence, $R(1) < 0$ implies $R(w) < 0$ for all $w \in [w^1_{min},1]$. This case contradicts the assumption that some contrat $(1,\hat{w},0)$ is feasible, and hence this case can be ignored. If $\delta \leq \bar{\delta}$, $R'(w^1_{min}) \leq 0$, and then $w^* = w^1_{min}$, where $w^1_{min} \equiv \lambda^{1/\beta}$. So R(w) is strictly decreasing in w for all $w \geq w^1_{min}$. But $R(w^1_{min}) = \bar{U}(1,w^1_{min},0) - (1-\delta)\hat{U}(1,w^1_{min},0,0) \leq 0$, since $\hat{U}(1,w^1_{min},0,0) = 0$ by the definition of w^1_{min}, and $\bar{U}(1,w^1_{min},0) < \hat{U}(\alpha,w^1_{min},0,0) = 0$ given $U_z(\alpha,w,0) < 0$ for all $w \geq w^1_{min}$ by A3. Hence, $R(w) < 0$ for all $w \in [w^1_{min},1]$. Again, no contract with $\alpha = 1$ is feasible.

The remaining possibility is $\min[\delta^*,\hat{\delta}] > \delta > \bar{\delta}$. In this situation, $w^* = \{\beta\Theta[(n^\beta - (1-\delta)(n-1)^\beta]/vn\}^{1/(1-\beta)} \in (w^1_{min},1)$. There are three possible cases: (a) $R(w^*) < 0$, then $R(w) < 0$ for all $w \in (w^1_{min},1]$. In this case there is also no feasible contract with $\alpha = 1$.

(b) $R(w^*) > 0$, and $R(w) \geq 0$ holds for all $w \in [\underline{w},\bar{w}]$, where $\underline{w} > w'_{min}$ and $\bar{w} < 1$ and $R(\underline{w}) = R(\bar{w}) = 0$. All contracts $w \in [\underline{w},\bar{w}]$ are feasible when $\alpha = 1$. It is easy to show that this case can be reduced to the situation described in Theorem 8 where $\alpha < 1$ and $\delta < w$. Consider $\bar{w}$.[49] Let $w' = \bar{w} - \epsilon_1 > w^*$, where $\epsilon_1 > 0$ is chosen so that $R(1,w',0) > 0$. Then choose some $\epsilon_2 > 0$ such that $\alpha' = 1 - \epsilon_2 < 1$ and $R(\alpha',w',0) > 0$. This is possible by continuity. Thus, the contract $\alpha' = 1-\epsilon_2$, $w' = 1-\epsilon_1$, $C = 0$ is feasible and $\alpha' < 1$. Moreover, it can be shown that $R(w^*) > 0$ implies $w' > \delta$. Substituting w^* into $R(w)$ gives us

$$(41) \qquad R(w^*) = [\frac{\beta\Theta}{vn}(n^{\beta}-(1-\delta)(n-1)^{\beta})]^{1/(1-\beta)}\frac{v}{\beta}(1-\beta) - \delta D/n.$$

$R(w^*) > 0$ implies $w^*\frac{vn}{\beta D}(1-\beta) > \delta$; therefore, $w^* > \delta$ since $\frac{vn}{\beta D} \leq 1$ by A2 and $(1-\beta) < 1$. Hence, as long as $R(w^*) > 0$, $w' > w^* > \delta$. Now applying Theorem 8, we can show that if $w' < 1$ is feasible when $\alpha' < 1$, $C' \geq 0$ and $w' > \delta$, then $w = 1$ is also feasible for $\alpha' < 1$ and $C' \geq 0$.[50]

(c) $R(w^*) = 0$; and $R(w) < 0$ for all $w \neq w^*$. In this case the only feasible contract with $\alpha = 1$ has $w^* < 1$. But this can only occur when the parameters β, Θ, v, n, δ and D strictly define a relationship such that $[\frac{\beta\Theta}{vn}(n^{\beta}-(1-\delta)(n-1)^{\beta})]^{1/(1-\beta)}\frac{v}{\beta}(1-\beta) - \delta D/n = 0$. This equality holds only for a region of measure zero in the parameter space. Assuming that this relation does not hold, we can ignore case (c) in the following discussion. Q. E. D.

---

[49] By A2, w is always preferred to $\underline{w}$ when both $\bar{w}$ and $\underline{w}$ are feasible.

[50] Case (b) presents a situation where the team must make a choice between (i) saving monitoring cost by tolerating some degree of shirking and (ii) enforcing maximum effort by investing on labour supervision. The optimal choice depends upon the relative strength of dynamic monitoring and static monitoring. If a high level of effort, although not the possibile maximum, can be ensured merely by informal monitoring, the team may prefer (i) to (ii), otherwise (ii) to (i). In the discussion of case (b), for simplicity we rule out a might-be optimal labour contract $\alpha = 1$, $w' < 1$ and $C \geq 0$ at $\delta' < \delta^*$, where w' is fairly close to 1 and δ' is close to δ^*. It is argued that this type of contract can be well approximated by an optimal contract $\alpha = 1$, $w = 1$ and $C \geq 0$ at $\delta \geq \delta^*$, because to some extent that effort is a relative matter.

Theorem 10: Any optimal contract with $0 < w \leq \delta < 1$ must have $\alpha = 1$.

Proof: Suppose $(\hat{\alpha}, \hat{w}, \hat{C})$ is an optimal contract with $\hat{w} \leq \delta$ and $\hat{\alpha} < 1$. We have

$$(42) \qquad \frac{\partial R}{\partial C} = -\Phi(\alpha)w[1-w-(1-\delta)(1-\hat{z})]$$
$$= 0 \text{ if } \hat{z} = 0 \text{ and } \delta = \hat{w}$$
$$< 0 \text{ if } \hat{z} > 0 \text{ or } \delta > \hat{w}.$$

If $\frac{\partial R}{\partial C} < 0$, reduction in C will increase $R(\alpha,w,C)$ without decreasing the objective function. It is then possible to increase α, raising utility. This contradicts the optimality of the original contract.

If $\frac{\partial R}{\partial C} = 0$, $\hat{z} = 0$ must hold. $R(\alpha,w,C)$ then remains unchanged when C is reduced from $C_0(\hat{w}) \geq \hat{C}$ to zero. Therefore, we can assume $C = 0$ holds whenever $\frac{\partial R}{\partial C} \leq 0$, since the objective function is nondecreasing for small reductions in C. Such reductions are always permitted by CRED. But if $C = 0$, then

$$(43) \qquad R(\alpha,w,0) = \frac{\Theta}{n}(\alpha n w)^{\beta} d - wv - (1-\delta)[\frac{\Theta}{n}(\alpha(n-1)w)^{\beta} - d],$$

and

$$(44) \qquad \frac{\partial R(\alpha,w,0)}{\partial \alpha} = \frac{\beta\Theta}{n}\alpha^{\beta-1}w^{\beta}[n^{\beta} - (1-\delta)(n-1)^{\beta}] > 0.$$

If $\alpha < 1$ is feasible at $C = 0$, then $\alpha = 1$ must also be feasible because R is strictly increasing in α. The team will then increase α, since the objective function is strictly increasing in α. Such increases are permitted by IC and CRED. This contradicts optimality.

Q. E. D.

Having discussed the feasible effort contract, we are now able to characterize all the parameters of an optimal labour contract.

Theorem 11: Assume the conditions of Theorem 8, 9 and 10 are met, $n > n^*$, and there is a feasible contract with $\tilde{U}(\alpha,w,C) > 0$. Define $\delta^* = [v-\frac{\Theta}{n}(n^\beta-(n-1)^\beta)][\frac{\Theta}{n}(n-1)^\beta - d]^{-1}$.

(a) If $\delta \geq \delta^*$, any optimal contract (α^*, w^*, C^*) has $\alpha^* = w^* = 1$ and $C^* = 0$. In this case, $\tilde{U}(1,1,0) \geq (1-\delta)\hat{U}(1,1,0)$, $\frac{\gamma}{n}Y(1,1,0) > 0$ and $\hat{z}(1,1,0) = 0$.

(b) If $0 < \delta < \delta^*$, then there is $\tilde{n}$ such that whenever $n > \tilde{n}$, any optimal contract has

(i) $w^* = 1$, $\alpha^* < 1$ and $C^* > 0$;

(ii) $\tilde{U}(\alpha^*,1,C^*) = (1-\delta)\hat{U}(\alpha^*,1,C^*)$;

(iii) $\frac{\gamma}{n}Y(\alpha^*,1,0) = C^*$;

(iv) α^* is the largest $\alpha \in (\alpha_{min},1]$ such that (ii) and (iii) hold for some $C^* > 0$;

(v) $\hat{z}(\alpha^*,1,C^*) = 0$, and $C^* = C_i(1)$.

(c) If $\delta = 0$, any feasible contract has

(i) $w^* = 1$, $\alpha^* < 1$, and $C^* > 0$;

(ii) $\hat{z}(\alpha^*,1,C^*) = 1$;

(iii) $\frac{\gamma}{n}Y(\alpha^*,1,0) = C^*$;

(iv) α^* is the largest $\alpha \in (\alpha_{min},1]$ such that (ii) and (iii) hold for some $C^* > 0$.

Proof: (a) Since $\tilde{U}(\alpha,w,C) > 0$ for some feasible (α,w,C), an optimal contract must yield positive utility. By Theorem 7, it follows that an optimal contract satisfies the IC condition and part (b) of the original credibility constraint, since otherwise the team has zero utility. Hence, $w > w^\alpha_{min}$ and $\alpha > \alpha_{min}$. Because $\tilde{U}_{iw} > 0$ given A2, $\tilde{U}_\alpha > 0$, and $\tilde{U}_C \leq 0$, we have, $\tilde{U}(1,1,0) > \tilde{U}(\alpha,1,C) > \tilde{U}(\alpha,w,C)$ for any $\alpha < 1$ and $w < 1$ and $C > 0$. Also $\tilde{U}(\alpha,1,C)$ is independent of C for all $\alpha \in [0,1]$. Hence, whenever a contract (α^*,w^*,C^*) with $\alpha^* = w^* = 1$ and $C^* = 0$ is feasible, it is optimal. But $C^* = 0$ implies $\hat{z}(1,1,0) = 0$. When $C^* = 0$ and $\hat{z} = 0$, $\frac{\gamma}{n}Y(1,1,0) > 0$ because $\gamma > 0$ and $\frac{1}{n}Y(1,1,0) \equiv U(1,1,0,0) > U(\alpha,w,0,0) \geq 0$ for any $\alpha < 1$ and $w < 1$. $\alpha^* = w^* = 1$ and $C^* = 0$ satisfy the IC condition if and only if $R(1,1,0) = \frac{\Theta}{n}n^\beta - d - v - (1-\delta)(\frac{\Theta}{n}(n-1)^\beta - d) \geq 0$. But this holds whenever $\delta \geq \delta^*$.

(b) If $\delta \in (0,\delta^*)$, then $R(1,1,0) < 0$ so that the contract $(1,1,0)$ is infeasible. From Theorem 9, we know that a contract with $\alpha = 1$ is not optimal because this would contradict the infeasibility of $(1,1,0)$. Hence, an optimal contract must have $\alpha^* < 1$. But if $\alpha^* < 1$, then $C^* > 0$ must hold. If $C^* = 0$, the team could increase α to raise the objective function without violating IC or CRED. This contradicts optimality. From Theorem 8, there exists some n^* such that whenever $n > n^*$ if some (α,w,C) is feasible with $\alpha < 1$, $w < 1$, and $C \geq 0$, then $(\alpha,1,C')$ is also feasible for some C'. Since we already showed in the proof of part (a) that $\bar{U}(\alpha,1,C') > U(\alpha,w,C')$ for any $w < 1$ and any C, any optimal contract must have $w^* = 1$. This proves (i).

To prove (ii), assume an optimal contract (α^*,w^*,C^*) with $\alpha^* < 1$, $w^* = 1$ and $C^* > 0$ has strictly inequality in the IC condition. Then we can choose some $\epsilon > 0$ such that the IC condition holds at $(\alpha^* + \epsilon,1,C^*)$ and so does the credibility constraint. But utility is higher at $(\alpha^* + \epsilon,1,C^*)$, contradicting the optimality of (α^*,w^*,C^*). So IC holds with equality.

To prove (iii), consider an optimal contract with $\alpha^* < 1$, $w^* = 1$ and $R(\alpha^*,1,C^*) = 0$, which must hold by (ii) above, then $\frac{\partial R}{\partial C} = (1-\delta)\Phi(\alpha^*)(1-\hat{z}) \geq 0$ with $= 0$ if $\hat{z} = 1$ and > 0 otherwise. But if $\hat{z} = 1$, then $R(\alpha^*,1,C^*) = \delta[\frac{\Theta}{n}(n\alpha)^\beta - d - v] > 0$, since an optimal contract must yield positive utility. This contradicts $R(\alpha^*,1,C^*) = 0$, so we conclude that $\hat{z} < 1$ and $\frac{\partial R}{\partial C} > 0$. Now suppose the optimal contract $(\alpha^*,1,C^*)$ has $C^* < C_{max}(\alpha^*,1)$. We can then choose some $\epsilon_1 > 0$ such that $C^* + \epsilon_1 < C_{max}(\alpha^*,1)$, so that CRED continues to hold, but $R(\alpha^*,1,C^* + \epsilon_1) > 0$. It is now possible to choose ϵ_2 such that $(\alpha^* + \epsilon_2,1,C^* + \epsilon_1)$ satisfies both IC and CRED with $\alpha^* + \epsilon_2 < 1$. But utility from the latter contract exceeds $\bar{U}(\alpha^*,1,C^*)$, contradicting optimality. Hence, CRED holds with equality at an optimum.

To prove (iv), observe that $\bar{U}(\alpha,1,C)$ is stricly increasing in α and independent of C. Hence, if $(\alpha^*,1,C^*)$ is optimal, there cannot be any contract $(\alpha',1,C')$ with $1 \geq \alpha' > \alpha^*$ such that IC and CRED hold with equality; otherwise the latter contract would be feasible and would dominate $(\alpha^*,1,C^*)$, contradicting optimality.

To prove (v), denote C_1 such that $R(\alpha^*, 1,C_1) = 0$ for $\hat{z} > 0$ and C_2 such that $R(\alpha^*,1,C_2) = 0$ for $\hat{z} = 0$. Recall that $\hat{z} > 0$ if and only $\Phi(\alpha^*)C_1 > v - \frac{\beta\Theta}{n}\alpha^{*\beta}(n-1)^{\beta-1}$

and $\hat{z} = 0$ if and only if $\Phi(\alpha^*)C_2 \le v\frac{\beta\Theta}{n}\alpha^{*\beta}(n-1)^{\beta-1}$. Therefore, $C_1 > C_2$. The objective function is independent of C when w = 1, but C_2 is feasible whenever C_1 is feasible, while the converse is not true. It is plausible to argue that the team prefers C_2 to C_1, because intuitively it is easier to enforce a small penalty than a big one, while both C_1 and C_2 generate the same level of utility. Hence, $C^* = C_2$ and $\hat{z}(\alpha^*,1,C^*) = 0$. From $R(\alpha^*,1,C^*) = 0$ and $\hat{z}(\alpha^*,1,C^*) = 0$, we have $C_0(\alpha^*,1) \ge C^* = C_i(\alpha^*,1)$.

(c) When $\delta = 0$, the IC condition requires that the utility from $\hat{z}$ must be identical to that from w, i.e. $w = \hat{z}$. This follows from the uniqueness of the maximizing $\hat{z}$, which follows from the strict concavity of $U(\alpha,w,C,z)$ in z. The remaining assertions follow from the analysis of the static model. This completes the proof of Theorem 11. Q. E. D.

Theorem 11 specifies three types of optimal labour contracts. Part(a) of Theorem 11 defines the optimal labour contract in a pure dynamic monitoring system. It is shown that with a high discounting factor, the team can achieve an efficient effort level without conducting costly labour supervision. Part(b) of Theorem 11 characterizes the labour contract where the discounting factor is strictly positive but not very high. In this case, informal mutual monitoring supplements rather than entirely replaces formal labour supervision. Part(c) of Theorem 11 describes the optimal labour contract in a pure static monitoring system. By summarizing those different contracts, Theorem 11 presents us a broad picture of labour supervision in a worker team.

We are now in the position to show that a feasible contract with w = 1 always dominates a feasible contract with w = 0 in this general mutual monitoring model.

Theorem 12: An optimal contract which satisfies the IC condition and the CRED constraint must yield positive utility.

Proof: Consider an optimal contract $(\alpha^*, 1, C^*)$ derived from part (b) of Theorem 11. When $\alpha^* < 1$, $w^* = 1$, $C^* > 0$, $\hat{z} = 0$, and $\frac{\gamma}{n}Y(\alpha^*,1,0) = C^*$, we have

$$\hat{U}(\alpha^*,1,C^*) = \frac{\Theta}{n}(\alpha(n-1))^{\beta} - d - \Phi(\alpha)C \tag{45}$$
$$= [\frac{\Theta}{n}(\alpha(n-1))^{\beta} - d] - \Phi(\alpha)\gamma[\frac{\Theta}{n}(\alpha(n-1))^{\beta} - d]$$
$$= (1-\Phi(\alpha)\gamma)[\frac{\Theta}{n}(\alpha(n-1))^{\beta} - d] > 0,$$

since $(1-\Phi(\alpha)\gamma) > 0$ and $C^* = \frac{\Theta}{n}(\alpha(n-1))^{\beta} - d > 0$. But from IC, $\bar{U}(\alpha,1,C) \geq (1-\delta)\hat{U}(\alpha,1,C)$; therefore, $\bar{U}(\alpha,1,C) > 0$. It follows that $\bar{U}(1,1,0) > 0$, because $\bar{U}(1,1,0) > \bar{U}(\alpha,1,C)$ for $\alpha < 1$ and $C > 0$ by Theorem 11. Q. E. D.

6.4.2 Existence of a Feasible Contract

In the following discussion, we will focus on the optimal contract specified by part(b) of Theorem 11. The optimal contract in a pure dynamic monitoring system defined by part(a) of Theorem 11 is ignored because it is more common in reality that informal mutual monitoring supplements rather than entirely replaces formal labour monitoring. The static labour contract defined by part(c) of Theorem 11 was already examined in chapter 4 and 5.

According to part(b) of Theorem 11, when $\delta^* > \delta > 0$, an optimal labour contract implies that $w = 1$, $\alpha < 1$, $C = C_i(1) > 0$ and $\hat{z} = 0$, and that IC and CRED hold with equality. Substitute the IC condition into the CRED constraint. The team optimization problem can be written as

$$\max_{\alpha} \quad \frac{\Theta}{n}(\alpha n)^{\beta} - d - v$$

$$\text{Subject to: } \gamma[\frac{\Theta}{n}(\alpha(n-1))^{\beta} - d] - \Phi(\alpha)^{-1}[\frac{\Theta}{n}(\alpha(n-1))^{\beta} - d - \frac{1}{1-\delta}(\frac{\Theta}{n}(\alpha n)^{\beta} - d - v)] = 0.$$

The issue of existence of an optimal labour contract then boils down to whether there is an optimal α. Denote the objective function as $\bar{U}(\alpha)$, and multiply the credibility constraint by $\Phi(\alpha)$ and then denote the expression as $F(\alpha) \geq 0$. $F(\alpha) \geq 0$ specifies the feasible set $M \equiv \{ \alpha: \alpha_{min} < \alpha < 1, F(\alpha) \geq 0 \}$.

Fact A: Assume A1-A3 hold. If Θ, γ and n are large, v and D are small but positive, and $0 \leq \delta < \delta^*$, the feasible set $M \equiv \{\alpha: \alpha_{min} < \alpha < 1, F(\alpha) \geq 0 \}$ is non-empty and compact.

Proof: 1.Nonemptiness: Choose some arbitrary fixed α_0 such that $\alpha_{min} < \alpha_0 < 1$, and then $\Phi(\alpha_0) > 0$. Take the largest values of Θ, γ, and n and the smallest values of v and D which are permitted by A1-A3. We have that $D = \Theta$, $v = \frac{\beta\Theta}{n}$ by A2 and A3, n is large but finite because A1 and $v > 0$, $\beta > 0$ and $\Theta > 0$, and $\delta < \delta^* < 1$. Hence,

$$(46) \quad F(\alpha_0) = [\gamma + \frac{\delta}{1-\delta}\Phi(\alpha_0)^{-1}][(\alpha(n-1))^{\beta} - 1] - \Phi(\alpha_0)^{-1}\frac{1}{1-\delta}[\beta - \alpha_0^{*}(n^{\beta}-(n-1)^{\beta})] \geq 0.$$

2.Closedness: $F(\alpha)$ is continuous on the interval $\alpha \in (\alpha_{min},1)$. Now consider the limiting cases when α approaches its boundary:

$$(i) \qquad \lim_{\alpha\to\alpha min}F(\alpha) = -\frac{1}{1-\delta}[\frac{\Theta}{n}\alpha_{min}^{\beta}n^{\beta} - d - v] < 0$$

and

$$(ii) \qquad \lim_{\alpha\to 1}F(\alpha) = [\frac{\Theta}{n}(n-1)^{\beta} - d - \frac{1}{1-\delta}[\frac{\Theta}{n}n^{\beta} - d - v] < 0$$

The first inequality holds because $\delta < 1$ and $\frac{\Theta}{n}\alpha_{min}^{\beta}n^{\beta} - d - v = U(\alpha_{min}^{1},1,0,1) < 0$, since $U(\alpha_{min}^{1},1,0,0) = 0$ and $U_z(\alpha_{min}^{1},1,0,z) < 0$ by Theorem 7. The second inequality holds given $0 \leq \delta < \delta^*$. Hence, all points in some neighbourhood of the boundary $[\alpha_{min},1]$ have $F(\alpha) < 0$. The boundary of the feasible set must lie entirely in the interior of $[\alpha_{min},1]$. Since $F(\alpha)$ is continuous, the feasible set M is a closed set.

3.Boundedness: Because $\alpha \in [\alpha_{min},1]$ is bounded, its interior is also bounded.

Q. E. D.

Theorem 13: There exists some feasible solution for the team's optimum if the conditions in Fact A hold.

Proof: The objective function $\bar{U}(\alpha)$ is continuous throughout $[\alpha_{min},1]$. Also, according to Fact A, given the conditions, the feasible sets M are non-empty and compact. From the

Weierstrass theorem, there exists a solution for the present team optimization problem.

Q. E. D.

Introducing a dynamic monitoring mechanism does not alter the fundamental fact that the team must possess sufficient rents in order to conduct effective labour supervision. It is straightforward that without rents, the team cannot punish a shirker, and workers are also unable to retaliate the defector in a repeated game.

6.4.3 The Optimal Solution and Comparative Statics

By part (b) of Theorem 11, when $0 \leq \delta < \delta^*$, the credibility constraint must be binding in an equilibrium. So, we can set up the Lagrangean as,

(47) $$L = \bar{U}(\alpha) + \mu Q(\alpha)$$

Where μ is the Lagrangean multiplier, and $Q(\alpha) = \Phi(\alpha)F(\alpha)$ is the credibility constraint. Maximizing L with respect to α and μ respectively, we get the first order conditions as follow.

(48) $$\frac{\partial L}{\partial \alpha} = \frac{d\bar{U}(\alpha)}{d\alpha} + \mu\frac{dQ(\alpha)}{d\alpha} = 0$$

(49) $$\frac{\partial L}{\partial \mu} = Q(\alpha) = 0$$

At the team equilibrium, we cannot have $\frac{dQ(\alpha)}{d\alpha} \geq 0$. Because $\mu > 0$ when the contraint is binding and $\frac{d\bar{U}(\alpha)}{d\alpha} > 0$, $\frac{\partial L}{\partial \alpha} = 0$ holds only if $\frac{dQ}{d\alpha} < 0$. The optimal α is thus given by the binding credibility constraint, along with the fact that $\frac{dQ(\alpha)}{d\alpha} < 0$ at the optimum.

Theorem 14: If the conditions of Fact A hold, the team optimization problem has a unique solution of the form

(50) $$\alpha^* = \alpha^*(\Theta, v, \gamma, n, d, \delta)$$

$$(51) \qquad C^* = \Phi(\alpha^*)^{-1}\{\frac{\Theta}{n}(\alpha^*(n-1))^{\beta} - d - \frac{1}{1-\delta}[\frac{\Theta}{n}(\alpha^* n)^{\beta} - d - v]\}.$$

Proof: First, the credibility constraint $Q(\alpha)$ has continuous partial derivatives with respect to α and all the parameters if Theorem 13 holds. Second, the derivative of $Q(\alpha)$ with respect to α is less than zero at the equilibrium. From the implicit-function theorem, $Q(\alpha)$ defines the unique optimal solution of α which is an implicit function of the parameters, and has continuous partial derivatives with respect to all the parameters.

The optimal penalty is obtained by substituting α^* into the IC conditions with equality. Since the optimal solution of α is unique, the optimal penalty is also unique. Any deviation from the penalty defined by (51) will violate either the credibility constraint or the no-shirking conditions.

Q. E. D.

Now we turn to discussing the relationships between the monitoring cost, $1 - \alpha$, and the parameters such as Θ, γ, d, n, v, δ, and the technical efficiency of monitoring. The results of a simple comparative static study are given below.

$$(1) \qquad \frac{\partial\alpha}{\partial\Theta} = -[\frac{1}{n}\alpha^{\beta}(\gamma(n-1)^{\beta} + \Phi(\alpha)^{-1}(n^{\beta-1}\beta + \frac{\delta}{1-\delta}n^{\beta}))][\frac{dQ}{d\alpha}(\alpha^*)]^{-1} > 0$$

$$(2) \qquad \frac{\partial\alpha}{\partial\gamma} = -[\frac{\Theta}{n}(\alpha(n-1))^{\beta} - d][\frac{dQ}{d\alpha}(\alpha^*)]^{-1} > 0$$

$$(3) \qquad \frac{\partial\alpha}{\partial d} = [\gamma+\Phi(\alpha)^{-1}\frac{\delta}{1-\delta}][\frac{dQ}{d\alpha}(\alpha^*)]^{-1} < 0$$

$$(4) \qquad \frac{\partial\alpha}{\partial n} = \{\frac{\Theta}{n}\alpha^{\beta}[\gamma(n-1)^{\beta}(\frac{1}{n} - \frac{\beta}{n-1})] + \Phi(\alpha)^{-1}\Theta n^{\beta-2}\alpha^{\beta}[(2-\beta)\beta/n + \frac{\delta}{1-\delta}(1-\beta)]\}[\frac{dQ}{d\alpha}(\alpha^*)]^{-1} < 0, \quad \text{if } 0 < \beta < \frac{n-1}{n}$$

$$(5) \qquad \frac{\partial\alpha}{\partial v} = \frac{1}{1-\delta}[\Phi(\alpha)^{-1}][\frac{dQ}{d\alpha}(\alpha^*)]^{-1} < 0$$

$$(6) \qquad \frac{\partial\alpha}{\partial\omega} = -[\Phi(\alpha)^2]^{-1}[v - \frac{\beta\Theta}{n}\alpha^{\beta}n^{\beta-1} - \frac{\delta}{1-\delta}(\frac{\Theta}{n}(\alpha n)^{\beta} - d - v)]\frac{\partial\Phi}{\partial\omega}[\frac{dQ}{d\alpha}(\alpha^*)]^{-1} > 0,$$

where ω is a shift parameter that increases the effectiveness of monitoring, so that $\frac{\partial\Phi}{\partial\omega} > 0$.

$$(7) \qquad \frac{\partial\alpha}{\partial\delta} = -\Phi(\alpha)^{-1}\frac{1}{(1-\delta)^2}[\frac{\Theta}{n}(\alpha n)^{\beta} - d - v][\frac{dQ}{d\alpha}(\alpha^*)^{-1}] > 0.$$

The impact of changing the parameters on the team's monitoring cost in the dynamic model can be easily interpreted. While $\frac{\partial\alpha}{\partial\Theta}$ has the same positive sign as in the static case, its value is bigger due to the additional term of $\frac{\delta}{1-\delta}n^{\beta}$. A rise in Θ, i.e. increasing the price of the team product, or/and improving production efficiency, allows the team to increase both its static enforceable penalties and the dynamic informal punishment. As a result, the team is able to allocate more resources from monitoring to production without violating the no-shirking conditions.

$\frac{\partial\alpha}{\partial\gamma}$ is the same as in the static case. Because a change in γ affects α through changing the size of the penalty, it has no impact on the dynamic informal punishment. Again a change in the proportion of the income distributed based on labour performance, γ, has a positive impact on the team's productive time and a negative effect on the ratio of time spent on monitoring. The high proportion of the need distribution limits the team's ability of penalizing free-riders. As the proportion $(1-\gamma)$ is raised, the magnitude of the penalties on a shirker is curtailed. Consequently, the team has to reallocate the productive labour input to monitoring in order to reduce the effect of deline in the feasible penalties through improving information accuracy. A rise in the proportion of the needs distribution $(1-\gamma)$ will thus increases the monitoring cost of the team.

A change in production cost has a larger negative impact on α in the dynamic model than in the static case. This happens because a rise in d, i.e. increasing production cost, implies a lower net income available for distribution, permitting not only a smaller static penalty but also a lower dynamic punishment. As a result, the team has to invest more resources in formal labour monitoring to increase the chance of penalizing shirkers.

A change in the size of the team also has a larger positive effect on monitoring cost, $1-\alpha$, and a negative effect on the allocation of productive labour input in the dynamic situation. This occurs because an increase in the team size not only dilutes the negative impact of shirking on the shirker himself, but also reduces the average product of labour. This will curtail both the feasible static penalty and the dynamic informal punishment. In consequence, the team has to increase its monitoring input to control the internal incentive problems.

A change in the disutility of effort, v, has a larger negative impact on α in the dynamic case. An increase in the opportunity cost of productive work and monitoring, v, enhances the workers' motives for shirking both in the current period and in the future. To prevent workers from shirking, the team must reduce its productive labour input, and then raise monitoring cost.

A change in the efficiency of monitoring has a smaller impact on α in the dynamic case than in the static situation. This is consistent with the fact that in a team with a long time horizon, formal labour supervision is suplemented by informal mutual monitoring, and therefore it plays a less significant role in labour control than in a one-shot game. An improvement on the efficiency of monitoring technology allows the team to reduce the required monitoring input, and then increases productive input for any given intensity of monitoring and amount of the feasible penalties.

An increase in the discounting factor δ implies that workers appreciate their future welfare more than before, or/and that the chances the workers react to a defector rise. This in turn aggravates the severity of the penalties in the repeated game. As the dynamic self-enforcing mechanism becomes more effective, the first best team production can be retained with the utilization of less monitoring input.

6.5 Conclusion

In this chapter a model of dynamic mutual monitoring in a worker team has been developed. By integrating informal mutual monitoring with formal mutual monitoring, the model provides a comprehensive approach to the issues of labour supervision in a worker team. The pure dynamic monitoring system and the pure static monitoring system that have been discussed in the literature are special cases of this general monitoring model. Compared with established monitoring models in the literature, the present approach is argued to be more realistic, because it is observed that formal labour supervision is conducted in most production teams, while peer group pressure plays a part in disciplining workers in cooperative production.

The main results of this chapter are summarized as follows. It is demonstrated that in a single-period noncooperative game, the equilibrium is often inefficient , and obtaining the efficient level of effort can be quite costly. However, if the single-period game is repeated infinitely, the optimal level of effort can be attained by a production team with a low monitoring cost or even in the absence formal labour monitoring. An effective dynamic mutual monitoring system depends upon three factors. The first of these is the interest interdependence among workers which results from a residual sharing payment scheme. Only with joint interest, would the workers have a motive to keep an eye on their co-workers and care about peer group pressure from other workers. In this regard, residual-sharing teams have an institutional advantage over capitalist firms in the area of labour supervision. Another factor is institutional stability. High confidence in the survival of the team and low turnover, characterized by $\delta > 0$, is a necessary condition for the existence of dynamic mutual monitoring. The third factor is the possibility of economic rent. Unless the worker can be better off by following a cooperative rule than deviating from it, the workers are unable to penalize a defector by a future return to a noncooperative equilibrium. These characteristics of a dynamic mutual monitoring system provide additional insight into the reasons why labour supervision was a difficult task for Chinese production teams.

Chapter 7

Labour Supervision in a Chinese Commune (II)

7.1 Introduction

In this chapter the dynamic incentive system is examined with the same communal data that were used in chapter 5. The Chinese recognition of mutual interdependence among the individual members of a collective unit can be traced back to the early 1950s. In the collectivization movement, "only if I serve for other people, would others serve for me" was one of the popular slogans which the government proposed to inspire the Chinese peasants' collective consciousness. In the early 1960s, the Chinese government decided that the production teams, with 20-30 households, were the basic units of collective ownership and accounting, providing an institutional setting for group incentives. The Dazhai work point system was introduced to Chinese production teams in the middle of the 1960s. As discussed in chapter 2, one general aspect of the Dazhai renumeration system was that individual workers' performance was subject to workers' mutual assessment at public meetings. This practice minimized potential information asymmetry among workers and made individual workers' performance a matter of public knowledge for all the team members.

Although having attempted to reduce the difficulty of labour monitoring in the production teams through the exploitation of group-individual interaction, the Chinese government seemed to have not realized that voluntary cooperation among individual peasants resulting from their mutual interdependence was vulnerable to institutional instability. Its frequent switch of policy regimes and endless rural institutional adjustments had serious consequences for the incentive system of the Chinese production teams. To examine the economic and political constraints on Chinese collective farms' incentives, we will in the following study focus on the period 1970-76, when China was undergoing the Cultural Revolution. The empirical investigation of this chapter thus parallels that of chapter 5.

The remainder of the chapter is organized as follows. The next section presents the derivation of the estimating equation and the hypotheses regarding dynamic mutual monitoring in Chinese production teams. The third section discusses the results of estimating the model. The final section summarizes the findings of the chapter.

7.2 Estimating Equation

In section 6.4.3, we develop a team's optimal time allocation equation,

$$(1) \qquad \gamma[\frac{\Theta}{n}(\alpha(n-1))^{\beta} - d] - \Phi(\alpha)^{-1}[\frac{\Theta}{n}\alpha^{\beta}(n-1)^{\beta} - d - \frac{1}{1-\delta}(\frac{\Theta}{n}\alpha^{\beta}n^{\beta} - d - v)] = 0.$$

To compare the present dynamic model with the static model developed in chapter 4 and estimated in chapter 5, we write (1) as

$$(2) \qquad \gamma[\frac{\Theta}{n}(\alpha(n-1))^{\beta} - d] - \Phi(\alpha)^{-1}[v - \frac{\beta}{n}\Theta\alpha^{\beta}n^{\beta-1} - \frac{\delta}{1-\delta}(\frac{\Theta}{n}\alpha^{\beta}n^{\beta} - d - v)] = 0.$$

(2) is derived by rearranging the terms in (1) and using the fact that $n^{\beta} - (n-1)^{\beta} = \beta n^{\beta-1}$ when n is large. [51] Comparing (2) with equation (1) in chapter 5, they differ only through the term involving $\frac{\delta}{1-\delta}$ due to the different specification of IC in a dynamic model. When $\delta = 0$, we are back to the original static model estimated in chapter 5. The following discussion parallels that in chapter 5, but is repeated here for easy reference.

To estimate this equation, the functional forms of production function and monitoring technology must be specified. As in chapter 5, a Cobb-Douglas agricultural production function is assumed,

$$(3) \qquad Q = \lambda_0 L^{\lambda_1} K^{\lambda_2} T^{\lambda_3} F^{\lambda_4}$$

Here Q is gross output in value terms. $L = \alpha n w$ is labour input. By part (b) of Theorem 11 in chapter 6, the optimal level of productive effort is $w = 1$, so $L = \alpha n$. K, T and F are

[51]According to the data and the estimated β n = 83 and beta = 0.70. The difference between the two sides of the equality lies only after the third decimal point.

capital stock, sown area and current farm expenditures, respectively. λ_0 is the product of the price of the output and the technical efficiency of production, and λ_i, i = 1,2,3,4 define output elasticities. Thus we have, $\Theta = \lambda_0 K^{\lambda_2} T^{\lambda_3} F^{\lambda_4}$, and $\beta = \lambda_1$. Assuming that the production function displays constant returns to scale, i.e. $\Sigma_{j=1}^{4} \lambda_j = 1$, the production function at the optimal work effort w = 1 can be written as

$$q = \lambda_0 \alpha^{1-\Sigma\lambda} k^{\lambda_2} t^{\lambda_3} f^{\lambda_4}. \tag{4}$$

Here q, k, t, and f represent gross value of output, capital stock, sown area and current farm expenditures per worker, respectively, and $\Sigma\lambda = \lambda_2 + \lambda_3 + \lambda_4$. From equation (4), we obtain the identity,

$$\alpha = \{q/(\lambda_0 k^{\lambda_2} t^{\lambda_3} f^{\lambda_4})\}^{1/(1-\Sigma\lambda)}. \tag{5}$$

Define the monitoring technology as

$$\Phi(\alpha,\omega) = 1 - \exp[-\omega(1-\alpha)], \tag{6}$$

where $\omega > 0$ is the technical coefficient of monitoring. Substituting (3), (4), (5) and (6) into (2) gives us an estimating equation

$$\gamma_i(q_i\{(n_i-1)/n_i\}^{1-\Sigma\lambda} - d_i) - [1 - \exp(\omega((q_i/(\lambda_0 k_i^{\lambda_2} t_i^{\lambda_3} f_i^{\lambda_4}))^{1/(1-\Sigma\lambda)} - 1))]^{-1} \\ [v - q_i(1-\Sigma\lambda)/n_i - \beta(q_i - d_i - v)] = u_i, \tag{7}$$

where $\beta \equiv \frac{\delta}{1-\delta}$, i is the team index, u_i is a stochastic term, γ_i is the proportion of the income distributed according to performance, and d_i represents cost and saving deduction per worker. In equation (7), the λ_k, v, ω and β are the parameters to be estimated. q_i is the endogenous variable, and γ_i, d_i, k_i, t_i, f_i and n_i are exogenous variables. The justification of the exogeneity of n_i, k_i, t_i, f_i, d_i, and γ_i is given in Chapter 5. The variables involved in the estimation are defined in Chapter 5 as well. The only new parameter to be estimated is β, which indirectly gives an estimate of the discounting rate δ. When $\delta = 0$ and hence $\beta = 0$, equation (7) is identical to the estimating equation (8) in chapter 5.

By estimating equation (7), we are able to examine the hypotheses concerning labour supervision in Chinese production teams. The central issue concerning the dynamic monitoring model is whether informal mutual monitoring among workers plays an effective role in disciplining workers in team production. It is conjectured that the future reaction of the other workers had a positive impact on the individual workers' effort decisions. In other words, the discounting factor δ is statistically significant and greater than zero.

The hypotheses tested in chapter 5 will be re-examined in the dynamic setting. For easy reference, those hypothese are briefly re-stated as follows. First, it is asserted that a team faces free-riding problems. This hypothesis is constructed as

$$H_{01}: v - \{(1-\Sigma\lambda)/n\}\lambda_0\alpha^{1-\Sigma\lambda} k^{\lambda_2} t^{\lambda_3} f^{\lambda_4} = 0$$
$$H_{a1}: v - \{(1-\Sigma\lambda)/n\}\lambda_0\alpha^{1-\Sigma\lambda} k^{\lambda_2} t^{\lambda_3} f^{\lambda_4} > 0$$

where H_{a1} is consistent with free-riding and H_{01} is the null of no free-riding. Note the difference between H_{a1} and A3 in Chapter 6. H_{a1} presents a necessary and sufficient condition for free-riding, while A3 is sufficient but not necessary for the existence of free-riders in team production. The test for free-riding specified by H_{01} and H_{a1} is more powerful in discriminating the competing hypotheses than that given by A3.

Second, when $\delta^* > \delta \geq 0$,[52] the team must conduct formal labour supervision; therefore, both the resources invested in labour supervision and the probability of catching and penalizing a shirker in the team were greater than zero. That is,

---

[52] Recall that δ^* is the critical value of the discounting factor below which the optimal level of effort cannot be sustained without conducting formal labour monitoring.

H_{02}: $1 - \alpha = 0$

H_{a2}: $1 - \alpha > 0$

where H_{a2} implies non-zero monitoring effort and H_{02} is the null of no labour supervision. These hypotheses imply, respectively,

H_{03}: $1 - \exp(-\omega(1-\alpha)) = 0$

H_{a3}: $1 - \exp(-\omega(1-\alpha)) > 0.$

Third, the Chinese production teams were economically rational. Denote equation (2) as $Q(\alpha) = 0$. It is asserted that the derivative of the credibility expression $Q(\alpha)$ with respect to α at the team optimum is negative. That is,

H_{04}: $\frac{dQ(\alpha)}{d\alpha}|_{\alpha^*} = 0$

H_{a4}: $\frac{dQ(\alpha)}{d\alpha}|_{\alpha^*} < 0.$

Fourth, following the comparative static study conducted in section 6.4.3., we conjecture that the team's monitoring input $(1-\alpha)$ is negatively related to the price of output and/ or production efficiency, the proportion of the income distributed based on work, the efficiency of monitoring technology, and the discounting factor, and positively related to the cost and saving deduction of the team's gross revenue, the team size, and the disutility of production and monitoring, i.e. $\frac{\partial\alpha}{\partial\lambda_0} > 0$, $\frac{\partial\alpha}{\partial d} < 0$, $\frac{\partial\alpha}{\partial\gamma} > 0$, $\frac{\partial\alpha}{\partial n} < 0$, $\frac{\partial\alpha}{\partial v} < 0$, $\frac{\partial\alpha}{\partial\omega} > 0$, and $\frac{\partial\alpha}{\partial\delta} > 0$. The empirical specification of the derivative of the expression $Q(\alpha)$ with respect to α and the partial derivatives of α with respect to the parameters are given in Appendix 7.1.

7.3 Estimation and Results

The estimation was performed for each of seven cross-sectional data sets just as in Chapter 5. Equation (7) was estimated by the nonlinear two-stage least squares (NL2S) method using the Gauss-Newton algorithm in TSP version 4.1. The estimation procedure was outlined in Appendix 5.1. The NL2S estimator is consistent and is asymptotically normally distributed under certain assumptions.

The Koenker heteroskedasticity test was conducted for each regression. The test results are reported in Table 7.1. The hypothesis of homoskedastic variances of disturbances cannot be rejected at the 5% level of significance by in any of the regressions except 1976. To correct the inconsistency of the NL2S estimates of standard errors caused by the heteroskedasticity problem, the heteroskedasticity consistent variance-covariance matrix estimates of the NL2S are reported for the year 1976, while those reported for the rest of the regressions are the NL2S estimates.

The estimation results are presented in Table 7.1.[53] The estimated discounting factor δ is the focus of our attention in this chapter. The estimates of the discounting factor δ have the predicted sign. The estimated value of δ lies between 0.02 and 0.09. The estimates of δ are statistically insignificant for 1970,1971, and 1972, but significant at the 5% level for 1973 to 1976. It shows that the future reaction of the other workers had a significant impact on each worker's own effort decisions during 1973-76, while the workers cared little about the consequences of their behaviors in future periods during 1970-72. This variation in the workers' intertemporal perceptions can be explained by the change in the political climate facing the Chinese production teams during the Cultural Revolution period. In 1970, the leftist leadership of the CCP, represented by the Minister of Defence, Lin Piao, and the "Cultural Revolution left" launched a program of transferring the basic accounting units from production teams to brigades in order to establish advanced socialist production relations in China's rural area. Despite confronting the strong resistance of the Chinese peasants and local leaders, this policy imposed institutional uncertainty on Chinese collective farming units, and encouraged Chinese peasants to engage in short-run behavior. The leftist rural institutional experiments stopped at the end of 1971 with the fall of Lin Piao and the purge of his

[53] A test of constant returns to scale was attempted through the estimation of an unrestricted model. Unfortunately, the estimation of an unrestricted dynamic monitoring model was not able to reach its convergency. As a result, constant returns to scale was treated as a maintained hypothesis, and only the results from the restricted model are reported.

followers. The CCP leadership reassured peasants that the production teams were the basic accounting units, and emphasized that the principle of "distribution according to work" must be complied with by the teams. The new agricultural policies were enhanced by the rise of Deng Xiao-ping, who was rehabilitated in early 1973 as the acting Prime Minister until the winter of 1975. The resulting moderate agricultural policies gradually stabilized the rural situation. As a result, we observe an increase in the weight placed by workers on future outcomes during the period 1973-76.

To discuss the size of the repeated game effect, we calculated the critical value of δ for each regression using the estimated parameters and the sample mean of the variables involved. The empirical expression of δ^* is given in appendix 2 following this chapter. By part (a) of Theorem 11 in chapter 6, when $\delta \geq \delta^*$, formal labour supervision becomes redundant. The calculated δ^* is reported in Table 7.1. It ranges from a low of 0.033 to a high of 0.096, indicating that the required discounting factor needs not to be very high when sustaining the efficient level of effort in the absence of formal labour supervision. This peculiar outcome can be explained by a number of factors. First of all, the length of one period assumed in this model is not a week or a month, but a year. Obviouly, the longer the time interval is, the lower the discounting rate. The choice of a relatively long length of time interval such as a year ensures the accuracy of the assumption of perfect observability for the teams with more than eighty workers. It takes time to spread information to every worker in the team. Secondly, it has been argued in the literature that in a peasant society the need to secure subsistence is always the prodominant concern to most peasant families (Scott, 1976 and Popkin, 1979). If unilateral cheating on collective effort agreement could trigger universal laziness in the team and threaten the security of subsistence in future, the peasants would behave themselves to avoid this grave consequence, even if the likelihood of such an outcome was perceived not to be very high or/and the peasants put a low weight on future action.[54]

[54]Recall the discounting factor δ is defined as the product of the probability that the game is believed to be repeated in future and the worker's intertemporal preference.

Comparing δ with the calulated δ^*, we note that except for 1975, the estimated discounting factors are lower than the corresponding δ^*. These results suggest that although the repeated game effect is statistically significant and the required discounting rate is not too high, the size of the effect is still not big enough to sustain the efficient level of effort without conducting formal monitoring. A low value of δ could be explained in part by a low weight of individual workers on the future and in part by a high probability perceived by the peasants that their teams would be absorbed into brigade level accounting units. However, it could also be regarded as empirical evidence for the limita . of the dynamic model specification or of the empirical estimation permitted by the data available. Therefore, the quantitative aspect of the estimate should be interpreted with more caution.

Now we look at the estimated parameters of the production function. In the seven estimated cross-sectional regressions, all but three of the 35 parameters of the production function have the correct signs.[55] Among those, 22 are significant at the 1% level and 5 at the 5% level. The estimated output elasticity of capital in two instances and land in one has the "wrong" sign, but one of them is statistically insignificant. The unsatisfactory estimates for capital and land inputs may again be explained by the inadequate data using a stock measure of capital and land, while the theoretically correct measure is the service flow from the capital and land stock. Comparing these results with those in chapter 5, we note that the two sets of the estimates of parameters on production function are very close for the year 1970, 71 and 72 when the estimated δ is not significant. Nonetheless, there are apparent variations between the two sets of estimates for the year 1973, 74, 75 and 76 when the estimated δ is statistically significant. As in the static model, the estimated parameters on the production function vary over the cross-sections, indicating that some structural change in production technology may have taken place during the period 1970-76.

For the purpose of comparison, the marginal products of inputs are calculated and reported at the bottom of Table 7.1. The marginal product of each input was defined in Chapter 5. Only those with significant test statistics are calculated. As in chapter 5, those

[55] λ_1 is calculated from λ_r, $r = 2,3,4$.

estimates are computed at the sample means for output and input levels. The estimated marginal product of capital ranges from a low of 0.05 to a high of 0.10, with an average over the 3 estimates of 0.08. The estimated marginal product of land lies between 6.03 and 17.16 *yuan*, with an average over the 3 estimates of 12.[illegible] *yuan*. These estimates are slightly lower than those in chapter 5. The estimated marginal product of expenditures on current inputs shows a low value of 0.16 to a high of 0.60, and an average value over the 6 estimates of 0.42 *yuan*. As in chapter 5, the value of the marginal product of the current inputs fell far below the marginal factor cost, showing an excessive use of current farm inputs in China's farming teams.

From the coefficients of capital, land and current farm inputs, the output elasticity of labour can be calculated. It lies between 0.68 and 0.82 except for the outlier of 0.96. The estimated marginal product of labour shows a low value of 251.7, a high of 423.8, and an average of 353.5 *yuan* across seven regressions. The estimated marginal products of labour are in general higher than those in chapter 5. They are, therefore, higher than the average revenue net of non-labour cost per unit of labour which ranges from 256.4 to 306.3 *yuan* with an average of 272.2 *yuan* over the 7-year period, except for 1970 and 1971. It is again evident that Chinese peasants were not able to receive a full compensation for their marginal contribution due to resource misallocation.

Like those in chapter 5, the estimates of disutility and the technical coefficient of monitoring are quite satisfactory. Summarizing over all runs, these coefficents have the predicted signs. Six of the estimates of disutility v, and all the estimates of ω are significant at the 10% level or better. The estimated v lies between 11.07 and 24.46, and ω varies from 0.21 to 0.44. As in chapter 5, the results are fairly stable over the period. The size of the estimated v and ω are generally larger than those in chapter 5. However, the standard errors of the estimated v and ω appear to be larger than those in chapter 5 as well. By introducing δ into the model, we have to estimate more parameters with the same number of variables. This inevitably increases the burden of the estimation and reduces the efficiency of the estimates.

With the estimates of the parameters of the model, we can check the legitimacy of the assumptions made in A1-A3, Lemma 2, 4, and 8 in chapter 6. The empirical specifications of these assumptions are given in appendix 7.2. The calculated results are presented in Table 7.4. The results show that the assumptions about the relationship between the size of a team and the other parameters made in the derivation of the dynamic monitoring model are valid for all the regressions.

We now re-examine the hypotheses that were tested in chapter 5. Those hypotheses specified in Section 7.2 are formulated as nonlinear functions of the estimated parameters. The test statistics for these hypotheses are calculated at the mean values of the variables involved. The test statistic in each case has an asymptotic chi-squared distribution with 1 degree of freedom. The test statistics reported in Table 7.2 and Table 7.3 are the values of the standard normal variables. (See appendix 5.1 for details.) [56]

Let us first discuss the results reported in Table 7.2. The workers' motive to shirk in production is confirmed by the positive value of all the test statistics for the free-riding problem; 5 of the tests are significant at the 5% level or better. The hypothesis that labour supervision is zero in the team is rejected at the 5% level of significance in all cases. The estimated time fraction spent on production, α, is always less than unity, lying between 80 and 97 percent of the team's total working time. This result indicates that the production teams allocated 3 to 20 percent of their working time to labour supervision in order to prevent workers from shirking. [57] The estimated time fraction spent on monitoring appears to be higher for 1970 to 1972 than that for 1973 to 1976. Recall that the estimated δ are

---

[56]The statistical problems associated with the use of the mean value of the dependent variable q in the calculation of the test statistics and the test on a boundary of the parameter space were discussed in the footnotes of chapter 5.

[57]As in chapter 5, the test of the hypothesis that $\alpha = 1$ is also carried out for each individual estimated $\hat{\alpha}_i$, (see Table 7.2). The rejection rate of the hypothesis that $\alpha = 1$ against $\alpha < 1$ varies from 57% to 81% of the total $\hat{\alpha}_i$s for each cross-section, with an average of 65.1%, in contrast to 53% to 79% and an average of 64.1% in chapter 5. It again confirms the reliability of the test results for $\alpha = 1$ based on the mean values of the variables involved. Nevertheless, we are not able to ensure that every individual $\hat{\alpha}_i$ falls in the range from zero to one. About 8% to 34% of the calculated $\hat{\alpha}_i$ exceed one, with an average of 24.9%. A proportion of the sample points fails to satisfy the economic constraint $\alpha \leq 1$ indicating the limitations of the present model specification and empirical estimation.

insignificant for the period of 1970-72, but significant for the period of 1973-76. The pattern of the change in the time allocation indicates that an effective informal mutual monitoring system helped the teams reduce the cost of labour monitoring. More precisely, when workers cared little about others' future actions, the team had to spend 11 to 20 percent of labour resources on labour supervision. However, formal monitoring consumed only 3 to 7 percent of the working time when informal mutual monitoring became effective.

The estimated probability of catching and penalizing a shirker consistently has the correct sign; 6 of the estimates are significant at the 5% level or better. The estimated chance of catching and punishing a shirker who provides zero effort while the other workers work at $w = 1$ in a team is 1 to 5 percent. These results are consistent with the findings in chapter 5 that labour monitoring in team farming is not technically impossible, but is nonetheless imperfect and costly. There is a mild improvement on the estimated α which theoretically should be less than one. The estimated α for 1975 was equal to unity in chapter 5, while it is significantly less than one in the present model. As a result, the estimated monitoring probability for 1975, which was negative in chapter 5, has the correct sign now, although the estimate is not statistically significant in either case.

The test for the rationality of the teams and the comparative static results are reported in Table 7.3. Concerning the negativity of the derivative of the credibility expression $Q(\alpha)$ with respect to α, three of the estimated test values have the expected signs. All the tests are, however, statistically insignificant. This result provides only mild support for the economic rationality of the Chinese production teams in a dynamic setting. Since the negativity of the expression $Q(\alpha)$ with respect to α is a necessary and sufficient condition for the comparative static results to be valid, the controversial result of the test for $\frac{dQ(\alpha)}{d\alpha} < 0$ implies an ambiguous outcome in the test of the comparative static predictions. Three sets of the estimated partial derivatives have the correct sign. All the estimates are statistically insignificant. The elasticity of the time fraction spent on production with respect to the discounting factor was calculated for those cases where the estimated δ had the correct sign. It is found that if the change in δ had a significant effect on α, a 1% increase in δ would have resulted in 2.28% to 4.77% of reduction in working time spent on monitoring (an average of 3.12%).

In sum, the estimated results in this chapter are in general close to those found in chapter 5, showing the stability and robustness of the estimates. Allowing $\delta > 0$ leads to some mild improvements in the estimates of v, ω and α. Introducing an additional parameter into the model, however, apparently increases the burden of estimation and results in a reduction in efficiency. In consequence, the standard errors of the estimates of the dynamic model are generally larger than those found in the static model, and the tests for the team rationality and the comparative static results are not statistically significant.

7.4 Conclusion

A 1970's panel data set from a commune in Hebei province, China is used to test hypotheses regarding the impact of dynamic mutual monitoring upon the incentives of Chinese production teams. It is found that the recognition of mutual interdependence among individual workers had a statistically significant impact on the workers' effort decisions. The fear of retaliation could dampen the workers' motive for shirking in team production and thus reduce the difficulty facing a team in obtaining the optimal work incentives. However, an effective dynamic monitoring system requires that workers have a stable long-term relationship with one another; therefore, institutional instability could undermine the incentives of a production team. There is some evidence that this actually occurred during the period 1970-76. As a result, the repeated game effect found is not large enough to sustain the proper work effort without conducting formal labour supervision in Chinese production teams.

The hypotheses that were tested in chapter 5 are re-examined under the dynamic specification. Once again we find that (1) the teams faced free-rider problems; (2) the teams invested resources on monitoring to cope with the problem; (3) monitoring in group farming is not technically impossible, but is imperfect and costly. However, the test results of the team rationality and the comparative static predictions are inconclusive.

Table 7.1
THE NL2S ESTIMATION RESULTS
(Constant Returns to Scale)

	1970	1971	1972	1973	1974	1975	1976
λ_0	180.9 (7.00)a	352.6 (3.94)a	152.1 (1.88)b	795.9 (1.81)b	250.8 (3.32)a	184.6 (4.74)a	119.1 (3.87)a
λ_1	0.68 (12.50)a	0.82 (18.34)a	0.81 (11.46)a	0.96 (11.47)a	0.75 (16.44)a	0.76 (24.03)a	0.71 (16.37)a
λ_2	0.04 (1.98)b	-0.15 (-2.06)	0.03 (0.45)	-0.24 (-4.68)	0.04 (2.63)a	0.01 (0.26)	0.03 (1.74)b
λ_3	0.23 (5.99)a	0.18 (3.52)a	-0.06 (-1.22)	0.22 (2.97)a	0.16 (3.06)a	0.06 (1.69)b	0.02 (0.74)
λ_4	0.05 (6.25)a	0.15 (2.42)a	0.22 (3.74)a	0.06 (0.78)	0.05 (0.64)	0.17 (2.94)a	0.24 (4.56)a
v	11.07 (1.09)	13.74 (2.30)b	14.89 (1.46)c	24.46 (2.61)a	13.38 (3.03)a	22.68 (2.15)b	12.91 (2.81)a
ω	0.22 (1.55)c	0.27 (3.65)a	0.35 (2.47)a	0.44 (3.20)a	0.21 (3.66)a	0.33 (2.23)b	0.28 (3.54)a
δ	0.031 (0.67)	0.025 (0.96)	0.032 (0.68)	0.091 (1.77)b	0.037 (1.92)b	0.089 (1.65)b	0.041 (1.79)b
δ^*	0.033	0.035	0.034	0.096	0.039	0.088	0.042
KH	(1.84)	(12.34)	(4.77)	(6.76)	(4.80)	(2.94)	(18.33)b
MPK	0.10	---	---	---	0.09	---	0.05
MPT	13.30	11.38	---	17.16	14.70	6.03	--
MPF	0.16	0.48	0.60	---	---	0.37	0.49
MPL	251.7	314.8	366.9	423.8	378.9	389.3	349.2
APL	256.4	263.8	286.8	266.0	306.3	277.1	248.8

Notes: (1) This set of NL2S estimation results is obtained with the use of γ, n, k, t, f, and d as the instrumental variables.
(2) T statistics are reported in parentheses below the coefficient estimates.
(3) c denotes that the test is significant at the 10% level.
(4) b denotes that ....at the 5% level.
(5) a denotes that ....at the 1% level.
(6) δ^* is the critical value of δ below which the efficient level of effort cannot be sustained without conducting formal monitoring.
(7) KH is the Koenker heteroskedasticity test value.
(8) MP and AP stands for marginal and average product, respectively.
(9) λ_1 is calculated from λ_r, r=2,3,4.

Table 7.2
RESULTS OF THE HYPOTHESES-TESTING (I)

	1970	1971	1972	1973	1974	1975	1976
H_1	8.05	10.03	9.81	19.24	8.62	17.96	8.97
(z)	(0.81)	(1.68)b	(0.93)	(2.00)b	(1.94)b	(1.69)b	(1.95)b
α(mean)	0.85	0.80	0.89	0.94	0.97	0.97	0.93
H_2	0.15	0.20	0.11	0.06	0.03	0.03	0.07
(z)	(7.33)a	(8.76)a	(9.32)a	(3.92)a	(4.52)a	(2.04)b	(6.34)a
No.1	22	30	24	33	33	49	48
%	58	81	63	70	57	63	64
No.2	7	4	5	2	5	6	5
%	18	11	13	4	9	8	7
No.3	9	3	9	12	20	23	22
%	24	8	24	26	34	29	29
H_3	0.03	0.05	0.04	0.03	0.01	0.01	0.02
(z)	(1.97)b	(3.25)a	(2.93)a	(2.75)a	(2.63)c	(1.16)	(2.51)a

Notes: (1) The value of standard normal variables are reported in parentheses below the coefficient estimates.
(2) H stands for the test of free-riding in production.
(3) H_2 is the test for the hypothesis that $1-\alpha$ is equal to zero with the use of the mean values of the variables involved.
(4) No.1 is the number of teams having $\alpha < 1$ at the 5% level of significance.
(5) No.2 is the number of teams having $\alpha = 1$ at the 5% level of significance.
(6) No.3 is the number of teams having $\alpha > 1$ at the 5% level of significance.
(7) H_3 is the estimated value of the probability of catching and penalizing a shirker who provides zero effort while the other workers work at $w = 1$.

Table 7.3
RESULTS OF THE HYPOTHESIS-TESTING (II)

	1970	1971	1972	1973	1974	1975	1976
H_4	115.3	-37.87	-41.78	991.10	-490.08	1558.04	194.05
(z)	(0.43)	(-0.24)	(-0.10)	(0.95)	(-0.35)	(0.40)	(0.59)
P_{λ_0}	-0.03	0.03	0.11	-0.003	0.03	-0.02	-0.06
(z)	(-0.52)	(0.22)	(0.09)	(-1.63)	(0.36)	(-0.43)	(-0.60)
P_d	0.01	-0.02	-0.03	0.004	-0.01	0.01	0.01
(z)	(0.60)	(-0.22)	(-0.09)	(1.64)	(-0.33)	(0.43)	(0.67)
P_γ	-1.67	4.98	5.20	-0.20	0.42	-0.13	-0.96
(z)	(-0.43)	(0.24)	(0.10)	(-0.95)	(0.35)	(-0.40)	(-0.59)
P_n	0.03	-0.05	-0.09	0.004	-0.04	0.02	0.03
(z)	(0.45)	(-0.24)	(-0.09)	(1.18)	(-0.33)	(0.41)	(0.58)
P_v	0.28	-0.50	-0.66	0.04	-0.29	0.08	0.26
(z)	(0.36)	(-0.25)	(-0.10)	(0.93)	(-0.32)	(0.36)	(0.50)
P_ω	-2.34	9.81	5.30	-0.16	1.09	-0.18	-1.26
(z)	(-0.33)	(0.25)	(0.11)	(-0.60)	(0.41)	(-0.28)	(-0.49)
P_δ	-53.16	92.73	141.4	-8.42	59.67	-16.15	-48.19
(z)	(-0.36)	(0.25)	(0.10)	(-0.93)	(0.32)	(-0.36)	(-0.50)
size	38	37	38	47	58	78	75

Notes: (1) H_4 is the test for the negativity of the first derivative of the credibility expression $Q(\alpha)$ with respect to α.

(2) P_{λ_0}, P_d, P_γ, P_n, P_v, P_ω and P_δ are the estimates of the partial derivatives of α with respect to the corresponding parameters.

Table 7.4
THE ASSUMPTIONS OF THE DYNAMIC MONITORING MODEL

	1970	1971	1972	1973	1974	1975	1976
A1	235.6	267.9	261.7	220.1	221.5	208.3	213.7
A2	108.4	142.1	171.6	205.3	207.6	213.6	204.8
A3	6.07	8.27	8.21	18.75	7.49	17.01	7.76
L2	4.47	7.29	6.94	18.52	6.02	15.65	6.27
L4	0.01	0.001	0.01	0.07	0.01	0.06	0.02
L8	1.92	1.23	1.45	6.25	2.19	6.31	2.90

Notes: A1, A2 , A3, L2, L4 and L8 are defined in appendix 2. They are calculated at the sample means of the variables involved in each year. According to Assumption A1-A3 and Lemma 2, 4, 8 in chapter 6, A1 ≥ 0, A2 ≥ 0, A3 ≥ 0, L2 ≥ 0, L4 ≥ 0, and L8 ≥ 0.

Chapter 8

Conclusions and Policy Recommendations

The monitoring capability of a residual-sharing team is one of the major issues which divides economists interested in labour-managed firms. It has been argued that an income-sharing team faces potential free-rider problems because this type of firm cannot offer adequate incentives to monitors (Alchian and Demsetz, 1972) and/or subject agents to effective disciplinary power (Holmstrom, 1982). These internal incentive problems are considered to be multiplied in farming teams, because labour supervision is thought to be especially difficult in agricultural production (Bradley and Clark, 1972). This view is justified by pointing to the marginal status of producers' cooperatives in the free-market economy, as well as China's rural institutional reform and improved agricultural performance after 1978.

However, there is a growing body of econometric evidence showing that successful cooperatives enjoy higher productivity than similar capitalist firms in the West (Jones, 1982 and Svejnar, 1982). It has also been found that the incentive problems facing Chinese rural teams are frequently exaggerated in the literature, and the explanation of internal incentive problems merely by the technical characteristices of team farming is overly simplistic, when the issue of incentives is placed into an historical perspective (Lardy, 1983, Blecher, 1985, Putterman, 1985 and 1987, and Chapter 2 of this thesis). To obtain an appropriate assessment of the economic potential of residual-sharing teams in general and Chinese agricultural teams in particular, those sympathetic to workers' self-management have approached the issues of internal incentives from new angles.

Bradley and Gelb (1981) identify the existence of horizontal monitoring among workers as one central characteristic of producers' cooperatives. By formulating a model of informal mutual monitoring in a worker team, they depict the informational advantages of worker teams over capitalist firms in the area of labour supervision.[51] Bonin and Putterman (1987) and Putterman and Skillman (1988) point out the impact of the income distribution scheme on the efficacy of a team's labour supervision system, while assuming an ad hoc

[51]Mirrlees (1976), Cable and FitzRoy (1980a), Putterman (1984) and Ben-Ner (1988b) also discuss the possibility of mutual monitoring in a worker team.

approach to the payment rule in their analysis. Sen (1967) and Riskin (1974) stress the importance of group incentives in a residual-sharing team. Tyson (1979) and MacLeod (1984) discuss the implications of interest interdependence and long-term relationships among workers in the framework of a dynamic repeated game. Specialists in the study of the Chinese economy, such as Lardy (1983), Nolan (1983), Putterman (1985,1987) and Blecher (1985) address the economic, political and ideological constraints imposed on the incentive system of Chinese collective farms.

Drawing upon this progress in the literature, this thesis develops a comprehensive model to address the incentive problem of China's rural teams. The major innovation of the thesis is as follows. On the theoretical side, a static model of formal mutual monitoring has been developed in chapter 4. With this approach, we can go beyond the concept of "peer group pressure" to trace the consequences of mutual monitoring for workers' material incentives. The model endogenizes the determination of the size of penalties as well as the choice of monitoring intensity in the context of team optimization. This approach provides us with an analytical framework which can be used to evaluate the influence of the policy regime on internal incentives, and to highlight the incentive problems of China' rural teams from an historical and developmental perspective. Furthermore, in chapter 6, we integrate the processes of formal and informal mutual monitoring, and individual and group incentives, by extending the static mutual monitoring model to a production team with an infinite time horizon. This set-up is more realistic than established dynamic monitoring models in the literature, stressing how group incentives and long-term relationships among workers supplement rather than replace the static formal monitoring system.

Empirically, the static and dynamic versions of an implicit function determining the team's optimal allocation of labour time between production and monitoring were estimated in chapters 5 and 7. The estimated results appear quite satisfactory, providing empirical support for the highly abstract theoretical model of mutual monitoring. These chapters represent the first attempt in the labour supervision literature to estimate monitoring cost directly. Having accounted for the fraction of working time spent on labour monitoring, we correct a common measurement error where number of workers is used as a proxy for labour input in the estimation of a production function. It was also shown in chapter 7 that the long-term relationships among workers had a statistically significant impact on individual workers' effort decision-making.

To keep the model theoretically tractable and empirically manageable, the analysis employed a number of restrictive assumptions. First, the cost of labour contract negotiation was not addressed here. This cost is treated as a lump sum cost. It is assumed that once an agreement about job assignments and payment schemes has been reached among workers, it will be routinized every year. With a long time horizon, the set-up costs associated with annual contract negotiations are negligible.

Second, it is assumed that workers are identical. This assumption is justified by the fact that the team must standardize the contribution and rewards of potentially heterogenous workers, because the direct measurement of individual workers' marginal contribution to the team is almost impossible. (For details, see footnote 18 of chapter 4.) However, analytically, this means that the model cannot accommodate the effects of diverse worker preferences. Third, it is assumed that workers are risk neutral with respect to income and that the disutility of effort is linear. We ignore the possibility of risk averse preferences. Putterman and Skillman (1988) illustrate the possibility that if effort can only be measured with a large variance, risk-averse workers are likely to work harder, the poorer is the quality of monitoring and/or the lower is the monitoring investment. Thus, introducing a strictly concave utility and risk averse preferences could well strengthen rather than weaken the main theoretical results of the thesis, which rely on the finding that workers exert maximum productive effort under an optimal team contract.

On the empirical side, one of the major difficulties was the unavailability of data on effort and monitoring cost. It is extremely difficult, if not impossible, to estimate the two unknown variables simultaneously. One practical approach to this problem is to focus on one variable and keep the other from directly entering the estimation by making some assumptions. For instance, Putterman (1987, 1989) estimates Chinese peasants' response to material incentives through a production function approach. The concept of monitoring was not brought into the regression because the egalitarian distribution of income rather than the lack of monitoring was argued to be the major factor responsible for the incentive failure in Chinese rural teams. McMillan et al.(1989) directly measure the effort slack under the Chinese communal system by comparison with the effort made under the household responsibility system. The issue of monitoring in the communal system was not dealt with explicitly, since the focus of this study is the post-1978 rural reform era.

I have taken a different tack in this thesis by directly estimating the monitoring cost that is incurred when a team enforces an optimal labour contract with the maximum effort level $w = 1$. This approach is justified by the fact that the so-called incentive problem is essentially a matter of contract enforcement. It was shown in chapters 4 and 6 under fairly general conditions that an optimal labour contract will in fact have $w = 1$. As noted above, the results of Putterman and Skillman (1988) show that this conclusion is likely to be reinforced rather than weakened, if worker risk aversion is added to the model.

According to the theory developed in chapters 4 and 6, Chinese production teams can be classified into two categories: (1) prosperous teams which were able to conduct effective labour monitoring, and (2) poor teams where peasants struggled for subsistence and no feasible monitoring system could be implemented. The production teams in Dahe commune that we investigated in chapters 5 and 7 belong to the first category. However, the results obtained for such a "moderately prosperous" commune provide insight into the reason why many Chinese teams in backward areas suffered severe incentive problems. Without measuring effort directly, we were able to show in chapters 5 and 7 that monitoring was indeed costly in terms of foregone output. Poorer communes would necessarily be obliged to put more weight on monitoring activities relative to penalties for shirkers, and would therefore suffer a greater output loss due to imperfect work incentives. [59]

[59]The assumption that the labour contract with the maximum level of effort $w = 1$ was enforceable in relatively rich teams may be challenged by the fact that the peasants appeared to work more energetically in the post-1978 rural reform era. However, it is necessary to distinguish the labour slack caused by the lack of willingness and by the lack of opportunities. The first type of slack is called the

We also confronted a number of problems in the estimating and testing procedures. First, to test the null hypothesis that the team's monitoring investment was negligible,i.e. the time fraction spent on production $\alpha = 1$, we had to use the conventional $\chi^2(1)$ distribution as a rough approximation to the unknown distribution of the test statistics on a boundary of the parameter space. Secondly, having used a highly nonlinear estimating equation, we were unable to test for the accuracy of our assumed functional forms: a Cobb-Douglas production function, an exponential monitoring technology, and a linear utility of leisure. Thirdly, there is no guarantee that every individual $\hat{\alpha}_i$ falls in the economically meaningful range from zero to one. Solutions of these problems would be useful in a number of econometric applications, and should therefore be investigated in future research. Nonetheless, most of the results reported in chapters 5 and 7 are plausible in sign and magnitude, and seem unlikely to change substantially even with such refinements in the estimation procedures used.

The major results of the thesis and their policy implications can be briefly stated as follows. It was shown that Chinese production teams faced free-rider proiems which necessitated labour supervision. The institutional design of cooperative production provides worker teams with a fairly flexible monitoring structure. In principle, there is no institutional barrier that prevents a residual-sharing team from selecting the most efficient monitoring structure (central, mutual or a mix) according to the informational characteristics of the

[59](cont'd) incentive problem which is the subject of this study. The second type of slack is related to the problem of involuntary unemployment or underemployment. For the sake of tractability, the issue of unemployment was not addressed here. The post-1978 rural reform offered new economic opportunities to Chinese peasants with such policies as crop diversification, market liberalization and development of rural enterprises. These activities were something that Chinese peasants would be definitely eager to participate in, but were unable to engage in under the old policy regime. A good deal of the "increased work effort" observed after 1978 appears to be related to this expansion in economic opportunity through the relaxation of policy constraints, rather than a higher level of work intensity in previous cereal production, especially in such teams as those in Dahe commune.(For details, see Putterman 1988.)

The conjecture that peasants made reasonable effort in team production can be further justified by the economic achievement of Chinese agricultural collectives under a harsh external environment. Observing the experiences of Dahe commune, Butler (1985) raises a question, "Is there any other country in the world that could achieve, as it did in Dahe, a doubling of yields in six years, while rural producers actually earned less money?"

production process.[60] The shared residual claimancy also creates interest interdependence among workers and introduces the possibility of informal mutual monitoring. Informal mutual monitoring based on group incentives enables the teams to efficiently use intra-firm observable but non-verifiable information in the area of labour discipline.

In such an institutional setting, the technological characteristics of team farming play an important role in shaping the monitoring structure. But it is not the only factor that determines the failure or success of labour supervision in a farming team. This is in part because, at least theoretically, formal labour supervision is not always necessary. When workers have strong commitment to the team, an optimal level of effort can be enforced through workers' self-discipline. In any case, the empirical evidence suggests that verification of work performance in team farming may be costly, but it is not technically impossible.

An effective monitoring system is premised on the existence of rents. Without economic rents, the team cannot impose penalties on shirkers, and workers are also unable to retaliate against a defector in future periods. If shirkers were not punished, few people would care about monitoring. There is also a trade-off between required monitoring cost and the size of feasible penalties. The operation of a monitoring system can be very costly, when the economic environment and government policies depress the financial returns to the team and restrict the size of the penalties which can be imposed on shirkers.

The empirical results suggest that Chinese peasants in the rural teams were rational economic agents. Their practices were sensible responses to prevailing internal incentive problems, given the constraints they faced. However, under the extractive policy regime, state intervention, anti-incentive ideology and acute economic reality, labour supervision taxed away 2 to 20 percent of the team's working time.

Based on the findings of this thesis, it is plausible to argue that the observed problem of labour supervison and incentives in Chinese rural teams is not an intrinsic characteristic of cooperative production. There are indeed agency costs associated with team farming. Yet the

[60]The potential advantages of mutual monitoring in a farming team were discussed in chapter 3, although this thesis does not address the issue of choice of an optimal monitoring structure among the set of all feasible monitoring schemes.

agency costs incurred by Chinese teams would have been much smaller if the Chinese government had consistently adopted sound agricultural policies, provided correct price signals to the teams and encouraged material incentives for individual peasants. In this regard, this thesis supports the view advanced by Lardy (1983), Blecher (1985) and Putterman (1986,1987) that the key obstacles to China's agricultural development lie primarily not in the realm of rural institutions, but in the Chinese government's inappropriate development strategies and policies in the areas of pricing, crop planning and marketing, and anti-incentive ideology.

Although I have focused the analysis on China's agricultural sector, the approach developed above could be used to address the issue of incentives from a broader perspective. The central theme of this thesis is that the presence of rents is a precondition for the implementation of an effective incentive system. This is not only true for capitalist institutions (Shapiro and Stiglitz 1984 and Bowles 1985), but also for socialist institutions as well. In a more general sense, sufficient economic surplus is a necessary condition for economic reforms which aim at stimulating individual enthusiasm by raising the level of inequality (or potential inequality).

However, in an underdeveloped socialist country such as China, the size of the economic surplus available for implementing individual material incentives is limited by three crucial economic factors. First, China is a low income country; the size of the pie is small. Second, the basic human needs of China's mass population have somehow to be satisfied. The anti-incentive ideology should be condemned. But it would be inappropriate to press this argument too far. Penalizing shirkers, who often can hardly be discriminated from the incapable, by letting them starve lacks a moral basis and is bound to cause severe social unrest. However, welfare programs as simple as merely guaranteeing subsistence food consumption could consume a significant proportion of the income distributable to individuals in a poor nation. Third, China's economic development requires a huge amount of capital investment. The industrial sector must expand at a reasonable speed to provide employment outside of agriculture, so that the population pressure on farmland can gradually be relieved.

The rural infrastructure must also be developed to support farming modernization. Under China's current resource allocation system, the required investment funds cannot be raised without squeezing income distributable to individual workers. All these factors create an egalitarian tendency in China's income distribution practice.

The demise of teams may help circumvent the difficulty of monitoring effort in China's agricultural sector, but it does not change the fundamental nature of the incentive problem in China. This is because individual households' enthusiasm about farming is dictated by the profitability of crop production which is determined by the government's policies of pricing and crop planning. All these policies derive from government decisions about how to handle the relationships between capital investment and current consumption, and between industry and agriculture. Without proper attention to the macro-policy regime, household farming alone is unlikely to provide sustained growth in agriculture. This point is clearly evident from the stagnation of China's agricultural production after 1985.

Switching to household operations is obviously not a feasible option for modern industry. Could a capitalist solution ease the difficulty of providing adequate individual material incentives in China? The answer to this question indicated by the existing literature seems not particularly promising. It has been recognized that under capitalist institutions, the internal incentive problem induced dramatic labour-saving technological change in the western industrialized countries (Bradley and Clark, 1972). According to the efficiency-wage literature, supervision problems have also forced the employers of capitalist firms to pay high wages, so they can rely on the threat of involuntary unemployment to achieve a no-shirking equilibrium (Shapiro and Stiglitz 1984, and Bowles 1985). The price paid for a capitalist approach to the incentive problem is thus persistently high rates of involuntary unemployment. It is clear that such a solution would be very costly and painful for a country like China which is unfortunately endowed with exceedingly abundant labour and scarce capital and land.

All these factors make economic reforms which are designed to improve individual material incentives a difficult task for China. It is not surprising that China takes a more cautious approach to free-market-oriented economic reforms than the richer east European countries. Economists and politicians in the West excel in criticism of the conservative approach of the Chinese "hardliner" government to radical economic institutional changes. But insufficient attention has been paid to the economic constraints imposed on the economic reform process in China. It would be more constructive and helpful to provide the Chinese with an accurate estimate of how far and fast China could go if the Chinese government had a full commitment to a free-market-oriented reform, given these economic and social constraints.

With the approach developed by this thesis, we could draw some lessons from the experiences of China's rural teams. First of all, China must pursue a development strategy which properly balances not only industrialization and rural development, but also dynamic growth and static incentives. The issue of incentives is an important economic constraint on overall economic growth; however, it can only be dealt with properly when the economy has a sound development strategy. To have a balanced development strategy, Chinese policy-makers must have a realistic estimate of the nation's economic capability. China is a country which has a glorious past, but fell far behind the western countries in modern times. As a result, from its leadership to its people there is a strong aspiration for rapid development to catch up with the rest of the world. However, an unduly high rate of economic growth can only be sustained by a correspondingly high rate of capital accumulation. This inevitably squeezes the income distributable to individual workers, and thus undermines work incentives. A high target growth rate for industrialization will also put great pressure on agriculture and lead to stagnation of agricultural production. This is what happened in China during the Great Leap Forward, the Cultural Revolution and the post-1985 era. Thus, it is very important to choose an optimal rate of growth on the basis of realistic expectations about the various economic constraints the economy faces.

Secondly, China must continue to reform its economic system. The reform should be designed to establish product, factor and financial markets. This will provide correct price signals for firms and allow them to maximize the economic returns generated by their scarce resources. Within the firms, measures must be taken to encourage maximization of individual material incentives within the limits permitted by the economic and social constraints. It is plausible for rural cooperatives to have flexible forms of labour management according to their managerial ability, peasants' commitment to the cooperatives and the nature of production. In the areas where no feasible monitoring system can be devised, individual household farming makes sense and should be encouraged.[61]

However, it is necessary to help individual households in poorer areas establish some collective decision-making mechanisms in order to improve rural infrastructure, establish rural enterprises and provide basic social services such as primary education, health care and family planning. In those areas, collective decision-making can be more efficient than individual decision-making in terms of mobilizing abundant labour and pooling scarce financial resources for local development. The local self-reliance program invented by Chinese collective farms has not lost its significance for third world countries. Many developing country governments are struggling to cut budget deficits and to service heavy foreign debts; they are unable to provide the poor with adequate financial aid. The local self-reliance development programs, together with possible financial assists from the government and international donors, may be a promising approach to breaking the vicious cycle of poverty in rural areas.

As far as the restructuring of economic institutions is concerned, China should encourage experiments involving various types of firms such as those run by the state, workers as a group, capitalists and private individuals. Any type of economic institution has its own advantages and disadvantages. The choice of the institutional design for firms should be based on the technological characteristics of the industry, the local factor endowments, and the stage of economic development, rather than ideological faith in uniquely "best" solutions. It is hoped that this thesis can help correct the bias against residual sharing teams in the

[61]China's rural institutional reform introduced a fairly diversified pattern of farm management. It is reported that of Chinese production teams, 5% (probably containing nearly 10 million peasant households) continue to act as basic production units with unified production plans, accounting and distribution, while roughly 60% adopted a system of two-tier management with both household and cooperative dimensions, and the remaining 35% let peasant households be the basic units of production and accounting (Lu, 1987).

existing literature.

China needs a plausible, workable and socially responsible blueprint for its economic reform. The reform program should also be coordinated with a sound economic development strategy. It is impractical -indeed suicidal- to attempt a transition from the old system to a new one "overnight" in a poor and vast country like China which has nearly a quarter of the world population, 80% of whom live in rural areas, and a territory as large as Europe.

The third suggestion is that China should pay more attention to group incentives which can be used to supplement potentially inadequate individual material incentives. The promotion of cooperative behavior and group norms can reduce the likelihood of conflict and ease the implementation of individual material incentives. In terms of internal labour management, China can learn from Japanese firms. China has the cultural heritage of Confucius which advocates individual commitment to the family, local community and the state. Thus, it is easier to promote collectivism in China than in the West. China is also a less developed country; there are many natural barriers to labour mobility which could reduce labour turnover for firms. Low labour turnover helps establish a long term relationship among individual workers which has proven to be a necessary condition for an effective group incentive system.

In the 1950s, the Chinese leadership did correctly initiate an experiment with the implementation of group incentives. Unfortunately, this experiment was eventually used by leftists as a tool to attack individual material incentives, and therefore it was not very fruitful. Learning from these past mistakes, the implementation of group incentives must be based on full recognition of individual interests. Group incentive schemes should be used to supplement rather than replace individual material incentives.

The last conclusion is that China must uphold the policy of opening to the outside world to encourage foreign investment and to participate in international trade. International assistance can help China break the acute domestic economic constraints to speed up its economic reform and growth. Entering the world market and developing labour-intensive manufacturing industry may be the only hope for China's mass surplus labour in the rural sector. For China to gain the full acceptance of the international community, it has a lot of catching-up to do in the area of economic and political reforms. Nevertheless, since the program of economic reform and opening to the outside world is a very difficult task, China needs international encouragement and support. One of the major obstacles to China's joining the world is the lack of understanding of the complexity of Chinese economic problems in the West. The misunderstanding introduces the danger that China might be penalized for its alternative approach to domestic problems by the western powers, even if this might be the only feasible option for the nation. Economic sanctions are unlikely to change China's course, yet they definitely slow down the process of modernization and democratization in China.

The lack of understanding of China is not the fault of economists. China has been isolated from the rest of the world for so long. It is governed by a different political system and remains at a different stage of economic development from the western industrialized nations. M[illegible] problems have often been presented by the Chinese using the terminology of Ma[illegible] an inevitable but unfortunate consequence, those problems have sometime bee[illegible]ted by the West in the context of ideological confrontation. This thesis documents the story of Chinese rural teams using neoclassical analytical tools. It is hoped that this approach can be interpreted as a sincere attempt to promote mutual understanding between the two sides of the world.

References

Alchian, A and Harold Demsetz, 1972, "Production, Information costs, and Economic Organization," **American Economic Review 62**, 777-795.

Adler, S., 1957, **The Chinese Economy**, New York: Monthly Review press.

Amemiya, T. 1985, **Advanced Econometrics**, Oxford: Harvard University Press.

Aumann, R. J., 1983, "Survey of Repeated Games" in Aumann et al. eds "**Essay in Game Theory**", Boulder: Westview Press.

Avner Ben-Ner, 1988a, "Comparative Empirical Observations on Worker-Owned and Capitalist Firms",**International Journal of Industrial Organization 6**, 7-31.

……, 1988b, "The Life Cycle of Worker-owned Firms in Market Economies: A Theoretical Analysis", **Journal of Economic Behavior and Organization 10**, 287-313.

Binswanger, Hans P and Mark R. Rosenzweig, 1984, **Contractual Arrangements, Employment and Wages in Rural Labour Markets in Asia**, Yale University Press, New Haven,CT.

Binswanger, Hans P., 1974, "The Measurement of Technical Change Biases with Many Factors of Production", **American Economic Review 64**, 964-976.

Marc Biecher, 1985, "The Structure and Contradictions of Productive Relations in Socialist Agrarian Reform: A Framework for Analysis and The Chinese Case",**The Journal of Development Studies 22**, 104-12.

Bonin, J. P., 1977, "Work Incentives and Uncertainty on a Collective Farm", **Journal of Comparative Economics 1**, 77-97.

Bonin, J. P. and L. Putterman, 1987, **Economics of Cooperation and Labour-Managed Economy**, New York: Harwood Academic Publishers.

Bonin, J. P. and Putterman, Louis, 1987, "Incentives and Monitoring in Cooperatives under Labour-Proportionate Sharing Schemes", **Working Paper No. 87-18**, Brown University.

Bowles, S, 1985, "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian models," **American Economic Review 75**, 16-36.

Bradley, M. and M. G. Clark, 1972, "Supervision and Efficiency in Socialized Agriculture," **Soviet Studies 23**, 465-73.

Bradley, Michael E., 1971, "Incentives and Labour Supply on Soviet Collective Farms", **Canadian Journal of Economics 4**, 342-352.

Bradley, Michael E., 1973, "Incentives and Labour Supply on Soviet Collective Farms: Reply", **Canadian Journal of Economics 6**, 438-443.

Bradley, K. and Gelb, A., 1981, "Motivation and Control in the Mondragon Experiment", **British Journal of Industrial Relations 19**, 211-231.

Browning, M. J. 1982, "Cooperation in a Fixed-membership Labour-managed Enterprise," **Journal of Comparative Economics 6**, 235-247.

Butler, S. B., 1985, "Price Scissors and Commune Administration in Post-Mao China" in **Chinese Rural Development: the Great Transformation** edited by William L. Parish, M. E. Sharpe Inc.

Buck, J. L., 1937, **Land Utilization in China**, Chicago: University of Chicago Press.

Bull, Clive, 1987, "The Existence of Self-enforcing Implicit Contracts", **Quarterly Journal of Economics 102**, 147-159.

Cable, John and Felix R. FitzRoy, 1980a,"Productive Efficiency, Incentives and Employee Participation: Some Preliminary Results for West Germany," **Kyklos 33**, 100-121.

....., 1980b,"Cooperation and Productivity: Some Evidence from West German Experience",**Economic Analysis and Workers' Management**, 14, 163-180.

Cable, John and Wilson N., 1989, "Profit-Sharing and Productivity: An Analysis of UK Engineering Firms", **The Economic Journal 99**, 366-375.

Calvo, Guillermo A. and Wellisz Stanislaw, 1978, "Supervision, Loss of Control, and the Optimum Size of the Firm", **Journal of Political Economy 86**, 943-952.

Cameron, Norman E., 1973a, "Incentives and Labour Supply in Cooperative Enterprises", **Canadian Journal of Economics 6**, 16-23.

Cameron, Norman E., 1973b, "Incentives and Labour Supply on Soviet Collective Farms: Rejoinder", **Canadian Journal of Economics 6**, 442-445.

Chao, K., 1970, **Agricultural Production in Communist China 1945-1965**, Madison: University of Wisconsin Press.

Chen chu-yuan, 1982, **China's Economic Development: Growth and Structural Change**, Boulder, Colorado: Westview Press.

China Handbook Editorial Committee, 1984 **Economy**, Beijing: Foreign Languages Press.

Chinese Statistical Bureau, **Chinese Statistical Yearbook** , 1984, 1985 and 1987.

CCP, Document, "Regulations on the Work of the Rural People's Commune", September 1962.

CCP, Document No. 1 on Rural Economic Policy, Oct. 5, 1979, **United States FBIS Daily Report, the People's Republic of China**.

Chinn, Dennis L., 1979,"Team Cohesion and Collective-labour Supply in Chinese Agriculture", **Journal of Comparative Economics** 3, 375-394.

........, 1977, "Land Utilization and Productivity in Prewar Chinese Agriculture: Preconditions for Collectivation", **American Journal of Agricultural Economics** 2, 559-564.

Conte, Michael A, 1979, "Comments" On Tyson's paper, **Journal of Comparative Economics** 3, 301-303.

Conte, M. and Svejnar, J., 1981, "Productivity Effects of Participatory and Worker-Ownership Schemes, Profit Sharing, and Trade Unions in US Manufacturing", Paper presented at the 1981 North American Meetings of the Econometric Society, Washington, DC, December 1981.

Jacques Cremer, 1982, "On the Efficiency of a Chinese-Type Work-Point System", **Journal of Comparative Economics 6**, 343-352.

......., 1986, "Cooperation in Ongoing Organizations", **The Quarterly Journal of Economics** February, 33-49.

Defourney, J., Estrin, S., and Jones, D. C., 1985, "The Effects of Workers' Participation on Enterprise performance", **International Journal of Industrial Organisation 3**, 197-217.

Dernberger, R. F. 1982, "Agriculture in Communist Development Strategy", in **The Chinese Agricultural Economy** edited by Randolph Barker and Radha Sinha with Beth Rose, Boulder, Colo.: Westview Press.

Dernberger, R. F. and David Fasenfest, 1978, "China's Post-Mao Economic Future", A

Compendium of Papers for the Joint Economic Committee, Congress of the USA, November 9, 1978, US Government Printing Office, Washington D.C.

Dittrich, Scott, R. and Ramon H. Myers, 1971, "Resource Allocation in Traditional Agriculture: Republican China, 1937-1940, **Journal of Political Economy 79**, 887-896.

Domar, Evsey D., 1966, The Soviet collective farm as a producer cooperative, **American Economic Review 56**, 734-757. .

Domes Jurgen, 1980, **Socialism in the Chinese Countryside**, Montreal: McGill-Queen's University Press.

Dow, G. K., 1986, "Control Rights, Competitive Markets, and the Labour Management Debate", **Journal of Comparative Economics 10**, 48-61.

Dow, G. K., 1987, "The Function of Authority in Transaction Cost Economics," **Journal of Economic Behavior and Organization 8**, 13-38.

..tein, A., 1978, "The Chinese Development Model", A Compendium of Papers for the Joint Economic Committee, Congress of the USA, November 9, 1978, US Government Printing Office, Washington D.C.

Estrin, S., Jones, D. C., and Svejnar, J., 1987, "The Productivity Effects of Worker Participation: Producer Cooperatives in Western Economies", **Journal of Comparative Economics 11**, 40-61.

Eswaran, M. and Ashok Kotwal, 1984, "The Moral Hazard of Budget-Breaking," **Rand Journal of Economics** 15: 578-581.

FitzRoy, F. R. and Kraft, K., 1987, "Cooperation, Productivity, and Profit Sharing", **The Quarterly Journal of Economics**, February 1987, 23-35.

Friedman, J. W. 1986, **Game Theory with Applications to Economics**, New York: Oxford University Press.

Furubotn, E. 1976, "The Long-Run Analysis of the Labour-Managed Firm: An Alternative Interpretation." **American Economic Review 66**, 104-123.

......, and Pejovich S., 1973, "Property Rights, Economic Decentralization and the Evolution of the Yugoslav Firm, 1965-1972", **Journal of Law and Economics 16** (October), 275-302.

Fudenberg, D. and Eric Maskin, 1986, "The Folk Theorem in Repeated Games with Discounting and with Incomplete Information", **Econometrica 54**, 533-554.

Georgescu-Roegen, N., 1960, "Economic Theory and Agrarian Economics", **Oxford Economic Papers, Vol. 12**, 1-40.

Godfrey, L.G., 1987, **Misspecification Tests in Econometrics: the Lagrange Multiplier Principle and Other Approaches**, Econometric Society Monographs, Cambridge: Cambridge University Press.

Hsueh T-T and Liu P-W, 1980, "Collectivization of Chinese Agricultural Production 1952-1957: A land-Saving Organizational Change," **Journal of Comparative Economics 4**, 378-398.

Holly, A., 1985, "Specification Tests: An Overview" in **Advances in Econometrics Fifth World Congress Volume I**, Edited by Truman F. Bewley, Cambridge: Cambridge University Press.

Holmstrom, B. 1982, "Moral hazard in teams," **Bell Journal of Economics 13**, 324-340.

Hayami, Yujiro and Michio Kikuchi, 1983, **Asian Village Economy at the Crossroads: An Economic Approach to Institutional Change**, Baltimore: Johns Hopkins University Press.

Howe, H., Pollak, R. A. and Wales, T. J., 1979, "Theory and Time Series Estimation of The Quadratic Expenditure System", **Econometrica** Vol.47, No.5:1231-1247.

Israelsen, L. Dwight, 1980, "Collective, Communes, and Incentives," **Journal of Comparative Economics** 4:99-124.

Jarque,C.M. and Bera, A. K., 1981, "An Efficient Large-Sample Test for Normality of Observations and Regression Residuals," Manuscript, Australian National University, Canberra.

Jensen, Michael C. and William H. Meckling, 1979, "Rights and Production Functions: an Application to Labour-managed firms and Codetermination", **Journal of Business, Vol.52**, 469-506.

Jiang Jinyong, Luo Xiaopeng, 1989, "Changes in The Income of Chinese Peasants since 1978". **China's Rural Development Miracle with International Comparisons**, Edited by John W.

Longworth, Queensland: University of Queensland Press.

Jones, D. C. and Sveinar, J., 1985, "Participation, Profit Sharing, Worker Ownership and Efficiency in Italian Producer Cooperatives", **Economica 52**, 449-465.

Jones, D. C., 1982, "British Producer Cooperatives, 1948-1968: Productivity and Organizational Structure" in Jones D.C. and J.Svejnar (eds), **Participatory and Self-Managed Firms: Evaluation of Economic Performance**. Lexington, Massachusetts: Lexington Books, pp. 175-198.

Jones, D.C. and Backus, D., 1977, "British Producer Cooperatives in the footwear Industry: An Empirical Evaluation of the Theory of Financing", **Economic Journal 87**, 488-510.

Kueh, Y.Y., 1989, "The Maoist Legacy and China's New Industrialization Strategy", **The China Quarterly 119**, 420-[illegible]

Kueh, Y.Y., 1986, "Weather [illegible]les and Agricultural Instability in China", **Journal of Agricultural Economics 37**, 101-104.

Lardy, N.R. 1983, **Agriculture in China's Economic Development**, Cambridge, England: Cambridge University Press.

........, 1985, "Prices, Markets and the Chinese Peasants", in **Agricultural Development in the Third World**, Edited by Carl,K. Eicher and John, M. Staatz, Baltimore: Johns Hopkins University Press.

Lin, J. Y., 1987, "The Household Responsibility System Reform in China: A peasant's institutional Choice", **American Journal of Agricultural Economics 69**, 410-415.

Lin, J. Y., 1988, "The Household Responsibility System in China's Agricultural Reform: A Theoretical and Empirical Study", **Economic Development and Cultural Change** April, 199-224.

Lu Wen, 1987, "Development Strategy in Rural China" in **China's Rural Development Miracle** Edited by John W. Longworth, Queensland: University of Queensland.

Ma Lizheng, 1989, "What to eat? How to eat?" **Chinese Women**, June 1989.

MacLeod, W. 1984, "A Theory of Cooperative Teams" CORE discussion paper No.8441, University Catholique de Louvain.

MacLeod, W. 1987, "Behavior and the Organization of the Firm", **Journal of Comparative Economics 11**, 207-220.

Markusen, James R., 1975, Efficiency aspects of profit-sharing versus wage systems. **American Journal of Agricultural Economics 57**, 601-612.

Markusen, James R., 1976, "Profit-sharing, Labour Effort and Optimal Distributive Shares", **Economica 43**, 405-410.

McMillan J, Whalley J, and Zhu L-J, 1987, "Incentive Effects of Price Rises and Payment-System Changes on Chinese Agricultural Productivity Growth," NBER working paper No. 2148.

Mirrlees, James A., 1976, "The Optimal Structure of Incentives and Authority within an Organization", **The Bell Journal of Economics**, 7 105-131.

Myers, R. H., 1975, "Cooperation in Traditional Agriculture and Its Implications for Team Farming in the People's Republic of China", in **China's Modern Economy in Historical Perspective** , edited by D. H. Perkins, Stanford, Calif.: Stanford University Press.

Niu Ruofeng and Peter H. Calkins, 1986, "Towards an Agricultural Economy for China in a New Age: Progress, Problems, Response, and Prospects,"**American Journal of Agricultural Economics** 445-450.

Nolan Peter, 1988, **The Political Economy of Collective Farms**, Cambridge: Polity Press.

Nolan Peter, 1983, "De-Collectivization of Agriculture in China 1979-82: a Long-Term Perspective." **Cambridge Journal of Economics** 7, 381-403.

Nove, Alec., 1971, "The Decision to Collectivization", in **Agrarian Policies and Problems in Communist and Non-Communist Countries,** edited by W.A. D., Jackson, Seattle: University of Washington Press.

Oi, Walter Y. and Elizabeth Clayton, 1968, "A Peasant's View of a Soviet Collective Farm", **American Economic Review 58**, 37-59.

Parish, W. L., 1985, "Introduction: Historical Background and Current Issues", in **Chinese Rural Development: the Great Transformation** edited by Parish, W. L., M.E.Sharpe, Inc.

Perkins, D.H., 1980, "The Central Features of China's Economic Development" in **China's**

Development Experience in Comaprative Perspective, Edited by Dernberger, R. F. Cambridge, Massachusetts, and London, England: Harvard University Press.

Perkins, D. H., 196?, **Agricultural Development in China 1368-1968**, Edinburgh: University of Chicago Press.

......, and Yusuf S., 1984, **Rural Development in China**, Baltimore: Johns Hopkins University Press.

Pindyck, R. S., 1979, "Interfuel Substitution and the Industrial Demand for Energy: An International Comparison", **The Review of Economics and Statistics 61**, 169-179.

Popkin S., 1979, **The Rational Peasant**, Berkeley: University of California Press.

Pejovich, S., 1969, "The Firm, Monetary Policy and Property Rights in a Planned Economy", **Western Economics Journal** 7, no. 3 (September): 193-200.

Prybyla, J. S. 1978, **The Chinese Economy: Problems and Policies**, Columbia: University of South Carolina Press.

Putterman, L. 1980, "Voluntary Collectivization: A Model of Producers' institutional Choice", **Journal of Comparative Economics 4**, 125-157.

......., 1981, "On Optimality in Collective Institutional Choice", **Journal of Comparative Economics** 5, 392-402.

......., 1984, "On Some Recent Explanations of Why Capital Hires Labour", **Economic Inquiry 22**, 171-187.

......., 1985, "Extrinsic versus Intrinsic Problems of Agricultural Cooperation: Anti-incentivism' in Tanzania and China." **Journal of Development Studies 21**, 175-204.

......., 1987a, "The Incentive Problem and the Demise of Team Farming in China," **Journal of Development Economics** Vol.1: 103-127.

......., 1987b, "Productivity and Material Incentives in a Chinese People's Commune, 1970-1979", Working Paper No. 87-13, Department of Economics, Brown University.

........, 1988, "People's Republic of China: Systemic and Sructural Change in a North China Township", **American Journal of Agricultural Economics**, May 1988: 423-430.

......., 1990, "Effort, Productivity and Incentives in a 1970's Chinese People's Commune",

Journal of Comparative Economics 14 No. 1: 88-104.

......., 1991, "Effort Observability in Collective-Era China" Implications of McMillan, Whalley and Zhu," **American Economic Review** forthcoming.

......., and M. DiGiorgio, 1985, "Choice and Efficiency in a Model of Democratic Semi-Collective Agriculture", **Oxford Economic Papers 35**, 33-53.

Putterman L. and G. Skillman, Jr., 1988, "The Incentive Effects of Monitoring under Alternative Compensation Schemes", **International Journal of Industrial Organization 6**, 109-119.

Radner, R. 1986, "Repeated Partnership Game with Imperfect Monitoring and No Discounting," **Review of Economic Studies 53**, 43-57.

......, Myerson, R. and Maskin, E, 1986, "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," **Review of Economic Studies 53**, 59-69.

Rawski, T. G., 1982, "Agricultural Employment and Technology," in **The Chinese Agricultural Economy**, Edited by R. Barker and R.Sinha with Beth Rose.

Riskin, C. 1974, "Incentive Systems and Work Motivations. The experience in China," **Working Papers for a New Society** 1: 27-31, 77-92.

......, 1975, "Surplus and Stagnation in Modern China", in **China's Modern Economy in Historical Perspective** Edited by D. H. Perkins, Stanford, Calif.: Stanford University Press.

Scott James C., 1976, **The Moral Economy of the Peasant** , New Haven: Yale University Press.

Sen,A.K. 1966, "Labour Allocation in a Cooperative Enterprise," **Review of Economic Studies 30**, 361-371.

Sen, A.K., 1967, "Isolation, Assurance, and the Social Discount Rate", **Quarterly Journal of Economics 81**, 112-124.

Sen, A., 1981, "Market Failure and Control of Labour Power: towards an Explanation of 'Structure' and Change in India Agriculture, Parts I and II", **Cambridge Journal of**

Economics , Vol.5 nos 3 and 4, September and December.

Sertel, M., 1982, **Workers and Incentives**, New York: North-Holland.

Shapiro, C. and Joseph E. Stiglitz, 1984, "Equilibrium Unemployment as a Worker Discipline Device", **American Economic Review 74:** 433-444.

Shi Jingtang, 1987, *Zhongguo nongye hezuoshi ziliao* (Historical material of agricultural cooperativizatior in China), Beijing: Agricultural Cooperativization in China Editorial Office.

Svejnar, J., 1982, "Codetermination and Productivity: Empirical Evidence from the Federal Republic of Germany", in D. C. Jones and J. Svejnar (eds), **Participatory and Self-Managed Firms**. Lexington, Massachusetts: Lexington Books, pp. 199-212.

Tang, Anthony, 1980, "Food and Agriculture in China: trends and Projections, 1952-77 and 2000", in **Food Production in the Peoples's Republic of China,** edited by Tang, A.M. and Stone, B. May, Washington D.C.: International Food Policy Research Institute.

......., 1985, "A Critical Appraisal of the Chinese Model of Development", in **Agricultural Development in the Third World** edited by Carl, K., Eider and John, Stuatz, Baltimore: Johns Hopkins University Press.

Tyson, Laura D Andrea, 1979, "Incentives, Income Sharing, and Institutional Innovation in the Yugoslav Self-Managed Firm", **Journal of Comparative Economics** 3, 285-301.

Unger, J., 1985, "Remuneration, Ideology, and Personal Interests in a Chinese Village, 1960-1980", in **Chinese Rural Development: the Great Transformation**, edited by Parish, W. L. M.E. Sharpe, Inc.

Vanek, J., 1970, **The General Theory of Labour-Managed Market Economies**, Cornell University Press.

Wang Xin, 1984, "A Peasant Maps his Road to Wealth", **Bejing Review** No. 46 November 12.

Wang Guichen, Zhou Qiren and others, 1985, **Smashing the Communal Pot** (English), New World Press.

Ward, Benjamin, 1958, "The Firm in Illyria: Market Syndicalism" **American Economic Review 48**, 566-589.

Watson, A, 1983, "Agriculture Looks for 'Shoes that Fit': the Production Responsibility System and Its Implications", **World Development 11**, 705-730.

Wiens, T. B., 1982, "Technological Change" in **"The Chinese Agricultural Economy** Edited by Randolph barker and Radha Sinha with Beth Rose, Boulder, Colo.: Westview Press.

Wittfogel, K.A. 1971, "Communist and non-Communist Agrarian Systems with Special Reference to the USSR and Communist China: a Comparative Approach, in Jackson, W.A. D.(ed), **Agrarian Policies and Problems in Communist and Non-Communist Countries**, Seattle, University of Washington Press.

Wu Quengan and Liu Shuinian, 1985, **Brief History of Chinese Socialist Economy**, Heilongjiang People Press, Harbin China.

Xiao Meng, 1987, "Are China's Advanced Producers' Cooperatives Simply a Copy of Foreign Models ?", **Mongye Jingji Wenti (Agricultural Economic Studies)**.

Xiu Dixin, 1982, "Transformation of China's Economy" in **China's Search for Economic Growth** edited by Xiu Dixin and others, Beijing: New World Press.

Xue Xin 1982, "Scientific Socialism or Agrarian Socialism?", **Social Sciences in China (English) 3**, no. 1, 70-93.

Yan Ling, 1982, "The Necessity, Possibility and Realization of Socialist Transformation of China's Agriculture" **Social Sciences in China (English) 3**, no. 1, 94-124.

Zhang Liuzheng, 1980, "Developing Agricultural Production, Improving the People's Living Standard", **Nongye Jingji Wenti (Agricultural Economic Studies)**, no. 1.

Zweig, David, 1985, "Peasants, Ideology, and New Incentive Systems: Jiangsu Province, 1978-1981", in **Chinese Rural Development: the Great Transformation**, edited by Parish W. L., Armonk, New York, London, England: M.E.Sharpe, Inc.

Appendix 5.1

The empirical expressions of the comparative static results are given as follows.

(1)
$$\frac{dQ(\alpha)}{d\alpha} = (1-\Sigma\lambda)\alpha^{-1/(1-\Sigma\lambda)}q[\gamma((n-1)/n)^{1-\Sigma\lambda} + \Phi(\alpha)^{-1}\frac{(1-\Sigma\lambda)}{n}]$$
$$- \Phi(\alpha)^{-2}[v - \frac{(1-\Sigma\lambda)}{n}q]\omega\exp(\omega(\alpha-1)) < 0$$

(2)
$$\frac{\partial\alpha}{\partial\lambda_0} = [\frac{dQ(\alpha)}{d\alpha}]^{-1}[\frac{-q}{\lambda_0}(\gamma((n-1)/n)^{1-\Sigma\lambda} + \Phi(\alpha)^{-1}\frac{(1-\Sigma\lambda)}{n})] > 0$$

(3)
$$\frac{\partial\alpha}{\partial d} = [\frac{dQ(\alpha)}{d\alpha}]^{-1}\gamma < 0$$

(4)
$$\frac{\partial\alpha}{\partial\gamma} = [\frac{dQ(\alpha)}{d\alpha}]^{-1}[d - ((n-1)/n)^{1-\Sigma\lambda}q] > 0$$

(5)
$$\frac{\partial\alpha}{\partial n} = [\frac{dQ(\alpha)}{d\alpha}]^{-1}\{\gamma q((n-1)/n)^{1-\Sigma\lambda}[(\frac{1}{n} - \frac{(1-\Sigma\lambda)}{n-1}) +$$
$$\Phi(\alpha)^{-1}((1-\Sigma\lambda)(1+\Sigma\lambda)\frac{q}{n^2})]\} < 0 \qquad \text{if } \frac{n-1}{n} > 1 - \Sigma\lambda$$

(6)
$$\frac{\partial\alpha}{\partial v} = [\frac{dQ(\alpha)}{d\alpha}]^{-1}\Phi(\alpha)^{-1} < 0$$

(7)
$$\frac{\partial\alpha}{\partial\omega} = [\frac{dQ(\alpha)}{d\alpha}]^{-1}[(v - \frac{(1-\Sigma\lambda)}{n})q](\alpha-1)\exp(\omega(\alpha-1))\Phi(\alpha)^{-2} > 0$$

where $\Phi(\alpha) = 1 - \exp(-\omega(1-\alpha))$.

Appendix 5.2

We write the estimating equation in matrix algebra notation,

(1) $$f(y, Z, B_0) = u$$

where y is a T-vector of endogenous variables, Z is a T x K matrix of exogenous variables, B_0 is a G-vector of unknown parameters, and u is a T-vector of disturbances. We shall sometimes write $f(y, Z, B_0)$ simply as $f(B_0)$ or f. The nonlinear least squares (NLS) estimate of the vector $\hat{B}$ is that value of B_0 which minimizes the residual sum of squares,

(2) $$S(\hat{B}) = [f(\hat{B})'f(\hat{B})].$$

According to Amemiya (1985), an estimator B which is obtained by minimizing a random function S(B) over the parameter space is consistent if plim $T^{-1}S(B)$ is minimized at the true value B_0. This condition, however, does not hold for the NLS estimator $\hat{B}$. Taking a Taylor expansion of $f(\hat{B})$ at B_0, we have

(3) $$f(\hat{B}) = f(B_0) + F_*(\hat{B} - B_0),$$

where F_* is the T x G matrix the ith row of which is $(\partial/\partial B^*)f_i(B)$ evaluated at B^* between $\hat{B}$ and B_0. From (1), (2) and (3) we get

(4) $$T^{-1}S(\hat{B}) = T^{-1}u'u - \frac{2}{T}(\hat{B}-B_0)'F_*'u + T^{-1}(\hat{B}-B_0)'F_*'F_*(\hat{B}-B_0)$$
$$= A_1 + A_2 + A_3.$$

Now look at plim $T^{-1}S(\hat{B})$. First, plim $A_1 = \sigma^2_0$ by a law of large numbers. Second, if plim $T^{-1}F_*'F_*$ exists and is nonsingular, then plim A_3 is uniquely minimized at $\hat{B} = B_0$. Clearly, the consistency of $\hat{B}$ crucially depends upon the behavior of plim A_2. Since F_* is a function of the endogenous variable y, plim $T^{-1}F_{*}'u \neq 0$. In this case, one may find a $\hat{B} \neq B_0$ which minimizes plim $T^{-1}S(\hat{B})$. Hence, the NLS estimator $\hat{B}$ for equation (1) is generally inconsistent.

To obtain a consistent estimator, we apply the nonlinear two-stage least squares (NL2S) method to the model (1). The minimand of the NL2S estimator, $\tilde{B}$ is defined as

$$(5) \qquad S(\tilde{B}\,|\,W) = f'W(W'W)^{-1}W'f$$

where W is a T x M matrix of instruments with rank at least equal to G ($M \geq G$).

According to Amemiya (1985), if

(i) $f(y,Z,B_0)$ is continuous in y, Z and B_0 and at least twice continuously differentiable with respect to B_0;

(ii) the parameter space is compact;

(iii) u is i.i.d with zero mean and variance $\sigma^2 I$.

(iv) plim $T^{-1}(W'W)$ is nonsingular;

(v) plim $T^{-1}W'F$ exists and is of full rank;

then,

(a) $\tilde{B}$ converges in probability to the true value B_0,

and

(b) $\sqrt{T}(\tilde{B}-B_0)$ converges in distribution to $N[0, \sigma^2 Q^{-1}]$,

where F is the T x G matrix the ith row of which is $(\partial/\partial B)f_i(B)$, $Q^{-1} \equiv [\text{plim } T^{-1}F'P_W F]^{-1}$ and $P_W \equiv W(W'W)^{-1}W'$. σ^2 and Q^{-1} can be estimated by

$$(6) \qquad \tilde{\sigma}^2 = [u(\tilde{B})'u(\tilde{B})]/(T-G),$$

and

$$(7) \qquad \tilde{Q} = [T^{-1}F(\tilde{B})'P_W F(\tilde{B})]^{-1}.$$

The efficiency of the NL2S estimator depends upon the choice of instruments. The variance-covariance matrix of the NL2S can be minimized by choosing $W = EF(B_0)$, where $EF(B_0)$ is the conditional mathematical expectation of $F(B_0)$ (Amemiya, 1985). Since it is very difficult to derive the explicit expression for $EF(B_0)$, we approximate $EF(B_0)$ by the third order polynomials of all the exogenous variables of the model, i.e $W = (1, Z, Z^2, Z^3)$ (Kelejian, 1971).

Write the hypotheses proposed in Section 5.2 in a general form (Gallant and Jorgenson, 1979),

H_0: $h(B_0) = 0$

H_1: $h(B_0) \neq 0$.

Here $h(B_0)$ is a nonlinear function of the elements in B_0.

Assume that

(a) $\sqrt{T}(\tilde{B} - B_0)$ is distributed asymptotically as the normal with zero mean and a variance $\sigma^2 Q^{-1}$,

and

(b) $h(B_0)$ is a function with at least two continuous derivatives with respect to B_0,

then asymptotically, the Wald test statistic under the null hypothesis is

$$(8) \qquad h(\tilde{B})'[(\frac{\partial h}{\partial B_0})'(\tilde{\sigma}^2\tilde{Q})^{-1}(\frac{\partial h}{\partial B_0})]^{-1}h(\tilde{B}) \sim \chi^2(1).$$

The presence of heteroskedasticity does not affect the consistency of the NL2S estimator $\tilde{B}$ but changes its limiting distribution. Denote $E[uu'] = \Omega$. Given the conditions stated before, $\sqrt{T}(\tilde{B}-B_0)$ converges in distribution to $N[0, C]$, where C is the asymptotic variance-covariance of $\tilde{B}$. C can be estimated by

$$(9) \qquad \hat{C} = T(\tilde{F}'P_W\tilde{F})^{-1}(\tilde{F}'P_W\hat{\Omega}P_W\tilde{F})(\tilde{F}'P_W\tilde{F})^{-1}.$$

where $\hat{\Omega}$ is the NL2S estimates of Ω, and $\tilde{F} = F(\tilde{B})$. $\hat{C}$ is called the heteroskedasticity consistent variance-covariance matrix. $\hat{C}$ is used to calculated the standard errors of the estimates and construct the hypotheses-testing for 1973, when the hypothesis of homoskedastic variances of disturbances is rejected by the KH test.

When the proportion of the income distributed according to performance, γ, is endogenously determined, the estimating equation will be written as

$$(10) \qquad f(y, Y, X, B_0) = u$$

where Y is a T-vector of endogenous variables, X is a T x (K-1) matrix of exogenous variables, and $(Y, X) \equiv Z$ in equation (1). The NL2S estimator, $\tilde{B}$ defined by (5) is generally inconsistent because of the possible dependence of Z on u. A consistent estimator of B_0 can be obtained by redefining the instrument matrix. Denote $\bar{Z} \equiv (g, X)$, where g is per capita grain ration and replaces γ. The minimand of the NL2S estimator $\bar{B}$ is defined as

$$(11) \qquad S(\bar{B}\,|\,\bar{W}) = f'\bar{W}(\bar{W}'\bar{W})^{-1}\bar{W}'f.$$

Here $\bar{W} \equiv (1,\bar{Z},\bar{Z}^2,\bar{Z}^3)$. We write $P_{\bar{W}} \equiv \bar{W}(\bar{W}'\bar{W})^{-1}\bar{W}'$. Define the null hypothesis that γ is exogenously determined and its alternative as

$$H_0\text{: plim } T^{-1}Y'u = 0$$

$$H_1\text{: plim } T^{-1}Y'u \neq 0.$$

Under H_0 $\tilde{B}$ and $\bar{B}$ are both consistent, but $\tilde{B}$ is more efficient than $\bar{B}$ because the instrument matrix W contains more accurate information than $\bar{W}$ does, judged by the model specification. Under H_1 $\bar{B}$ is a consistent estimator while $\tilde{B}$ is not. Denote the vector of contrasts $\hat{\theta} = \bar{B}\text{-}\tilde{B}$, then the Hausman test is written as

$$(12) \qquad H = T\hat{\theta}'V(\hat{\theta})\hat{\theta}.$$

Since under H_0, $\tilde{B}$ and $\bar{B}$ are both consistent and asymptotically normally distributed, $\sqrt{T}\hat{\theta}$ is asymptotically distributed as $N[0, V(\hat{\theta})]$, where $V(\hat{\theta})$ is the asymptotic variance of $\sqrt{T}\hat{\theta}$. Because the consistent NL2S estimates $\tilde{B}$ and $\bar{B}$ can be written as

$$(13) \qquad \tilde{B} \simeq B_0 + (F'P_W F)^{-1}F'P_W u$$

and

$$(14) \qquad \bar{B} \simeq B_0 + (F'P_{\bar{W}}F)^{-1}F'P_{\bar{W}}u,$$

$V(\hat{\theta})$ can be estimated by

$$(15) \qquad \hat{V} = TV(\bar{B}) + TV(\tilde{B}) - \tilde{\sigma}^2 T(F'P_W F)^{-1}(F'P_W P_{\bar{W}}F)(F'P_{\bar{W}}F)^{-1}$$
$$- \tilde{\sigma}^2 T(F'P_{\bar{W}}F)^{-1}(F'P_{\bar{W}}P_W F)(F'P_W F)^{-1}$$

where $V(\tilde{B})$ and $V(\bar{B})$ are the estimated asymptotic variance-covariance matrix of $\tilde{B}$ and $\bar{B}$, respectively. $\tilde{\sigma}^2 = [u(\tilde{B})'u(\tilde{B})]/(T-G)$ is the estimated variance of disturbances under H_0, and $F = F(\tilde{B})$ is the estimated first order derivatives under H_0. Under H_0 the Hausman test has an asymptotic χ^2 distribution with G degrees of freedom (Holly, A., 1985).

Appendix 7.1

$\frac{dQ(\alpha)}{d\alpha}$, $\frac{\partial\alpha}{\partial\lambda_0}$, $\frac{\partial\alpha}{\partial d}$, $\frac{\partial\alpha}{\partial\gamma}$, $\frac{\partial\alpha}{\partial n}$, $\frac{\partial\alpha}{\partial v}$, $\frac{\partial\alpha}{\partial\omega}$, and $\frac{\partial\alpha}{\partial\delta}$ are specified as follows:

(1) $$\frac{dQ(\alpha)}{d\alpha} = [\gamma((n-1)/n)^{1-\Sigma\lambda} + \Phi(\alpha)^{-1}((1-\Sigma\lambda)/n + \beta)]q(1-\Sigma\lambda)\alpha^{-1} - [v - \frac{1-\Sigma\lambda}{n}q - \beta(q-d-v)]\omega\exp(\omega(\alpha-1))\Phi(\alpha)^{-2} < 0$$

(2) $$\frac{\partial\alpha}{\partial\lambda_0} = -[\gamma((n-1)/n)^{1-\Sigma\lambda} + \Phi(\alpha)^{-1}((1-\Sigma\lambda)/n + \beta)]\frac{q}{\lambda_0}[\frac{dQ(\alpha)}{d\alpha}]^{-1} > 0$$

(3) $$\frac{\partial\alpha}{\partial d} = [\gamma + \Phi(\alpha)^{-1}\beta][\frac{dQ(\alpha)}{d\alpha}]^{-1} < 0$$

(4) $$\frac{\partial\alpha}{\partial\gamma} = [d - q((n-1)/n)^{1-\Sigma\lambda}][\frac{dQ(\alpha)}{d\alpha}]^{-1} > 0$$

(5) $$\frac{\partial\alpha}{\partial v} = [1 + \beta]\Phi(\alpha)^{-1}[\frac{dQ(\alpha)}{d\alpha}]^{-1} < 0$$

(6) $$\frac{\partial\alpha}{\partial n} = \{[\gamma q((n-1)/n)^{1-\Sigma\lambda}(n^{-1} - (1-\Sigma)(n-1)^{-1})] + [(1-\Sigma\lambda)(1+\Sigma\lambda)\frac{q}{n^2} + (\Sigma\lambda)\beta\frac{q}{n}]\Phi(\alpha)^{-1}\}[\frac{dQ(\alpha)}{d\alpha}]^{-1} < 0$$

(7) $$\frac{\partial\alpha}{\partial\omega} = -(1-\alpha)\exp(\omega(\alpha-1))[v - \frac{1-\Sigma\lambda}{n}q - \beta(q - d - v)]\Phi(\alpha)^{-2}[\frac{dQ(\alpha)}{d\alpha}]^{-1} > 0$$

(8) $$\frac{\partial\alpha}{\partial\delta} = \frac{-1}{(1-\delta)^2}(q-d-v)\Phi(\alpha)^{-1}[\frac{dQ(\alpha)}{d\alpha}]^{-1} > 0$$

where $\Phi(\alpha) = 1 - \exp(\omega(\alpha-1))$ and $\beta = \frac{\delta}{1-\delta}$.

Appendix 7.2

The assumptions made in A1-A3 and Lemma 2, 4, and 8 are respecified as

(1) $$A1 = \lambda_0 k^{\lambda_2} t^{\lambda_3} f^{\lambda_4} ((n-1)/n)^{1-\Sigma\lambda} - d \geq 0$$

(2) $$A2 = (1-\Sigma\lambda)d - v \geq 0$$

(3) $$A3 = v - \frac{1-\Sigma\lambda}{n}(\lambda_0 k^{\lambda_2} t^{\lambda_3} f^{\lambda_4})^{1/(1-\Sigma\lambda)}(1/d)^{(\Sigma\lambda)/(1-\Sigma\lambda)} \geq 0$$

(4) $$L2 = v - (1+\Sigma\lambda)(\frac{1-\Sigma\lambda}{n})(\lambda_0 k^{\lambda_2} t^{\lambda_3} f^{\lambda_4})^{1/(1-\Sigma\lambda)}(1/d)^{(\Sigma\lambda)/(1-\Sigma\lambda)} \geq 0$$

(5) $$L4 = ((n-1)/n)^{1-\Sigma\lambda} \cdot (1-\delta) - \frac{1}{n-1} \geq 0$$

(6) $$L8 = \delta(n-1)[1 - (\Sigma\lambda)(\lambda_0 k^{\lambda_2} t^{\lambda_3} f^{\lambda_4}/d)^{1/(1-\Sigma\lambda)} n^{-1}] - (1 - \Sigma\lambda) \geq 0$$

The critical value of the discounting factor δ^* is defined as

$$\delta^* = [v - \frac{1-\Sigma\lambda}{n}\lambda_0 k^{\lambda_2} t^{\lambda_3} f^{\lambda_4}][\lambda_0 k^{\lambda_2} t^{\lambda_3} f^{\lambda_4}((n-1)/n)^{1-\Sigma\lambda} - d]^{-1}.$$