Order Number 9300330

Incentive payment systems and work effort in Chinese state enterprises

Li, Binsheng, Ph.D.

University of Hawaii, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

INCENTIVE PAYMENT SYSTEMS AND WORK EFFORT IN CHINESE STATE ENTERPRISES

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION
OF THE UNIVERSITY OF HAWAII IN PARTIAL
FULFILLMENTS OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
ECONOMICS

AUGUST 1992

by

Binsheng Li

Dissertation Committee:

Calla Wiemer, Chairperson
Sumner J. LaCroix
James A. Roumasset
John Bauer
David Yang

ACKNOWLEDGEMENTS

My principal debt is to professor Calla Wiemer and Professor Sumner La Croix for their insightful comments and their painstaking effort in correcting errors in my dissertation.

I am also indebted to professor James Roumasset, John Bauer, David Yang, and Thomas Rawski for their valuable comments on my dissertation.

I am very grateful to the Would Bank and Professor Gary Jefferson for granting me permission to use the data of the 396 Chinese state enterprises.

I thank my wife Huimin Xia for her support and understanding of my career.

ABSTRACT

Egalitarianism was a basic feature of the payment system adopted by Chinese state enterprises before the economic reforms. Because workers' incomes were virtually independent of their performances, their enthusiasm for working was seriously dampened.

Since 1978, a series of wage reforms have been conducted to link workers' payments with their performances. These reforms stimulated work incentives, but failed to achieve their full potential and caused some other problems. In the profit sharing system, when compensation is based on individual performance, the marginal return of work effort is shared by the state and the worker, while the marginal cost of work effort is borne by the worker, so the optimal work effort is below the socially optimal work effort. When compensation is based on collective performance, if transaction costs are high, non-cooperative workers may behave like free-riders, which undermines the effectiveness of the incentive system. This system also has a horizontal equity problem, and it is less able to offset risk. While compensating workers according to their disutility of work effort may solve some of the problems above, its effectiveness is undermined by the formation of worker-worker and/or worker-supervisor coalitions and is limited by the measurability of individual work effort. This study puts the profit sharing system and three alternative incentive systems into perspective within a

principal-agent framework, analyses their problems, and points out some directions for improvement.

Empirical tests of the effects of fixed wage, profit sharing, and bonus on value added and profit (both serve as proxies of work effort) are conducted using cross-sectional enterprise data. The following results are generated: i) work effort is independent of fixed wages; ii) work effort is positively related to bonuses, the effect of bonuses is, however, decreasing over time; and iii) work effort is also positively related to profit sharing; and the effect of profit sharing on work effort is smaller than that of bonuses.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

LIST OF TABLES ix

LIST OF FIGURES x

PREFACE xi

CHAPTER 1. INTRODUCTION 1

1.1 Research Questions 1

1.1.1 Why People Supply Work Effort 2

1.1.2 Alternative Ways to Improve Work Effort 5

1.2 Methodology and Approach 8

1.3 Structure of the Dissertation 13

CHAPTER 2. ENVIRONMENT OF CHINESE STATE ENTERPRISES AND LITERATURE REVIEW 14

2.1 Environment of Chinese State Enterprises 14

2.1.1 Ineffective Monitoring 14

2.1.2 Waning Effect of Non-Material Incentives 21

2.1.3 Weak Relationship Between Work Effort and Worker's Payment 29

2.1.4 Other Related Features of the Chinese State Enterprises 39

2.1.5 Concluding Remarks 42

2.2 Literature Review of the Agency Theory 43

CHAPTER 3. THEORETICAL MODELS 69

3.1 The Fixed Wage System 71

3.2 The Profit Sharing System 73

3.2.1 Optimal Effort and Optimal Ratio of Share 74

3.2.2 Potential Gains of the Principal and the Agents 77

3.2.3 Analysis of the Free-Riding Problem ... 79

3.2.4 The Long-Run and the Socially Optimal Efforts 84

3.2.5 Incentive Compatibility Analysis 93

3.2.6 Exogenous s and Potential Inefficiency 99

3.2.7 Trade Off Between Efficiency and Equality 104

3.3 The Disutility of Effort Compensation System 106

3.3.1 Advantages of the DEC System Over the PS System 106

3.3.2 Disadvantages of the DEC System Over the PS System 108

3.4 The Extended Profit Sharing System 112

3.4.1 Optimal Effort, Optimal Ratio of Share, and Potential Gains 113

3.4.2 Risk and Uncertainty 117

3.5 The Lump-Sum Profit Turn Over System 123

3.5.1 Advantages of the LST System 123

3.5.2 Acceptability of the LST System 125

3.6 Concluding Remarks 134

CHAPTER 4. EMPIRICAL EVIDENCE AND POLICY IMPLICATION ... 140

4.1 Work Effort Under the Fixed Wage System 144

4.2 Work Effort Under the Profit Sharing System .. 151

4.3 Effect of Bonus on Work Effort 162

4.4 Test of Workers' Attitude Toward Risk 167

4.5 Concluding Remarks 170

CHAPTER 5. REAL SITUATIONS AND POLICY IMPLICATIONS 171

5.1 Wage Reforms and the Associated Problems 172

5.2 Analysis of the Current Wage System and Policy Implications 179

CHAPTER 6. CONCLUSIONS 195

APPENDICES 219

Appendix 1. Tables and Figures 219

Appendix 2. A More General PS System 244

Appendix 3. Expected Utility of the Agent Under the PS System 247

Appendix 4. A Model With Distinct Agents 251

REFERENCES 254

LIST OF TABLES

Table 1. Average Values of Key Variables in 1980 and 1987 219

Table 2. Effect of Fixed Wage on Value Added per Worker (by three-digit industry) 220

Table 3. Effect of Fixed Wage on Value Added per Worker (by two-digit industry) 221

Table 4. Effect of Fixed Wage on Profits per Worker (by three-digit industry) 222

Table 5. Effect of Fixed Wage on Profits per Worker (by two-digit industry) 223

Table 6. Effect of Wage on Value Added per Worker Under Incentive Payment system (by three-digit industry) 224

Table 7. Effect of Wage on Value Added per Worker Under Incentive Payment system (by two-digit industry) 225

Table 8. Effect of Wage on Profits per Worker Under Incentive Payment system (by three-digit industry) 226

Table 9. Effect of Wage on Profits per Worker Under Incentive Payment system (by two-digit industry) 227

Table 10. Effect of Bonus on Value Added per Worker Under Fixed Wage and Incentive Wage Systems .. 228

LIST OF FIGURES

Figure 1. Optimal Effort of the FW System 229

Figure 2. Optimal Efforts of the PS System 230

Figure 3. Optimal Effort of the DEC System 231

Figure 4. Optimal Effort of the EPS System 232

Figure 5. Optimal Effort of the LST System 233

Figure 6. Optimal Efforts of Various Payment Systems .. 234

Figure 7. The Effect of Risk Aversion on Optimal Effort 235

Figure 8. Utility of the Principal and Optimal s (β=2) 236

Figure 9. Utility of the Principal and Optimal s (β>2) 237

Figure 10. Utility of the Principal and Optimal s (β<2) 238

Figure 11. Optimal Efforts of Distinct Agents 239

Figure 12. Optimal and Socially Optimal Efforts (n>1) .. 240

Figure 13. Optimal Efforts of DEC System (Distinct Agents) 241

Figure 14. Optimal Efforts of the DEC system (Discrete Bonuses with two jumps) 242

Figure 15. Optimal Efforts of the DEC system (Discrete Bonuses with three jumps) 243

PREFACE

The work effort of Chinese state employees is motivated through monitoring, non-material incentives, and material incentives. As a stylized fact, ineffective monitoring in Chinese state enterprises has been recognized by both Chinese policy makers and foreign observers. Due to poorly defined property rights, the power of the "supervisor" to punish shirkers is very limited, as long as a minimum work effort is supplied; and the organization of the Chinese supervision system is not conducive to monitoring. As a result, monitoring of the supervisors can only motivate workers to supply a minimum work effort, x_{i0}, which is significantly below full work effort. Work effort can be increased through improved monitoring, this is not only costly, but also faced with political restraints. In this study, I treat monitoring as an exogenous variable, while acknowledging that improved monitoring leads to increased x_{i0}, the question of how to improve monitoring is left unanswered. Monitoring from coworkers is a special feature of socialist China. Combined with the effect of non-material incentives, it was very effective in motivating work effort during the 1950's and even the early 1960's. During this period, moral encouragement was emphasized to create a new type of selfless and pure communist person. Work effort was high in those days because many people believed that their work effort would be compensated

paradise." The revolution had just won a nationwide victory and people were still ecstatic. Over time, due to increased official corruption and perquisites and the stagnation of living standard growth, the effect of moral encouragement decreased as well as the effect of monitoring from coworkers. Some Chinese policy makers still hope to regain the effectiveness of moral encouragement, while others turned to material incentives to motivate work effort.

Even though the socialist distribution principle is "from each according to his ability and to each according to his work," this principle was not really implemented. In the Anti-Rightist-Campaign of 1957 and the Cultural Revolution, material incentives were regarded as revisionist, and payment was independent of work effort. Even under the current payment system, egalitarianism is still a very common feature which is officially derided as "eating from the same big pot."

To reduce egalitarianism and motivate work effort through material incentives, in 1977-78, the Chinese government approved that 40% of the state employees who were chosen by their work groups and then approved by the enterprise authorities on the basis of their contribution, technical level, labor attitudes, and political behavior, could receive a wage increase (Shirk, 1981). This wage adjustment, designed to increase work effort by linking payment to performance, led to serious conflict among workers. Workers' performances were very difficult to evaluate not only because there often were

no clear-cut criteria for its measurement, but also because the evaluated period was too long. The attempt to link payment with work effort through partial wage rise was unsuccessful.

Another attempt to link payment with work effort was reviving bonuses. In 1978, the year in which bonuses were revived, incentive pay (bonuses and piece rates) was only 3.1% of the total wage. By 1984, incentive pay increased to 24%. Link payment with work effort through bonuses has several advantages over partial wage rises. Evaluating work effort on a monthly basis is much easier than evaluating it on a basis of several years; and work effort has to be continuously higher, to receive a higher bonus continuously. Since bonus payment is based on work effort rather than the result of work effort, either work effort has to be easy to observe or good proxies of work effort has to exist. In the early stage, bonuses are often determined by the face-to-face small group discussions of co-workers, because co-workers have more information of work effort. By fixing the total bonus of the group, policy makers hoped that co-workers would compete for higher bonuses by supplying higher work effort. Due to the potential gains of coalitions, bonuses were distributed relatively equally, and the effect of bonus on work effort reduced over time.

By 1981, many enterprises, responding to government pressure, reportedly had abandoned the traditional small group

oral assessment and were trying to use more precise quantitative formulas to evaluate workers' performance. In practice, bonuses were tied to a given percentages of the total wage. If the firm has "black and white" evidence that a worker's performance was deteriorated, the worker's bonus would be canceled. So bonuses provided workers with an incentive to work carefully, work on time, and fulfill the production quota, this generally leads to higher productivity. The effect of a bonus on productivity was limited, because some important factors of work effort, such as labor intensity, and investment in human capital, were unaffected by bonuses. In addition, since the total bonus fund is fixed, workers had no incentive to improve firm performance.

To give workers an incentive to improve firm performance, in addition to the fixed bonus fund, a profit sharing (PS) system was introduced in which compensation depends on the collective performance of the firm. The profit retention system was introduced to achieve this target. Under this system, the enterprise retains a certain percentage of its after-tax profits for investment in plant capacity, housing, renovations of technology, and additional bonuses. By 1984, the vast majority of factories had adopted this system (Walder, 1987). Under this system, enterprise bonus funds grew much faster than productivity mainly because of the "soft budget constraint." Enterprises began using a number of legal and illegal means to inflate the wages of workers, including

evading accounting regulations (Chen, 1987). To stop the bonus inflation, a harsh bonus tax was levied on all bonus funds exceeding one-third of the annual wage bill. The tax rate could be as high as 300% (1984 Decision Document: 23). As a result, most enterprises were usually able to pay right up to the four-months limit and virtually nothing above it, regardless of enterprise performance (China Daily, 16 May 1986). The so-called profit sharing system was reduced to a fixed bonus system. To solve this problem, a new profit sharing system which links total wage funds (which are not covered by the bonus tax) with enterprise performance was introduced in 1985. By the end of 1987, 60% of the large and medium-sized state enterprises were under this new PS system (Xin, 1989). Under this system, in principle, only increases in profitability (or productivity) lead to higher labor income, high profitability level does not lead to higher incomes. Enterprises cannot inflate wages as in the profit retention system. Since individual compensation is based on collective performance, there is a free-riding problem which undermines the potential effectiveness of this system.

The agency theory is one part of a broad research program on problems involving asymmetric information. It focuses on utilization of compensation rules with which one player, the principal, seeks to motivate another (or others), the agent, to choose his activities in a way which is advantageous to the principal. In this study, I attempt to use agency theory to

analyze the mechanisms of the Chinese payment systems. In chapter 3, I construct mathematical models of five alternative payment systems: the fixed wage system, the profit sharing system, the disutility of work effort compensation system, the extended profit sharing system, and the lump-sum profit turn over system. A series of propositions are developed using these models. The main conclusions include:

1) Work effort is suboptimal under the FW system in Chinese state enterprises. The FW system is defined as follows: an explicit or implicit contract (agreement) between the employer (the principal) and the employee (the agent) is reached; for a wage payment the employee promises to offer a certain level of work effort. The wage payment is subject to changes when new contracts are negotiated, but is fixed during the current contract regardless of the real work effort level. With perfect information about job performance and sufficient power of the principal to fire or punish shirkers, the FW system is sufficient to generate optimal work effort through effective monitoring. Unfortunately, in Chinese state enterprises, not only is information often imperfect, but also the power of the "supervisor" to punish shirkers is very limited (Lee, 1990) as long as a minimum work effort, x_{i0}, is supplied.

2) If good proxies of work effort exist, compensating work effort through bonuses may improve work effort. Collusion between the managers and the workers may undermine the

effectiveness of this system. The optimal work effort in this system is not socially optimal. Workers have no incentive to investing in human capital and adopting advanced technology. Because work effort is not directly observable, the result of work effort (value added) is very often used as a proxy for work effort. The result of work effort, Px (assume it is linear with respect to x for simplicity), is composed of two factors: x, the work effort level, and P, the return to work effort. To determine actual work effort, the value of P has to be estimated. With asymmetric information on P, workers have strong incentives to misrepresent P. On the other hand, since the promotion of Chinese managers is based on both the approval of their supervisory state organs and the election of the Workers Congress (Hong and Lansbury, 1987), managers must consider both the interests of the state and the interests of the workers. As a result, P is often underestimated. In this system, risks caused by uncertainty, such as power breaks and material shortages are borne by the firm. When negotiating the p values, the effects of uncertainty are often exaggerated. Workers can reduce their work effort and still receive the bonus payment. When good proxies of work effort exist, there is no need to estimate the P value, collusion between managers and workers is unlikely to develop. Because only work effort is compensated, workers have no incentive to increase P by adopting advanced technology and investing in human capital. If work effort is observable, and workers are

identical in terms of P and the disutility of work effort, the manager can select the socially optimal work effort level as a reference point; when workers supply this work effort, a large enough bonus is paid such that the utility form money income is greater than the disutility from work effort. Since workers are not identical and the manager can select only one reference point, in general the optimal work effort in this system is not socially optimal.

3) If the result of individual work effort is observable, introducing the PS system improves work effort. If only the result of collective work effort of a production team is observable, free-riding behavior may undermine the effectiveness of the PS system. In addition, under the PS system the optimal work effort is not socially optimal; and there is a trade off between efficiency and equity. In the PS system, the profit (or value added) level in the FW system is taken as a reference point, extra profit is shared by the principal and the agents. Since compensation is based on Px instead of x, workers has no incentive to misrepresent the P value. The possibility of collusion between managers and workers reduced. If compensation is based on the result of collective work effort, when a worker increases his work effort, the increased profit, after deducting the share of the state, is shared by all coworkers in the firm, while the increased disutility of work effort is borne by himself. The optimal work efforts are determined when marginal benefit of

work effort equals marginal cost of work effort. If there are n workers in the team, the optimal work effort under free-riding consideration is only 1/n of the Pareto-optimal work effort (the optimal work effort without free-riding problem). The free-riding problem considers noncooperative workers involved in a one-period prisoner's dilemma game. If we extend the one-period model to a multi-period model, repeating games can yield cooperation according to the folk theorem (David, 1985). In a repeated game, worker i finds that if he reduces his work effort, his co-workers reduce their effort too. When the optimal work effort under free-riding consideration is below the minimum work effort, x_{i0}, all workers are only willing to supply x_{i0}. There are two possible equilibria, either all workers supply x_{i0}, or all workers supply the Pareto-optimal work effort. Obviously, workers are better off if they all choose the cooperative strategy. There is a potential gain between the cooperative and the noncooperative equilibria. If there is no transaction cost, or the potential gain is greater than the transaction cost, worker will reach agreements that no one should behave like a free-rider. There are three ways to solve the free-riding problem; one is breaking the result of collective work effort into the result of individual work effort through strengthened internal accounting; the second is reducing the transaction cost by reducing the team size and/or allowing production teams to be organized voluntarily, and the third is increasing

the potential gain by increasing the ratio of incentive payment to fixed wage. In the PS system, even though the result of individual work effort is observable, because the marginal return of individual work effort is shared by the state and the worker, while the marginal cost of work effort is borne by the worker, the optimal work effort is below the socially optimal work effort. In this system, there is an optimal share factor, s*, under which the firm's profit is maximized. If the principal set the share factor s equal to s*, the agent with higher P value is better off. The P value is determined by state investment and relative prices; it's unfair to reward a worker because he is using an advanced machine or he is producing a product favored by the price distortions. So there is a trade off between efficiency and equality.

4) If the result of individual work effort is observable, the optimal work effort in the LST system is automatically the socially optimal effort; and the horizontal equality problem in the PS system can be solved in the LST system. In the LST system, the agent pays a lump-sum, LS, and claims the residual revenue. Since the agent claims all the marginal return of work effort, the optimal work effort in this system is the socially optimal effort even when agents are not identical. And the horizontal equality problem in the PS system can be solved in the LST system by adjusting the LS values. The key issue in this system is how to determine the value of LS.

There are two critical values of LS, LS_{min} and LS_{max}, if LS is set between them and agents are risk neutral or uncertainty is low, both the principal and the agent would prefer the LST system to the FW system. There are two other critical values of LS, LS_l and LS_h, between them both parties would prefer the LST system to the PS system. The principal may, however, not select the LST system, because LS_l is not known automatically and LS_l is increasing over time as productivity increases. The principal may set LS below LS_l, in this case, he could have been better off if he had selected the PS system, even though work effort is still socially optimal. Allow LS to be adjusted period by period may solve this problem. If LS is fixed over time, the principal would have no incentive to improve management, this causes social losses. Allow LS to be adjusted over time can solve this problem too. In the LST system, risks are borne by the agents. If agents are risk averse and uncertainty is high, they demand a large risk premium to offset the risk. Setting LS below LS_h no longer guarantees that the agents prefer the LST system to the PS system. If the risk premium is very high, the agents may not willing to accept the LST system even when LS is below LS_{min} (in this case the principal prefers the FW system to the LST system). Uncertainty of workers' income can be reduced by building up an enterprise wage reserve fund which serves the function of a buffer. Extra income caused by uncertain factors can be saved for the "rainy days." This may reduce

the uncertainty level and the risk premium demanded by the agents.

In Chinese state enterprises, payment systems include multi-level structures. In the top level, the state government is the principal, the managers of the enterprises are the agents. In the second level, the manager is the principal, the shop directors or the workers (if the enterprise is very small) are the agents. The assumption that the principal cannot force the agents to supply work effort above x_{i0} may not hold in the top level. If LS is adjusted over time, and after several periods the LS value is set above LS_{max}, the agent would prefer the FW system and reduce work effort to x_{i0}. If the agent is the manager of an enterprise, he may find that it is difficult to reduce effort to x_{i0} because of the pressure from the state government. This phenomenon is commonly called "whipping the fast ox". So unless the LS value is fixed over time, the manager is very often reluctant to reveal the true information about the potential productivity of the enterprise by increase work effort to the optimal level.

Because the Pareto-optimal payment system depends on actual situations, empirical tests of the theoretical propositions are very useful. Unfortunately, I cannot test all theoretical propositions due to data unavailability. In chapter 4, only the effects of fixed wages, profit sharing, and bonuses on work effort are tested using enterprise level

data. Since work effort is unobservable, per worker value added and per worker profit are used as proxies for work effort.

Under the FW system, if workers are identical, we expect that value added or profit is independent of wages. In reality, workers are not identical. Regional differences, industrial wage differences, and differences in average tenure all cause differences in average wages. So value added or profit might be positively related to average wages. The regressions using 1980 data demonstrated, however, that value added and profit are independent of wages under the FW system.

The effect of the profit sharing system is tested using 1987 data. These regressions demonstrated that value added and profit are positively related to profit sharing. The effect of bonuses is tested using both 1980 and 1987 data. The regressions show that the effect of bonuses is decreasing over time, and the effect of profit sharing is smaller than the effect of bonuses.

CHAPTER 1

INTRODUCTION

1.1. Research Questions

The low labor productivity in Chinese state enterprises has long been recognized by economists all over the world. At the state level, the strongly centralized labor system, along with the mistaken belief that the real industrial wage and employment level could be fixed by political will, caused more disguised unemployment than in a typical less developed countries and considerable labor misallocation (Fei and Reynolds, 1988). The lack of factor markets in general and the inefficiency of the planned economy also contributed to low labor productivity. This study is, however, focused on the individual level. Through the analysis of work effort, we can determine the problems of the current labor system in Chinese state enterprises and make policy recommendations. The low level of work effort in the Chinese state sector has been one of the most serious problems in the Chinese economy (Wu 1988, Walder 1987, Chow 1985). To identify the reasons for suboptimal work effort and discover policies to increase it, I examine individual decisions to supply work effort.

1.1.1. Why People Supply Work Effort

Roughly, people supply work effort either voluntarily or in response to material incentives and/or outside pressures. If the work per se generates positive utility to the worker, he may supply work effort voluntarily. Unfortunately, most people do not have this taste. Since people's tastes over different jobs are different, if an individual can select his job in the market instead of being assigned to a job (as in the case of Chinese state enterprises), other things being equal, workers are more likely to supply work effort voluntarily. A workers' interest in a certain job depends on the way he is treated in the job. If a worker is treated as a machine, he is less likely to develop positive utility from the work per se. If, on the other hand, he is treated as a human being, other things being equal, he is more likely to supply voluntary work effort. Japanese firms have made great efforts to increase voluntary work effort. Large firms in Japan appear to offer lifetime employment to their regular employees; many comparative sociological studies have documented how well Japanese workers are treated by management (Weitzman, 1984). Another source of voluntary effort is political conviction or religious belief. Some politically or religiously motivated people are even willing to sacrifice their life for their beliefs. For these kind of people, it is not strange to see voluntary work effort if they believe that workers should work as hard as possible regardless of the

payment. The overthrew of the Chinese national government on the mainland was heavily dependent on such voluntary effort. If people are truly convinced that they should work as hard as possible, then monitoring and even money compensation are not necessary to guarantee high work effort. To understand the environment of the Chinese labor system, I analyze how people's political beliefs changed and how this affected their work effort (chapter 2).

The second reason why workers supply work effort is pressure from monitoring. If property rights are well defined and work effort is easy to observe, monitoring often leads to an optimal effort level. The Great Wall was constructed under such monitoring. The property rights were well defined --- the capital, the raw material, and even the body of the worker were owned by the emperor. The work effort was easily observed by the supervisors with their whips. Shirkers would be whipped or even killed. The total disutility of work effort might be greater than the total utility from the payment; the workers were, however, not free to leave the job. As a result, work effort was not a problem in this system. Optimal amount was obtained given constraints on mobility and freedom to monitor. In a typical classical Western firm, the property rights are also well defined --- the employer owns the firm or represents the owner. A contract between the employer and the employee is reached before work is actually started. In exchange for the payment, the employee promised

to supply a certain effort level; if he reduces his work effort below the agreed level and this is observed by the employer, he will be punished or fired. Since the employer owns the firm, he can fire a worker if he believes that the worker is shirking. If work effort is easy to observe, then reliance on monitoring leads to efficient work effort. Depending on monitoring to control work effort is inefficient when the property rights are not well defined or work effort is very costly to observe. In Chinese state enterprises, all workers are "owners" of the enterprises in principle. A owner might be criticized for shirking, but he cannot be fired from his ownership. In practice, Chinese workers are not real owners (neither they behave like real owners). Due to this vaguely defined property right, punishment on shirking behavior is possible, however, any punishment must based on "black and white" evidence and the punishment is trivial. The enterprise cannot punish workers as long as they supply a minimum work effort. Workers are very unlikely to be fired because of shirking.

Work effort is also affected by peer pressure from coworkers, since the relationship with coworkers and the behavior of coworkers affect a worker's utility. Coworkers are willing to put peer pressure on shirkers for two reasons: i) If some workers believe that shirking is an unacceptable behavior, they put peer pressure on shirkers. This pressure is positively related to the effect of nonmaterial incentives.

As fewer workers supply voluntary work effort, fewer workers place peer pressure on shirkers for political reasons. ii) If workers' payment depends on collective work effort, one worker's effort affects all coworkers' incomes, and then coworkers have incentive to put pressure on shirkers.

Of course work effort is related to incentive payments. If payment increases as work effort increases, the utility generated from the extra payment will offset the disutility of extra work effort.

1.1.2. Alternative Ways to Improve Work Effort

Based on the previous discussion, the state can improve work effort through four directions. First, the state can increase voluntary work effort as discussed before. Improving work effort through this direction is, however, beyond our concern. In this study I assume that work effort is only offered in response to material incentives and/or outside pressures. Second, work effort can be increased by improved monitoring. This is relatively easy if the property rights are well defined and inefficient monitoring is caused by asymmetric information. Monitoring can be improved through direct acquisition of information by the managers. Since monitoring is costly, maximizing profit requires choosing an optimal monitoring level. Optimal work effort depends on monitoring costs; this is an economic problem rather than a political problem. When the property rights are vaguely

defined, things become more complicated, and the problem turns into a political one. The reasons for ineffective monitoring in Chinese state enterprises are discussed carefully in chapter 2. The question of how to improve monitoring is, however, left unanswered. For our purposes, suffice it to say that monitoring in Chinese state enterprises does not produce the optimal work effort.

Increasing work effort through peer pressure has the advantage of generating additional information about work effort. By working together, coworkers can often observe each others' work effort without any extra cost. The problem in this case is how to motivate workers to put pressure on their shirking coworkers. This issue is discussed carefully in the discussion of the free-riding problem.

The focus of this study is how firms can increase work effort by changing incentive systems. Improving work effort through incentive payments has been the concern of agency theory. When perfect observation of actions and outcomes is impossible, complete contracting between a principal and agents is infeasible. With asymmetric information about the agent's effort level, three things can be done to improve work effort: i) direct acquisition of information by the uninformed party-- monitoring arrangements; ii) rearranging the pattern of allocation so that informational asymmetries have a smaller impact---the optimal assignment problems; and iii) design of

compensation rules when individual actions are not observed and basic incentives are in conflict (MacDonald, 1984).

In the Chinese situation, incentive payment systems are called for not only because of asymmetric information but also because managers have very limited authority to punish shirkers. If ineffective monitoring is caused by limited management authority, direct acquisition of information is not useful. And I am not dealing with optimal assignment problems in this study. The basic research question is how to improve work effort through optimal compensation rules.

One research area which deals with this issue is the design of optimal decision-making system or the architecture of an economic system (Sah and Stiglitz, 1985; Sah, 1991). To determine the payment system, should a polyarchical or a hierarchical or a mixture of the two structures in decision-making be adopted? If we adopt a mixture of the two structures, what are the optimal weights associated with each structure? Along with the economic reforms, Chinese decision-making system shifted from a basically hierarchical structure to a mixture of the two with the weights fluctuated back and forth. The debate over centralization or decentralization has been very intensive, and sensitive to politics. In this study, I sidestep this challenge by exogenize the decision-making system; the analyses are mainly based on the current decision-making system.

Even though the Chinese economy is shifting toward a decentralized economy, Chinese state sector is still a hierarchical system which includes the state government (the state government is also not an unity, the interests of various departments may be in conflict), enterprise manager, workshop directors, production team leaders, and workers. In this multi-level structure, each level is dependent on other levels. A multi-level model would improve our understanding of the Chinese system, but in this study, I have considered only a one level model.

1.2. Methodology and Approach

The basic research question is conducted using agency theory. Agency theory is one part of a broad research program on problems involving asymmetric information. It focuses on utilization of compensation rules with which one player, the principal, seeks to motivate another, the agent, to choose his activities to maximize the principal's utility. A detailed literature review of agency theory is conducted in chapter 2. Briefly, agency theory includes three parts:

1) Choosing optimal information systems which are used to measure agents' performance. For example, under certain conditions should the principal try to measure work effort directly or just measure the realized output?

2) Choosing the incentive payment system which generates Pareto-optimal work effort. For example, under given

conditions should the principal adopt a profit sharing system or an ordinary bonus system?

3) Choosing the value of optimal control variables in each incentive payment system. For example, under the profit sharing system, how is the optimal share ratio determined?

The standard practice in agency theory is to assume a determined utility function of the agent, which is a function of work effort (or the agent's actions in general) and compensation rules, and a utility function of the principal, which is also a function of the agent's work effort and payment schemes. For each payment scheme, to determine the conditional optimal work effort, the agent maximizes his expected utility in terms of work effort, conditional on the given payment scheme. To determine the optimal payment scheme, the principal maximizes his expected utility with respect to the payment scheme, subject to the agent supplying the conditional optimal work effort.

One difference between my theory and other agency theories is that standard agency theory treats monitoring as an endogenous variable. In this study monitoring is treated as an exogenous variable. I simply assume that monitoring is ineffective and exogenous. The so-called optimal work efforts in this study are in fact second-best solutions.

Most agency theories assume very general forms of utility functions. They correctly identify the major issues, but have very little positive content. If concrete policy implications

are to be derived, more specific forms of utility functions should be assumed. Using the concepts of agency theory, modelling alternative payment systems of Chinese state enterprises, analyzing these specific models, and generating concrete policy implications are the main contributions of this study. I start from a very simple model with one principal and n agents. The team size n is small enough so that each agent expects that all other agents in the same team supply identical work effort. The principal and the agent are both assumed risk neutral, and there is only one period. Information about work effort is asymmetric, as the principal only observes the final output of the agent. Monitoring is exogenous and ineffective and can only guarantee a minimum work effort which is well below the optimal work effort. A constant return to scale production technology is assumed. The principal can not affect the agent's productivity through his effort in management, that is I assume that the production function is separable in labor and management. Agents are assumed to derive utility from money income and disutility from work effort. To preserve the strict concavity of the agents' utility function and to simplify the analyses, I follow Cauley and Sandler (1991) in assuming that disutility is a quadratic function of work effort. Using this simple model, I calculate the conditional optimal work efforts and optimal payment scheme. The utility levels of the principal and the agent are calculated by substituting the optimal work

effort and optimal payment scheme into the utility functions. This allows me to compare utilities under different payment systems. The simple agency model is then extended in the following directions:

1) I assume that the team size is large enough so that at least part of the agents' work effort in the team is a constant when agent i determines his optimal effort level. The free-riding problem is analyzed in this model.

2) I extend the one-period model to a multi-period model. A solution of the free-riding problem through the cooperation of coworkers is discussed. The issue of one-period and multi-period optimal effort levels is also discussed.

3) I relax the assumption that agents are risk neutral. The effect of uncertainty and risk aversion on the agent's and the principal's utility functions and the selection of the optimal incentive payment system are analyzed.

4) I relax the assumption that the agent's productivity is independent of the principal's management effort and examine how this affects the selection of the optimal payment system.

5) I relax the assumption that the disutility of work effort is quadratic. With a more general form of disutility function, I analyze how the optimal effort and optimal payment scheme are affected.

Using these models, subject to boundary conditions and restrictions based on Chinese conditions, I compare the payment systems with respect to the following issues:

1) The individually optimal work effort in each payment system and the socially optimal work effort.

2) The distribution of the extra profit generated from the extra work effort.

3) The unobservability of productivity and incentive compatibility.

4) The exogenous vs endogenous compensation rules and the potential inefficiency associated with the exogenous compensation rule.

5) The trade-off between efficiency and equity.

6) The ability to offset risk in each payment system.

7) The acceptability of the incentive payment systems to the principal and agents.

To test the theoretical propositions empirically and generate policy implications, in this thesis:

1) I use empirical evidence from other papers, for example, the World Bank Questionnaire, to support the propositions developed from the mathematical models.

2) I runs cross-sectional regressions using the enterprise data (collected by the Chinese State Statistical Bureau and submitted to the World Bank) which includes data on 396 state enterprises in 1980 and 1987---one year before reform, another after reform.

1.3. Structure of the Dissertation

In Chapter 2, I first discuss the environment of Chinese state enterprises, and present intuitive explanations for ineffective monitoring and suboptimal work effort in Chinese state sector, then review the literature of agency theory. In chapter 3, I construct the mathematical models of five alternative payment systems (the fixed wage system, the profit sharing system, the disutility of work effort compensation system, the extended profit sharing system, and the lump-sum profit turn over system), analyze these models, and generate a series of propositions. In chapter 4, I conduct the empirical tests of some of the propositions generated in chapter 3. In chapter 5, I introduce some empirical evidence from other related papers, and offer some policy suggestions with respect to the real Chinese situations. In chapter 6, I summarize the main results and policy implications, list the limitations of this study, and discuss the directions of further research.

CHAPTER 2

ENVIRONMENT OF CHINESE STATE ENTERPRISES AND LITERATURE REVIEW

2.1. Environment of Chinese State Enterprise

The environment of Chinese state enterprises has been analyzed by many economists (Chow, 1985; Granick, 1987; Wong, 1989; Lal, 1990; Walder, 1987; Wu, 1988; Chen, 1987; Fei and Reynolds, 1988; Shirk, 1981; Burns, 1987). In this section, I focus on the microeconomic, especially labor-related, characteristics. By analyzing these characteristics, an intuitive understanding of why work effort is suboptimal in Chinese state enterprises is presented.

2.1.1. Ineffective Monitoring

As a stylized fact, ineffective monitoring in Chinese state enterprises has been recognized by both Chinese policy makers and foreign observers (Walder,1989). To understand why monitoring is ineffective in Chinese state enterprises, three aspects of the issue (the power of the supervisor to punish shirkers, the organization of the supervision system, and the incentive of the supervisors to supply monitoring effort) are discussed.

1) Chinese state supervisors have less power to punish shirkers. To see why it is the case, let's take a classical

Western private enterprise as a reference point. In the ideal classical model of the Western firm, property rights are well defined; the employer can fire the employee according to his will. If an employee is found shirking or the employer believes that he is shirking, the employer can easily fire the worker. In the Chinese firm, however, all workers are owners in principle. How could an owner be fired? In practice, Chinese workers are not really treated as the owner of their enterprise, the workers do not believe that they are true owners also. However, because workers are owners in principle, the relationship between the management and workers is different from that in a classical Western firm. For one thing, any punishment must based on "black and white" evidence that the worker is a shirker. The merit pay system which is very common in the West simply cannot be adopted in the Chinese case. Because the manager is not the owner, he cannot make decisions about the punishment according to his will. The situation is similar to the criminal prosecution where both parties have equal right. Decisions must be based on black and white evidence. Black and white evidence is, however, often unavailable or too costly to obtain. So, many shirkers are not punished even though the manager knows that they are shirking. Second, even though black and white evidence is available, excess forgiveness is very common. Workers are very unlikely to be fired because of shirking. A shirker may loss some of his bonus money, is less likely to be

promoted, or is more likely to be assigned to unpleasant jobs, but the punishment is too light from the classical Western firm's point of view. The underling cause is still poorly defined property rights---workers are owners. Not all the real Western firms fit the ideal classical model of Western firm, and not all Chinese state firms belong to the ideal Chinese state firm described here. The real Western firms and Chinese firms are not located in the two extreme poles, rather they are located somewhere between the two poles with the real Western firms closer to the ideal classical Western firm, and the real Chinese state firm closer to the ideal Chinese state firm. As a maintained hypothesis, I assume that the distance between the real Chinese state firms and the real Western firms is still very large.

2) The organization of the Chinese supervision system is not conducive to monitoring. It's common in the Western firm that supervision is conducted by the personnel whose job is solely monitoring workers' work effort. Work effort may be directly observed by these supervisors. This kind of organization cannot be found in the Chinese state sector, because Chinese workers are supposed to be "the masters" of the firm. It's "inappropriate" to have someone watching the masters' work. Supervision is conducted in the Chinese case by cadres who also have some other job to do. Work effort is not directly observed in this system; it's indirectly observed through the outcome of effort. Since most production

processes are subject to some degree of noise, black and white evidence of work effort is hard to obtain. Chinese cadres are divided into two categories, production cadres and political cadres. The production cadres are responsible for production. Since supervising workers' effort is only one of their tasks, the energy devoted to monitoring is very limited. Political cadres are belong to the communist party. Their mission is to follow the instructions of the party. If the party's objective is to increase production, they will keep their eyes on workers' effort. However, they are not supervisors, they can only "persuade" workers to increase effort, and the party's objective is not always to increase production.

3) The incentive of the "supervisors" to supply monitoring effort is weaker in Chinese state firms. Here we assume the "supervisor" is the principal. To analyze the effort of the principal spent on monitoring, I assume the principal maximizes his expected utility, $EV_p=Rx_p-x_p^2$, where x_p is the effort of the principal spent on monitoring. Rx_p is the personal return of this effort in terms of utility, the second term on the right hand is the disutility of effort (to preserve the strict concavity of the utility function, following Cauley and Sandler (1991), I assume that the disutility is a quadratic function of monitoring effort). The principal compares the utility generated from personal return of monitoring effort and the disutility caused by offering this effort, if the value of R is large, the principal

supplies more effort on monitoring, since the optimal monitoring effort is R/2 in this case. If we assume the principal is a manager, then the value of R is a positive function of the welfare associated with the position, r_1, the manager selection rules, r_2 (I define that r_2 is high if manager is selected based on the performance of the firm), and the power to punish shirkers, r_3. So we have $R=R(r_1,r_2,r_3)$. If the welfare associated with the position as a manager is high, people have incentive to be a manager and R is large; if the manager is selected based on the performance of the firm which is in turn positively related to workers' effort, r_2 is high, and the manager is more willing to supply monitoring effort; if the manager has the power to punish shirkers, so that for the same monitoring effort work effort is higher, then r_3 is high. Many literatures suggested that r_3 is low in the Chinese state enterprises (Walder, 1989). Now let's take a look of r_1 and r_2.

a) The welfare associated with the position. The answer to this question is ambiguous. For one thing, the wage difference between a manager and a worker is very small in the Chinese state sector. Wage difference is, however, only one benefit from being a manager. Nove (1977) pointed out that "the cadres value the power to control people with its accompanying status, prestige, and perquisites as witnessed by the notorious system of 'nomenclature' in most communist systems." So, the value of r_1 depends on how the manager

values such things as social status and prestige. It also depends on the perquisites associated with the position. In fact, many people in China are eager to be promoted as managers, the value of r_1 is high at least for some of the Chinese managers.

b) The manager selection criterion is more ambiguous. In China, there is almost no market for managers. "By the 1960's, China's cadre system had become a powerful, party-dominated one. A personnel dossier and cadre evaluation system was in place, and that cadre appointments, promotions, transfers and removals were deliberated and conducted in an atmosphere of secrecy" (Burns, 1987). During the economic reform, even though cadres may be elected by workers in some cases, authorities must first submit the names of the nominees to the relevant upper organs for approval. This cadre system endorses personal recommendation as the best method of filling organizational positions. Abuses like factionalism, nepotism, and localism can be easily accommodated in such system. Personal loyalty to the upper level leaders is most important. Better enterprise performance is also appreciated by the upper level leaders if personal loyalty is also offered. The relationship between personal loyalty and better enterprise performance is very subtle. On the one hand, the top leaders need to build up a good public image that under their leadership the economy is performing well. It's also a useful "card" in the political power struggle. On the other hand,

there is no objective criterion about enterprise performance. An enterprise which could have increased profit by 50%, but profit only increased by 20% because less monitoring effort was supplied may be treated as a good one; while an enterprise which could have increased profit under normal effort by 5%, but in fact profit increased by 10% because of over supplied effort may be treated as a worse one. The commonly accepted rule is that as long as the enterprise is increasing its profit (output value), however small the increase might be, the manager will not in trouble. The manager will supply a minimum effort level, which guarantees that the output value is not declining. After this effort level is supplied, following the instruction of top-authority seems to be the most important thing to do. A manager who always does whatever the top leader told him to do, will almost never loss his position. Managers may change positions from enterprise to enterprise, if the enterprise is unsuccessful, but they seldom loss their position as managers. In the case of a political power struggle, personal loyalty is more important than enterprise performance. So the value of r_2 is positive but much smaller than that in the Western firm; the value also depends on the political climate. As a result of smaller r_2 and r_3, the R value is smaller in the Chinese state enterprises, and so is the incentive of managers to supply monitoring effort. Walder (1989) points out that "there are two ways of increasing the enterprise's profits..... The

first is through superior management and efficiency. The second is through effective negotiation with superiors in the planning and financial system of the relevant ministry, province or city." The first way benefits both the state and the workers, and the second way only benefits the workers. If managers represent the interests of the state, they would spent their energy in the first way which includes improved monitoring. A 1985 survey of 900 enterprises found that 80% of employees believe that the manager represents the interests of the factory or its workers; 8% believed that the manager represents the interests of the state (Yang, 1987). This result suggests that managers rather spent their energy to negotiate with government agencies for better treatments with regard to taxation, investment, prices, credit and subsidies, instead of to monitor their employees.

2.1.2. Waning Effect of Non-Material Incentives

The Chinese Communist Party has for a long time pursued a policy of using non-material incentives or moral encouragement to generate voluntary work effort or create an environment where co-workers will check each other's work effort. "Moral encouragement has relied on a variety of different campaigns designed to strengthen the political and ideological education of workers so that they recognize their moral obligations to work hard for the benefit of the nation since, under

socialism, they are the masters of the house" (Henley and Nyaw, 1987).

The peer pressure of co-workers often occurs at meetings; a typical meeting (the self-criticism session) was described by Vepa (1979): "There is then a self-criticism session once a week where the erring employee is afforded an opportunity to analyze his own shortcomings and rectify them. ..... If a person does not utilize this opportunity, he is then subjected to criticism from others,". In some periods of the PRC's history, the effect of non-material incentives was significant. At this stage, however, the effect of non-material incentives is very small. Henley and Nyaw (1987) provided an empirical evidence for this hypothesis. When a simple questionnaire about the ranks of importance of twenty motivational factors was taken by workers and cadres in twelve industrial enterprises in 1984, the statistical results suggested that Chinese workers place great emphasis on material incentives such as wage increases and bonuses. These items were ranked first and second in terms of mean scores. Non-material stimuli such as recognition or a "model workers" award have low rankings. Similar result is also given by a survey conducted by Chinese psychologists Xu and Lin (1980). In this section, I will first briefly review the history in terms of the effect of non-material incentives, then analyze the underlying reasons of the waning effect of non-material incentives.

1) History of the effect of non-material incentives. The recent history of China can be divided into the following periods: economic rehabilitation (1949-52), First Five-Year Plan (1953-57), the Great Leap Forward(1958-60), economic recovery (1961-65), the Cultural Revolution (1966-76), and the Post-Mao periods after 1977. A notable feature of these periods is radical shifts between political extremism which emphasize non-material incentives and political moderation which focus more on material incentives.

In the economic rehabilitation period, the payment system was a partial rationing system which was similar to the pre-liberation "supply system." In this system, cadres from the old liberated areas and new workers who joined the labor force after 1949 were paid with goods in kind while other workers were paid a mixture of commodities and wages; non-material incentives were major factors in generating work effort. In the First Five-Year Plan period, the management system was, in large part, copied from the Stalinist Soviet model which included political campaigns and ideological indoctrination to conduct the function of moral encouragement. Industrial morale was reported to be high during these periods, and non-material incentives were effective in increasing work effort. "At that time people didn't care too much about their own gains and losses and much work got done quite easily. The revolution had just won a nation-wide victory and people were still ecstatic." (Yu 1985). In these periods, material

incentives were also increasing. The combination of material and non-material incentives caused a high work effort level, the people's enthusiasm compensated, to a certain degree, the defects of the economic system. Substantial industrial growth was observed during these periods, as annual industrial growth rate was 35.6% and 9.3% in these two periods respectively (Henley and Nyaw, 1987).

During the Great Leap Forward, moral encouragement was over emphasized. Mao believed that by creating a new type of selfless and pure communist person, China could rapidly achieve communism. An intensive nationwide campaign was launched. The coverage of bonus payments was reduced, and managers and technical staff were excluded from receiving bonuses (Li, 1985). The Soviet type work quota system and piece-rate system were abolished in 1958 (Encyclopedia, vol.1:553); they were replaced by a semi-wage, semi-rationing system in a number of enterprises (Feng and Zhao, 1984). The principle that "to each according to his work" was officially criticized. While the effect of moral encouragement was reduced, plus the fact that material incentives were eliminated, work effort reduced, and the annual industrial growth rate was -18.2%. After the collapse of the Great Leap Forward, Mao sidestepped into the "second line" in 1961. Liu Shaoqi become the top leader for reconstructing the economy. The bonus system was restored, and political campaigns were sharply reduced. By 1965, a comprehensive bonus system was

available for most workers (Xiang 1982). Annual industrial growth rate was estimated to be 15.1%.

China reverted to ideological extremism in 1966 when Mao came back from the "second line" and launched the Cultural revolution. The principle that "to each according to his work" was once again criticized for containing capitalist elements. Material incentives were eliminated, and egalitarianism became the guiding principle of distribution. There was little difference in income, regardless of work effort. The initiative of workers was seriously dampened (Xiang, 1982). This observation reflects the general consensus among Chinese assessments of wage and incentive policies during the Cultural Revolution (Henley and Nyaw, 1987). From 1966 to 1968, the annual industrial growth rate was -11%; from 1969 to 1976, the growth rate was 7.2%. The relatively lower growth rate reflected lower work effort and the effect that human capital is also being treated as capitalist element during the Culture Revolution. Professional engineers, scientists and managers were purged as capitalist elements. With less human capital, the same work effort led to lower labor productivity. Work effort per se was reduced because of the waning effect of non-material incentives and the elimination of the material incentives. The effect of non-material incentives was, however, not totally destroyed, as significant work effort was still being

supplied because of the intensive political education and the peer pressure.

In the Post-Mao periods, the socialist principle that "to each according to his work" was reinstated and material incentive were focused. Since 1978, there have been no political campaigns at enterprise level. The moral encouragement takes the form of emulations such as emulations for "advanced workers," "model workers," "labor hero," and "outstanding workers." Most of the moral incentives are also accompanied by monetary awards or benefits in kind. The effect of moral incentives are reduced significantly. Even the current Chinese leaders have to accept the fact that the effect of moral encouragement is very limited. One of the most popular political slogan in China is **Xiang qian kan** (look forward into the future), where the word **qian** has two meanings in Chinese, one meaning is future, another is money. Many people like this slogan, only the meaning of **qian** is money. Recent Chinese history shows a clear trend that the effect of non-material incentives is declining over time.

2) Underlying reasons for the waning effect of non-material incentives. There are three main reasons for the trend demonstrated above: stagnation in living standard growth, increased official corruption and perquisites, and increased political democracy.

Following the redistribution effect of the Land Reform in late 1940's, the living standards of many families increased

sharply. Many people expected that the "communist paradise" would be realized very soon as the Communist party declared. This expectation was supported by the success in the civil war. People trusted the new honest government which had severe discipline to punish official corruptions and there were almost no perquisites. During the civil war Chinese Communist Party (CCP) leaders earned no wages, they ate almost the same food as common soldiers, which was necessary to survive the civil war. The good relationship between the Party and the mass was critical to obtain people's support which was very important to win the civil war. Work effort was high in those days because people believed their effort would be compensated with future benefits---the realization of the communist paradise. This was a "check" written by the CCP which had a good credit history.

With the consolidation of the "dictatorship of the proletariat," the relationship with the common people was not so important any more. Party leaders who have been "fighting for the power" now want to "enjoy from the power," official corruptions and perquisites were increasing. The reputation of the government deteriorated. Some suspicions against the "communist dream" developed because of the stagnation of living standard growth. People's willingness to supply voluntary work effort declined. Under the dictatorship of the proletariat and the political campaigns, however, peer pressure was still very effective, as long as some workers

still believe in the communist dream, and put peer pressure on their co-workers, the other workers "dare not" significantly reduce their work effort. Shirkers might be blamed for trying to destroy revolution by sabotaging production, which was a serious political crime at that time. Work effort was supplied under political pressure. The degree of political democracy was very low at that time.

After the Cultural Revolution, many officials who were purged during the Cultural Revolution regained their power. The party believed that they should be compensated for their suffering and a lot of these officials felt the same way. Punishment of official corruptions was reduced and official corruption increased. The public image of the government was seriously damaged. Income growth was still stagnant, and the "open door" policy opened the view of the Chinese people toward the market economy. Chinese people were shocked when American tourists rushed into China with their abundant "greenbacks." How come the people in the "evil" capitalist country enjoyed such a high standard of living while the people in the "socialist paradise" were still fighting for their subsistence? The communist dream was smashed eventually. The "check" written by the CCP may never be honored, and the CCP no longer has a good credit history. Peer pressure was no long effective, because very few, if any, people still believe in the communist dream. People's struggle for political democracy made it impossible to

increase work effort by political pressure. Political campaigns have ruined its reputation, people were tired of politics. The only "check" they are willing to accept is the money check. Currently, the effect of non-material incentives on work effort has been reduced to a very low level for the above reasons. Even though the Chinese government still hopes to regain the effectiveness of moral encouragement, there is no evidence that the effect could be increased at least in the near future.

2.1.3. Weak Relationship Between Work Effort and Worker's Payment

Even though the socialist distribution principle is "from each according to his ability and to each according to his work", the principle is not really followed. In the Anti-Rightist-Campaign in 1957, and The Cultural Revolution, material incentives were regarded as revisionist (Wu, 1988), and payment has no relationship with work effort at all. Even under the current payment system, with its focus on the material incentives, egalitarianism is still a very common feature of the payment system. The Chinese policy makers have realized this fact and officially derided as "eating from the same big pot". In China, work related income including basic wages, subsidized commodities, bonuses, and welfare benefits such as privileged access to enterprise owned housing. The basic wage, even though very complicated, can be roughly

broken down into three sub-systems: (i) the system of wage brackets for workers; (ii) the post or job-type wage system; and (iii) the system of wage brackets for cadres. Under system (i), workers are paid according to an eight-grade wage system with new unskilled labor in grade 1. Promotion to a higher grade is based on work effort and performance in principle; in fact, promotion is based on seniority if a minimum level of work effort is supplied. The reason is that under the ineffective monitoring system, black and white evidence of work effort level is very often unavailable, so promotion is not related to work effort and is also not related to enterprise performance. China had in the past adopted a low-wage policy. Under this policy, China had been able to increase capital accumulation and investment in industrial development, the cost was paid by the labor force. As a result, workers' wage level was very low before the wage reforms. Between 1953 and 1978, the annual growth rate of real wages was only 0.3%. All workers "deserve" a wage upgrade. Seniority seems to be the easiest criterion for wage upgrading.

Chinese state enterprises are, in fact, multiple-function communities which providing housing, medical care, welfare benefits, and even employment of worker's offspring (Fei,1988). All these work-related welfare expenditures are basically independent of work effort, and the welfare

expenditure in real value is about the same as total wage expenditure.

Bonuses were the main form of incentive payment. In the past, state enterprises had to restrict annual bonus to a fixed ratio of annual basic wages. The bonus funds were independent of the performance of the enterprise. The main feature of this system is that bonus depends on effort level, not on the results of the effort supplied. If effort is clearly measurable, this system is basically feasible in principle. In most cases, however, the principal cannot directly measure the agent's effort level. An unambiguous and specific criterion for determining bonuses was hard to develop in most cases and workers became contentious. "..... [T]he common method used to assess workers,..... are face-to-face group discussions in which vague criteria are applied." (Walder, 1987). Because bonuses are decided by a "voting system", workers tend to vote his best friends, very often bonuses have not encouraged work effort, it encouraged effort toward common relationships. This "voting system" may end up with egalitarianism, because, first of all, the ratio of the bonus to fixed payment is too small, the marginal benefit of extra bonus is lower than the marginal cost which results from:

a) The time spent on the face-to-face meetings to decide the bonus distribution. When everyone is hoping to get a bigger share for himself or herself and there is no unambiguous

criterion, a consistent agreement is very time consuming. b) The tension among workers will be developed in this system. Given the expectation of life time employment in the same enterprise for most Chinese state workers, workers care about their relationships very much. The marginal cost to ruin their relationships may be too high. In some cases, bonuses depended upon the shop director's discretion; the director may use his or her power to reward friendship and other non-work related matters (Walder, 1986).

There is a critical difference between the Chinese state firm and the Western firm, that is the power of workers to against an "unfair" group leader and/or shop director. Under the Western merit pay system, a worker facing an "unfair" supervisor can report to the manager, but very often the manager would rather believe the supervisor instead of the worker. If the worker tries to disrupt production to get "revenge", he may be fired. If he quit, the cost of a "quit" to the firm is very limited. In the Chinese case, however, angry workers can disrupt the production process by shirking, striking or even engaging in riots. If workers believe that the "merit" pay is unfair, or even though that the merit pay is in fact fair but there is no unambiguous and specific criterion to prove that it's fair, they will take actions to disrupt production. The potential cost of such actions to the worker is at most the loss of all his bonus, which is low for an "angry" worker. The potential cost for the principal is,

however, very often too high. Among other things, the principal has created an enemy for himself. In order to avoid these potential disruptions, bonuses are distributed equally. So in the traditional bonus system, the relationship between payment and work effort was still very weak.

The overwhelmingly egalitarian nature of the wage system has seriously reduced the initiative and motivation of good workers. This problem has been realized by Chinese government leaders, as they have intuitively realized that workers' payment should be related to work effort. In order to link worker's payment to work effort, three things have been done: i) increase the proportion of incentive wage; ii) link the total funds of incentive pay to enterprise performance, and iii) link incentive payment to more precise quantitative formulas geared directly to performance.

From 1977 to 1984, the average industrial wage rose from 632 yuan to 1071 yuan, a 40% increase adjusting for the official inflation rate. "In 1978, the year in which bonuses were revived, time wages accounted for 85% of wages, incentive pay (bonuses and piece rates) were only 3.1% of the wages, and fixed subsidies another 6.5%. By 1984, straight time wages comprises only 58.5% of the wage bill, while incentive pay increased to 24% (fixed subsidies increases to 14.5%). Underlying this shift has been a continuous call to tie incentive pay closely to individual performance" (Walder, 1987).

Prior to 1979, factory bonus funds were set at 5% of the total wage bill. If the factory met its output targets, then the bonus funds were released. By 1984 most enterprises had adopted a new profit retention system, in which the enterprise retains a certain percentage of its after-tax profits for reinvestment in plant capacity, renovations and additional bonuses (Walder, 1987). Since the replacement of "profit delivery" by taxes in October 1984, enterprises which under the "contracted responsibility system" can make the decision about the amount of bonus payment, the state only collects a tax on the above-norm bonus (300% tax rate) from enterprises (1984 Decision Document:23). Bonus limits were lifted so that the state only collects tax for bonuses of over four months' basic wages. In principle, "enterprises which have failed to fulfil their quotas and paid less taxes and earned less profits must reduce or stop bonuses, or even withhold portions of their employee's wages" (Jin, 1984).

"By 1981, many factories, in response to government insistence, reportedly had abandoned the traditional small group oral assessment in favor of more precise quantitative formulas geared directly to output" (Shirk, 1981) in bonus distribution.

In order to link wages and bonuses with workers' performance, some pilot incentive payment systems were introduced:

1) The floating wage system. Under this system part of workers' standard wage and bonus becomes a "floating wage" which is then paid to workers on the basis of their performance. This seems to be a very popular wage system.

2) The floating promotion system. Under this system, enterprises can promote a greater number of the more capable workers financed from part of the wage and bonus fund which is dependent on enterprise performance.

3) The structural wage system. Under this system, wages are broken down into four parts: a base wage, a seniority wage, a position wage, and a flexible wage. The base wage is supposed to ensure that workers' basic needs are met. The seniority wage is based on years of working experience. The post wage is a responsibility payment to cadres, technical staff or highly-skilled workers. The flexible wage is based on a measure of the enterprise's performance and of the contribution of individual workers. It consists of a floating wage, bonus and subsidized benefits (Henley and Nyaw, 1987).

4) The profit sharing system. In some research institutions, a profit sharing system has been accepted since 1987. Under this system, the profit generated by each worker is shared by the institute and the worker according to a predetermined formula.

5) The lump-sum profit turn over system. Under this system, the enterprise signs a contract with the state in terms of the fixed profit turn over each year for several

years, the profit after the lump-sum turn over will be spent the same way as in the profit retention system.

The effect of these measures has been discussed by some economists (Cauley and Sandler, 1991; Wang Dingding, 1990; Xiao, 1990; Lee, 1990; Chen, 1987; Walder, 1987). All of them agree that these reform measures failed to achieve their potential. The reasons which undermined the potential effect of these reform measures can be summarized as follows:

1) In the absence of a labor market, when wages or bonus increased in one enterprise, workers in every other enterprise demand wage increase too. For example, after the taxi companies in Beijing achieved wage raises, the bus drivers applied to be transferred to taxi companies. When their applications were turned down, they were usually compensated by a wage increase. Since any one enterprise's wage increase is merely a result of the rigid economic mechanism, a call for "upward emulation" naturally is received with sympathy by factory managers, and local governments. According to an empirical study of 429 enterprises in 1984 (Chen, 1987), enterprises were able to raise the percentage of retained profit in gross profit from 19.36% to 21.59% and to increase the bonus disbursement in retained profit from 25.43% to 36.7% through "negotiations with their higher-ups. In the sample enterprises, more than 40 different methods were employed to dodge wage control and secure more bonuses and benefit. A survey in 1985 also found that many firms used all their

retained profit for bonuses and collective welfare (CESRRI, 1985, Report No.5). As a result, many firms increased bonus payment without increase work effort and improve firm performance. The firm which supplied high work effort may end up with lower percentage of retained profit. This phenomenon is called "whipping the fast ox," and it also can be observed within a firm. Because of the uncertainty of labor productivity, if a group of workers increased their output or profit significantly, there are two possible explanations, one is that they increased their work effort significantly, another is that their original quota is too low, so with a very little increase in work effort, output will increase significantly. Even if the output increase is caused by increased work effort, the other groups or firms may not acknowledge this. To maintain horizontal equity, the manager or the government authority is likely to rise the quota, so that the marginal benefit of work effort is reduced. In principle, the bonus payment is not limited from above, in reality, the horizontal difference is quite limited regardless of work effort.

2) Even though bonus distribution rules become more and more dependent on measurable and objective behaviors, precise quantitative formulas which reflect true work effort are not always available. Some proxies of work effort have to be used, for example, a bonus for full attendance. It does prevent workers from absent, but it does not prevent workers

from shirking. Very often, team output cannot be divided into individual output either because of a lack of internal accounting or because of economics of scale. In this case, the free-riding problem may prevent workers from offering optimal work effort. If unambiguous and specific criteria for determining work effort are hard to develop (especially when uncertainty of production is high), distributing bonuses according to work effort is infeasible, since work effort is unobservable. The result to insist the bonus be distributed according to work effort very often lead to egalitarianism in bonus distribution, and the intended incentive pay degenerated into a general wage supplements.

3) Under the profit retention system, the enterprises negotiate a contract with their supervisory authorities regarding retained profits, funding, and input and output quota. After a 15% energy and transportation tax, about half of the net retained profits can be used for welfare and bonus, and the rest of it used in productive investment. Even though the percentage of retained profit in gross profit has been increased, the value of this percentage is still too low, and half of the retained profit has to be used in productive investment which do not serve as incentive payment, and part of the other half has to be used in collective welfare which is also independent of work effort. So the ratio of potential incentive payment to total profit is still very low. For example, in the profit sharing system of the research

institutions, the share which goes to the worker is typically 20%, while the socially optimal share factor is one at the margin.

For all these reasons, the relationship between workers' payment and work effort is still rather weak.

2.1.4. Other Related Features of the Chinese State Enterprises

1) The Chinese state enterprises are under the central planning system. In recent years, some effort has been made to reform the state enterprises so that they will move toward the direction of the enterprises under the market economy, but they are still basically under central planning. Policies are determined by central planners, while ordinary workers have only weak and passive influences.

2) Laborers basically have no mobility in the state sector, and most of the workers expect life time employment in their current jobs. New high school graduates were given compulsory job assignments, and job shifts were seldom permitted. Even though workers may quit their job in principle, chances to find another job are rare.

3) Uncertainty of productivity is high. Many factors may contribute to the exogenous risk facing the Chinese state enterprises, especially in the wake of reforms. Tang (1987) and others have pointed to "systematic" shortcomings in the Chinese planning and supply systems that lead to a variability in the gross profits of enterprises. The inputs with

particular specifications required by the enterprises may not be supplied; insufficient input supply, especially energy, at a given time and place; input may not appear at all owing to altered transportation constraints or failure; input deliveries may be random owning to the presence of multiple leadership and ministerial and regional lines of authority that cut across one another (Tidrick, 1987; Tang, 1987). Price distortion, imperfect market caused by local protection, change of market situations, machine failure, and the possibility to take advantage of differences in demander's ownership in selling products, especially producer goods (chances of bribery) are all factors of risk.

4) Wide variance in capital labor ratio and average as well as marginal productivity of labor. Since China's capital market is not developed yet, capital allocation is not determined by market forces. According to Chen (1987), in the sample of 429 enterprises, the ratio between the highest and lowest per worker value of fixed assets in the sample is roughly 200.

5) The worker's fixed wage is independent of the performance of the enterprise (Alan Gelb, 1990). It's interesting to contrast a Japanese firm and a Chinese state firm. Even though both firms adopt life time employment, the former has fewer shirking problem than the latter. One explanation might be peer pressure from co-workers. First of all, the marginal benefit to prevent co-workers from shirking

may be larger in the Japanese firm, since widespread shirking may cause bankruptcy of the firm which means unemployment for the workers. Since the chance to find a new job is low because other firms may also adopt life time employment policy (Hashimoto, 1990), the cost of firm bankruptcy is very high. While Chinese state firms seldom go bankrupt. As a result, workers in the Chinese state firm have no incentive to care about the performance of the firm under the fixed wage system. Workers are basically indifferent with respect to the quality of their fellow employee's work. Which is quite different from the situation of the Chinese township, village, and private (TVP) enterprises. According to the World Bank Worker Survey Questionnaire, in the TVP enterprises, even though payment is based on the worker's opportunity wage which is not closely related to their effort, workers still care about the quality of their fellow employee's work. Only 14% of the workers in the survey are indifferent to poor performance of their co-workers or consider that it is a matter for managers alone; 32% would like to speak to the shirkers in their firm but feel inhibited; and 54% indicate that they would speak out (Alan Gelb, 1990). The reason is very clear---poor performance of co-workers increases the risk of bankruptcy which would lead to unemployment for TVP workers. For the state workers, either the government will subsidize the enterprise to prevent bankruptcy, or the government is responsible to find the workers alternative jobs in case of

bankruptcy do happen. So state workers bear no risk with respect to poor performance of their enterprises.

2.1.5. Concluding Remarks

Suboptimal work effort is one of the most serious problems in the Chinese state sector. The low work effort is caused by ineffective monitoring, waning effect of non-material incentives, weak relationship between workers' payment and work effort, and the independence of workers' fixed wage with respect to the performance of their enterprises. The Chinese state enterprises also have the following features: policies are determined by central planners, laborers basically have no mobility in selecting their jobs, uncertainty of production is high, and there is a wide variance in their capital labor ratios. Since the ineffective monitoring and waning effect of non-material incentives are more of political problems rather than economic problems. In order to increase work effort, the economic reform has focused its attention on strengthening the relationship between workers' payments and work effort. Some reform measures have been conducted in the 1980's, but these reform measures are still based on intuition and have failed to achieve their potential.

2.2. Literature Review of the Agency Theory

"The agency theory is one part of a broad research program on problems involving asymmetric information. It focuses on utilization of compensation rules with which one player, the principal, seeks to motivate another (or others), the agent, to choose his activities in a way advantageous to the principal." (MacDonald, 1984). By asymmetric information we mean that the agent's activities are known only to himself, and there is no immediate incentive for him to reveal the true information to the principal.

Under ineffective monitoring and waning effect of non-material incentives, compensation rules have to be used to motivate workers. One of the commonly used compensation rules is share profits or share tenancy. The traditional share tenancy system is defined in the early literature (Cheung, 1968) as two or more individual parties combining privately owned resources for the production of certain mutually agreed outputs, where the actual outputs will be shared according to certain mutually accepted percentages as returns to the contracting parties for their productive inputs. Even though share profits may motivate workers, some economists (Schickele, 1941; Heady, 1947; Heady and Kehrberg, 1952; Issawi, 1957; Georgescu-Roegen, 1960; and Sen, 1966) believed that share tenancy results in inefficient allocation of resources. The traditional approach to analyzing share tenancy may be called the "tax-equivalent approach" since

under share tenancy a portion of every output unit produced is taken away as rent which is similar to an ad valorem excise tax of the government. Share tenancy, therefore, is said to result in less intensive and less efficient farming because the tenant's incentive to work or invest in land is reduced. The debate over the efficiency of share tenancy was finished by Johnson (1950) and Cheung (1968), wherein it was shown that in a non-stochastic environment and under free entry, so long as production technologies are independent, i.e., there are no externalities, sharecropping and related compensation rules yield efficient outcomes. Under these assumptions there is no basis upon which to choose between the various compensation rules. The result generated by Cheung is, however, based on the premise of wealth maximization subject to the constraints of private property rights in a free market. It's obvious that these constraints do not hold for the Chinese state enterprises. More important, the result also depends on symmetric information. In Cheung's model the total amount of tenant inputs, including work effort, are contractually stipulated, "which is essential because the tenant would commit less if only the rental percentage were prescribed." (Cheung, 1968). The disutility of work effort was not discussed.

The early papers in the agency theory (Spence and Zeckhauser, 1971; Ross, 1973; and Stiglitz, 1974; 1975) and much of the later work focus on a two-player (principal and

agent) situation. The agent's action (effort level) x, together with a random state of nature Ω are assumed to generate output Y according to $Y=f(x,\Omega)$. Realized output $y=f(x,w)$, where w is actual Ω realized, is to be divided between the principal and agent, with the principal choosing the method of allocation and the agent choosing x. It's assumed that it's too costly for the principal to observe anything beyond realized output y which is freely available to all. The payment from the principal to the agent, denoted F(.), can be a function of output alone: F(y). F(.) cannot depend on x or w. Assume that both the principal and agent possess state independent Von Neumann-Morgenstern utility functions U=u(.) and V=v(.), respectively. u(.) depends only on income, while v(.) depends on income and x ($v_x<0$); both are weakly concave in income.

The focus of the agency problem is on the compensation rule F(y). There are several ways in which F(y) can be determined. A particularly straightforward one is the partial equilibrium approach of selecting the F(y) which maximizes the principal's expected utility, given that the agent must be able to achieve some specified level of expected utility when he chooses x optimally. The F(y) so obtained is thus ex ante Pareto optimal in the class of compensation rules which are functions of y alone. At this stage, essentially nothing is known about the form of the optimal F(y). The agent's preferences over income lotteries may interact with his choice

of effort in an arbitrary and complex fashion. In order to reach some meaningful results, the agent's preferences are restricted so that his attitudes towards income risk do not depend on x. This separability is accomplished by requiring v(.) to be of this form (Harris and Raviv,1979; Grossman and Hart,1983)

$$v(F,x) = -v_0(x) + v_1(x)v_2(F) \qquad (2\text{-}1)$$

where $v_1>0$ and v_0 and v_2 are monotone increasing and weakly concave. Harris and Raviv show that the principal optimally receives a quantity independent of output, and hence that F'=1. However, as MacDonald (1984) pointed out, "the theory of agency, as currently formulated, has little operational content." In order to use the agency theory to explain the Chinese situations, more specific utility functions have to be assumed and some restrictions on the agent's actions have to be specified. This will be done in chapter 3 of this study.

Given the basic agency model, many extensions were forthcoming. In the two-player, one-period set-up, Harris and Raviv (1979) show that if the agent is risk averse, then full observability of the agent's action will lead to Pareto improvement, which suggests potential gains to monitoring. Holmstrom (1979) also points out that only under a special condition monitoring is useless. In general there will be positive demand for monitoring arrangements. This line of argument was examined further in a much more general informational setting by Gjesdal (1982). It turns out that

the special condition is no longer necessary, although it remains sufficient, that monitoring is useless. All these analysis are, however, based on the market economy and well defined property right. Given the situation of Chinese state enterprises, I treat monitoring as an exogenous variable in the compensation rules. I will also point out the situations under which it's necessary to strengthen monitoring.

Another extension is provided by Lal and Staelin (1982). The case in which one agent produces two goods was analyzed. They point out that except under obvious and extremely implausible separabilities, the optimal compensation rule depends non-additively on both outputs.

Multiple-agent and dynamic versions of the agency problem differ fundamentally from the basic model. Three extensions were developed, the single-period-multiple agent models are the best developed of the three types of extended models, they are also most closely related to the single-period-two-player model. The dynamic-single-agent models are also currently the subject of considerable research activity. The dynamic-multiple-agent models are only under development.

In the multiple-agent extensions, the principal faces n agents. This kind of model can be further divided into three categories. One uses the agency framework to provide insight into why entrepreneurial firms exist; the second recognizes the inherent difficulty of examining fully optimal compensation rules and responds by seeking the optimal rule

from a restricted class; and the third approaches the same difficulty by searching for general results through explicitly modeling the problem as a game of incomplete information, perhaps with communication.

The most important papers in the first category are by Alchian and Demsete (1972), who raised some of the basic issues, and Holmstrom (1982) who provided a correct analysis of them. Because of economics of scale, a group can produce more than the sum of what each person could produce separately. In a team, however, these gains may be difficult to achieve when there is asymmetric information because of the free-rider problem. It has been pointed out that if all output is to be distributed among the n agents, none can have the correct marginal incentives. The Nash equilibrium effort level is suboptimal. The principal enters the picture at this point. If the compensation rule is determined in the following way: only if the output associated with Pareto-optimal action levels of all agents is produced the agents will be compensated, otherwise the principal will keep all output.

$$F_i(y)=F_i^* \quad \text{if } y=f(x_1^*,\ldots\ldots x_n^*)$$
$$F_i(y)=0 \quad \text{otherwise} \qquad (2\text{-}2)$$

where x_i^* is the Pareto-optimal action level of agent i, F_i^* is the optimal compensation rule to agent i. Holmstrom's key insight, in contrast to Alchian and Demsetz's emphasis on monitoring, is that the principal's participation offered a

credible threat. This feature gives the entrepreneurial firm a distinct advantage over pure teams and thus contributes to understanding why it exists. There is an efficiency-based rationale for having a principal in the problem from the outset. Such penalty contracts, however, are very often not feasible in the Chinese state sector because of the reasons discussed earlier, such as poorly defined property rights. Even in the Western firms, penalty contracts are also rare. Uncertainty might be one of the reasons.

As noted above, in an entirely deterministic setting the analysis of optimal compensation rules is straightforward in the market economy: forcing rules (penalty contracts) are optimal among other things. When production is subject to stochastic influences, however, the problem of analyzing optimal compensation rules becomes horrific. The "tournaments" literature grew out of responding to the inherent difficulty of choosing optimal compensation rules by restricting the space of feasible rules. The central works are Lazear and Rosen (1981), Nalebuff and Stiglitz (1983), and Green and Stokey (1983). When the values of the random output of agent i, Y_i, are independently distributed, allowing the ith agent's compensation $F_i(.)$ to depend on anything apart from the realized output of agent i, y_i, merely adds additional randomness to agent i's income stream. Under the assumption of a utility function of the form in (2-1), Holmstrom (1982) shows that a purely individualistic rule

$F_i(y_i)$ Pareto-dominates the interdependent schemes (i.e., $F_i(y_1, \ldots . y_n)$. The individualistic compensation rule, however, losses the peer pressure effect of the penalty contracts. Agents in this system have no incentive to check each other. And of course by assuming Y_i are independently distributed, the possibility of economics of scale is ignored.

The second category on multi-agent settings concentrated on the fundamental difference between single- and multiple-agent problems. Agents may be compared with one another. This possibility is critical because interactions among the agents become important and cause the problem to change in much the same way that the potential for entry alters monopoly problems. More possible compensation rules become available.

The analysis of the compensation rules is not as yet worked out in a general model, and the major papers do not assume a common format. However, the major results are available for a homogeneous contest: assume all agents are identical. In addition to restricting the agent's tastes to be the form (2-1), it's assumed that the principal is risk neutral and that production is subject to an additive common shock and an agent-specific idiosyncratic disturbance. Given this set-up, Green and Stokey provide the result that as the distribution of the additive common shock in the production function becomes less concentrated, $F_i(y_i)$ will eventually dominate $F_i(y_1, \ldots . y_n)$.

The third category includes works that place agency framework in a proper game theoretic framework. In particular, multiple-agent problems are posed as games of incomplete information (Harsanyi, 1967; Groves, 1973; Harris and Townsend, 1981). The analysis begins with an important restriction common to many games of incomplete information: altogether, the players in the game have complete information at the time decisions are taken, there is no "social risk". This means that agent i observes his random state of nature Ω_i prior to choosing his action, and there are no other random disturbances. The second assumption is that players may communicate with the principal, and thus there is the possibility that compensation may be based on the information they supply. Choices about the compensation rules and effort level made conditional on Ω_i are called "parameter contingent." The first result is that the only allocations that can be achieved in a game are parameter contingent. The second result is that only those parameter contingent allocations for which agent i rationally reveals Ω_i are achievable in a game. An important extension of this work is to situations involving social risk. When there is social risk, a natural candidate for a solution concept in a game with communication is the rational expectations equilibrium---when making their decisions, the principal and agent i make use of the correct conditional distribution of the unknown entities. However, in contrast to the case of no social risk,

there is simply no way for the players to check this ex post; one observation dose not generally discriminate among distributions. At present a good solution to this problem is not available. Even though I do not intend to develop a general solution to this problem, this study develops some propositions for specific situations.

The multiple-agent models also involve other issues concerning the interaction of agents. First, consider a simple tournament model with just two agents. It has been suggested (Lazear and Rosen, 1981) that there is scope for destructive activity. That is, for obvious reasons agent i may wish to destroy part of agent j's output. Second, the agent can also interact by colluding, and in a tournament setting such incentives are strong. The basic agency problem arises because effort is costly to observe. Analysis of collusive arrangements supposes the agents to have an advantage in observing each other and to be unamenable to the principal's offers to buy his information. In this study, I will take into account the possibility of communications between agents. I also try to find other ways, besides penalty contracts, to solve the free-rider problem.

The basic issue in the dynamic models is whether the fact that the agency relationship lasts over time can generate more efficient outcomes. Recall that the single-period game, like most non-co-operative games, rarely generates fully efficient outcomes (Dubey, 1983). Put simply, when one agent chooses

among actions given the other players' choice, he does not take into account the implied changes in the other player's pay-off. Essentially there is an uninternalized externality, naturally leading to inefficiency. Allowing for more than one period allows one player's action to generate a future response by the other player. There is at least the possibility of a more efficient outcome.

There are three basic kinds of approaches to dynamic single-agent models. Radner (1981)(see also Malueg, 1981; Rogerson, 1982; Rubenstein and Yaari, 1983; and Lambert,1983) shows that in a T-period repetition of the single-period-agency problem, a Pareto-optimal pair of compensation rule and effort level can be sustained by the strategy of permanently reverting to the one-period Nash action next period, should the other player not choose the Pareto-efficient action in the present period. Obviously the principal must have a method for deciding when suboptimal effort has occurred. In the presence of agency-type informational asymmetries, rules of inference can be constructed that allow efficient outcomes to be sustained.

The second dynamic approach is that of Becker and Stigler (1974). Therein it is shown that the agent can be induced to behave efficiently by underpaying him early in the game and overpaying him later. The rules of the game are then that should the agent behave inefficiently at some point, he loses the future overpayment through being dismissed.

Finally, Fama (1980) argues that competition alone will handle most agency problems if managers are concerned with the capital value of their reputations. Holmstrom (1983) shows that this proposition is correct under particular assumptions, but risk aversion and discounting generally render it false.

A very small amount of work on dynamic multiple agent models is presently available. The approaches in the dynamic-single-agent models are not conducive to analyzing the Chinese situations. This study develops a different approach.

Another extension of the basic agency theory is the theory of hierarchy. Agency-type notions are prominent in early work on hierarchies (Tuck, 1954; Simon, 1965; Williamson, 1967) but are never incorporated in a manner sufficiently precise to generate testable predictions. The leading example of modern work on hierarchies is Rosen (1982). Therein pyramidal structures emerge as a result of the public good aspect of management, treated as a factor of production. Management interacts with workers to produce output used as an input in conjunction with higher management time one hierarchical step up. The more recent works on multi-level structures include Tirole (1986), and Cauley and Sandler (1991) which discussed the formation of coalitions between the supervisor and workers. It has been pointed out that multiple levels may be plagued with agency costs owing to coalitions, even when intervening supervisory levels have more information than the principal. Coalitions between the supervisory and agent level

may form such that both parties may gain at the expense of the principal. The agency theory reviewed above correctly identified the major issues, but it has very little positive content. If concrete results are to be derived, more specific models should be developed with limited feasible payment schedules, a specified type of uncertainty, and the utility functions of both agent and principal. Cauley and Sandler (1991) extended theory in this direction. By extending the so-called Linear-Exponential-Normal (LEN) model, presented by Spremann (1987), Cauley and Sandler assume that:

1) There is one principal (e.g., a manager) and a team of n agents (e.g., workers). The principal can view the final outcome, in terms of output or profit, of agents' action or effort, but is unable to observe the team's actual action. The payments will depend partly on the collective output of the team.

2) All agents are identical in ability and behavior.

3) The principal is risk neutral. The agents are risk averse. The agents possess an exponential utility function. The disutility of effort is assumed to be a quadratic function.

4) The payment schedule and the output function are linear. The principal's gross (random) wealth is a linear function of the agent's total effort, X, and the exogenous risk u which is normally distributed with a zero mean and $var(u)=\sigma^2$.

5) Agent i takes the effort level of the rest of the team as constant (Nash assumption) when he or she is making the decision of what effort level to offer.

The model includes the following equations

$$Y=X+u \tag{2-3}$$

$$X=\sum_{i=1}^{n} x_i=x_i+X_{-i}=x_i+\sum_{j\neq i}^{n} x_j \tag{2-4}$$

$$F(Y)=w_0+s(Y/n) \tag{2-5}$$

$$U_i=-\exp(-\alpha w_i) \tag{2-6}$$

$$w_i=w_0+s(Y/n)-x_i^2 \tag{2-7}$$

$$\text{or} \quad w_i=w_0+s(x_i+X_{-i}+u)/n-x_i^2 \tag{2-8}$$

where Y: The principal's gross wealth

X: The total effort of the team of n agents

u: The random shock of Y

x_i: The effort level of agent i

X_{-i}: The effort level of the team without agent i

$F(Y)$: The payment schedule for each agent

w_0: The fixed wage

s: The ratio of share of Y, sY is the part of Y which goes to the team

U_i: The utility of agent i

w_i: Agent i's random wealth

n: The team size

The certainty equivalent utility of agent i is

$$U_i(X, x_i, w_0, s) = w_0 + \frac{sX}{n} - x_i^2 - \frac{\alpha s^2 \sigma^2}{2n^2} \qquad (2\text{-}9)$$

where the last term is a risk premium.

Maximize U_i with respect to x_i and assume X_{-i} is a constant, in case of an interior solution, we get

$$x_{if} = s/2n \qquad i=1,2,3,\ldots\ldots\ldots n \qquad (2\text{-}10)$$

where x_{if} is the optimal work effort under free-riding consideration and for given s. The Pareto-optimal effort level is developed as follows

$$\max_{x_i} \sum_{i=1}^{n} U_i = nw_0 + s(x_i + X_{-1}) - \sum_{i=1}^{n} x_i^2 - \frac{\alpha s^2 \sigma^2}{2n} \qquad (2\text{-}11)$$

The first-order condition for (2-11) implies that

$$x_i' = s/2 \qquad i=1,2,3,\ldots\ldots\ldots n \qquad (2\text{-}12)$$

where x_i' is the Pareto-optimal effort level of agent i for exogenous s. And x_i' is n times as large as x_{if}.

Two results, among others, can be generated from this model:

Result 1: Under the LEN model, if s=0, agents' effort is minimal (zero).

Result 2: For a linear payment scheme, individual effort decreases in proportion to the team size. The larger the team, the more inefficient is individual effort owing to free-riding considerations.

Here the concept "free-riding" is similar to but a little different from the same term used in the public goods literature. In the public goods literature, every individual's contribution to the supply of public goods is consumed by all people in this system, other people's consumption of the public goods will, however, not reduce the consumption of the supplier. If a supplier of public goods takes other's supply as fixed, and his marginal utility of public goods supply is greater than the marginal disutility, we will end up with a free-rider. In the n-agents model, every individual's contribution to the total profit, after deducting the share of the principal, is shared by all agents in the team. For a specific agent, if his utility generated from reduced work effort is greater than the disutility generated from the reduced personal income which is only 1/n of his reduced contribution, we will also end up with a free-rider. Even though in this case other people's "consumption" of the agent's contribution will exclude his consumption, which is the difference from the public goods model, the similarity between the two models is based on the comparison of the utility and disutility discussed above. In both cases, as long as the utility is greater than the disutility, we will have a free-riding problem, so some of the results generated from the public goods literature can also be used here. One of the results is the "Hume theorem" which says that free-riding will more likely to be a problem as the number of

potential cooperators increases (Roumasset and Laine, 1991). As the number of cooperators increases, the share of each supplier's contribution become smaller and smaller. If suppliers take other's contribution as a constant, when a supplier reduces his public goods supply, the marginal utility of reduced cost is more likely greater than the marginal disutility of his reduced public goods consumption. As Roumasset and Laine pointed out, in the Nash non-cooperative solution the marginal social benefits of public goods are n times the marginal social costs, which is a similar result as result 2 in the LEN n-agents model.

From this model we can see that the risk premium has no effect on x_{if} in case of interior solutions. And the optimal s which maximizes agents' utility is 1. Formally, since $X_f=nx_{if}=s/2$, we have

$$U_i(x_{if})=w_0+\frac{sX_f}{n}-x_{if}^2-\frac{\alpha s^2\sigma^2}{2n^2}=w_0+\frac{s^2}{2n^2}(n-\frac{1}{2}-\alpha\sigma^2) \qquad (2\text{-}13)$$

If $(n-1/2-\alpha\sigma^2)<0$, then $U_i(x_{if})<w_0=U_{i0}$. Where U_{i0} is the utility level of agent i when $x_i=0$, but in this case $x_{if}=s/2n$ is no longer optimal, since agents prefer $x_i=0$. This situation will happen if α or σ^2 is very high, that is the degree of risk aversion and/or uncertainty is very high.

If we assume that $\alpha\sigma^2$ is not very large and $U_i(x_{if})>U_{i0}$, then the risk premium will not affect the choice of x_{if}, so we can use the expected utility function of agent i instead of the certainty equivalent utility of agent i in the following

analysis. The expected utility of agent i is

$$EU_i = w_0 + sX/n - x_i^2 \qquad (2\text{-}14)$$

The optimal effort level, x_{if}, developed in this model is critically dependent on the Nash assumption that X_{-i} is a constant. This assumption is very unrealistic for the following reason:

According to this model, the working effort of each agent in a two worker team will be only 1/2 of that in a one worker team. This is unlikely to be true. Because the transaction cost in a two person team is likely to be small. As a rationale of this assumption, Olson (1965) cited that two neighbors may agree drain a meadow, but it is very difficult that a thousand people should agree in any such action.

In the public goods literature (Roumasset and Laine, 1991; Gardner et al.,1990), the Nash non-cooperative solution is also believed to be an extreme and clearly exaggerates the problem. In the "chicken" model, if the consequences of all players' defecting are too severe, then individual players will choose the cooperative strategy. Similarly, if the rewards of cooperation are sufficiently high such that each player knows that the others will cooperate then the free-riding problem may be solved.

In fact, the potential gain to each agent from offering the Pareto-optimal effort (the reward of cooperation) is

$$PG_i = EU_i(x_i') - EU_i(x_{if}) = s^2\left(\frac{1}{4} - \frac{1}{2n} + \frac{1}{4n^2}\right) \qquad (2\text{-}15)$$

For $n>1$, $PG_i>0$, and it will increase with n but not exceed $s^2/4$. Where $x_i'=s/2$, is the Pareto-optimal effort level; PG_i is the potential gain of agent i, if the agreement that all agents will offer Pareto-optimal effort can be reached.

If there is no transaction cost, such agreement will be reached, and agents will offer x_i' rather than x_{if}. If there is transaction cost, the result will depend on the relative value of PG_i and per capita transaction cost, TC_i. It's rational to assume TC_i is a function of n and TC_i will increase with n. There is a critical value of n defined as n_c, for $n>n_c$, we have $TC_i>PG_i$, workers will not reach any agreement and $x_i=x_{if}$; for $n<n_c$, the potential gain of agent i is greater than per capita transaction cost ($TC_i<PG_i$), agreement will be reached and $x_i=x_i'$. The enforcement of the agreement is through peer pressure of co-workers either actively or passively.

One possible extension of Cauley and Sandler's model is to relax the Nash assumption, and develop models which include peer pressure of co-workers. In the case of passive reactions of co-workers, we can assume that X_{-i} takes the following general form

$$X_{-i} = a(n) + b(n)x_i \qquad (2\text{-}16)$$

$$a'(n)>0, \qquad b'(n)<0$$

The first term, a(n), is independent of x_i, and as n increases, a(n) become larger; the second term, $b(n)x_i$, is

directly related to x_i, and as n increases, this part will become smaller. The reason to chose this general form is that only part of the team can observe agent i's effort level. In a small team, all or most of the workers are working together, so most of the workers will know the effort level of worker i; if worker i reduces his effort, the other workers will reduce their effort accordingly. In a large team, however, only a small portion of the workers know the effort level of worker i; if worker i reduce his effort, only a small number of workers will take action and reduce their efforts too. If n is large enough so that b(n)=0, the model reduces to the free-riding situation with the optimal effort level equals to s/2n. If n is small enough so that a(n)=0, the optimal effort level will be the Pareto-optimal effort level, x_i'. In fact, if a(n)=0 we have

$$\max_{x_i} EU_i = w_0 + \frac{sx_i}{n} + \frac{sX_{-i}}{n} - x_i^2 = w_0 + \frac{sx_i}{n} + \frac{sb(n)x_i}{n} - x_i^2 \qquad (2\text{-}17)$$

The first-order condition implies that

$$x_i' = \frac{s+b(n)s}{2n} = \frac{s+s(n-1)}{2n} = \frac{s}{2} \qquad (2\text{-}18)$$

Since $X_{-1}=b(n)x_i=(n-1)x_i$, so $b(n)=n-1$, $x_1=x_2=\ldots\ldots\ldots x_n$

In general, the optimal work effort is

$$\max_{x_i} EU_i = w_0 + \frac{sx_i}{n} + \frac{s[a(n)+b(n)x_i]}{n} - x_i^2$$
$$x_i' = \frac{s[1+b(n)]}{2n} = \frac{s[1+(n-1)-a(n)/x_i]}{2n} = \frac{s}{2} - \frac{sa(n)}{2nx_i} \qquad (2\text{-}19)$$

So we have

$$x_{if}=\frac{s}{2n}<x_i'=\frac{s[1+b(n)]}{2n}<\frac{s}{2} \qquad (2\text{-}20)$$

If workers in a team reduce their efforts when they find other workers reducing effort, and if all workers expect their co-workers behaving this way, the effort level will be greater than the optimal effort under free-riding considerations, but smaller than the pareto-optimal effort level. As team size decreases, the optimal effort will be close to the Pareto-optimal level.

This result points out two directions for increasing effort level. One is making the team size as small as possible. If for some reasons (economics of scale, for example), the team size cannot be reduced further, then we should let as more workers (in the same team) as possible work together, so that b(n) will be relatively larger, and a(n) will be smaller.

Passive reaction is only one of the possible reactions; co-workers may also actively react to stop shirking behaviors. There are many ways co-workers can actively stop or reduce shirking behavior, such as report to managers, directly speak to the shirker, exclude the shirker from the team, isolate the shirker, or put other social pressures on the shirker. The key question is whether or not the benefits of such reactions is greater than the cost. First of all, let's look at the

potential gain of agent j if he can stop agent i from shirking

$$MBP_j = EU_j(x_i') - EU_j(x_{if}) = \frac{s}{n}(x_i' - x_{if}) \\ = \frac{s}{n}\left(\frac{s}{2} - \frac{s}{2n}\right) = \frac{s^2}{2n}\left(\frac{n-1}{n}\right) \approx \frac{s^2}{2n} \qquad (2\text{-}21)$$

where MBP_j is the maximum potential benefit of agent j by put pressure on agent i. MBP_j will increase as n decrease, and it will increase very fast as s increases.

The actual benefit of active response of agent j depends on the effectiveness of such pressure. Agent i may increase effort level but may not increase all the way to x_i'. We assume that the actual benefit to agent j from offering an active response to shirker i is a fraction of MBP_j, so we have

$$ABP_j = CMBP_j \qquad 0<C<1$$

where ABP_j is the actual benefit of agent j if he offers an active response to shirker i. And C is a coefficient which reflect the effectiveness of peer pressure. C may depends on such things as social norms, the punishment from a manager if reported shirking, and measurement accuracy of effort level.

The cost of agent j to offer peer pressure on shirker i basically comes from deteriorated relationship, and the effort to offer this pressure. Most of the costs are independent with respect to team size, only the cost of tense relationship may decrease as n increases. In a larger team, a relationship with one specific person may be less important than that in a smaller team. Roughly, it's still rational to assume the cost of peer pressure, CP, is constant with respect to team size.

If $ABP_j > CP$, worker j will actively put peer pressure on shirker i. Since worker j represents all co-workers of shirker i, the collective effect of all co-workers will be much greater than that of one co-workers. So C will be close to 1. The effectiveness of peer pressure in the Chinese state enterprises is likely to be greater than that of a Western enterprise for the following reasons: i) Since most Chinese state workers expect life time employment in the same job, the relationship with co-workers become more important as workers expect that they will work together for their whole life. ii) Chinese people often care more about their relationship with co-workers. The economic reason is that some social functions which in the West are the job of the society, are taken care by friends in China. For example, if a worker need to borrow money, in the West, very often he will go to the bank and asks for loans; the Chinese worker, however, generally cannot do that, he has to borrow from his friends. If a worker has a conflict with someone, the Western worker may go to the court to settle this conflict, while the Chinese worker depends on his friends to settle this conflict. Because many functions are not marketized in China, workers depends heavily on their friends, the quasi-rent generated from good relationship with co-workers is higher in the Chinese case. There is also a cultural reason which explains why these relationships are more important in the Chinese society (Zhuang, et al 1991). The high cost of tense workplace relations is a "two-edged

knife." In one hand, if ABP_j is not high enough, such that $ABP_j<CP$, agent j will not put peer pressure on shirker i, because the damage caused by shirker i is not high enough; on the other hand, since ABP_j increases as s increases and/or n decreases, if $ABP_j>CP$, agent j will put pressure on shirker i. Since the shirker also care about workplace relations, when all co-workers start put pressure on him, he will take it seriously. In China, CP is high, however, it's still possible to have $ABP_j>CP$, such that co-workers are willing to put pressure on shirkers (Alan, 1990). The effect of peer pressure is not necessarily additive, but as more co-workers put peer pressure on one workers, the collective effectiveness will be greater than that of a single co-worker. And in some cases, the teams are organized voluntarily. If most workers in a team do not like someone, they can simply exclude the worker from the team. If no team wants this worker, he will have to accept some unattractive job. In this case the collective effect will be much greater than the effect of a single co-worker.

The benefit of active reaction of agent j against shirker i is positively related to s^2 and negatively related to n. The cost of such reaction is basically a constant. So as team size decreases (or s increases) to a certain level, co-workers start to put peer pressure on shirkers. And x_i increases toward the Pareto-optimal level.

This model also points out several important directions to reduce shirking or free-riding behavior:

1) Reduce team size so that the potential gain of co-workers to stop shirking behavior is high enough to offset the cost of offering peer pressure on shirkers.

2) The same result can be reached by selecting a higher s, but the principal may be worse off if s is too high. If n is too large, increasing s may not cause $ABP_j>CP$. Thus n should be reduced to be as small as possible, when n cannot be reduced any more and ABP_j is still greater than CP, increasing s may lead to $ABP_j>CP$.

3) Increase C, the coefficient of effectiveness of peer pressure. This can be done in many ways, for example, if the team can exclude shirkers through majority voting, the effectiveness of peer pressure will be bigger, if the social norm treat shirking the same way as stealing, and all workers look down upon shirkers, we will expect a higher C. Or if co-workers just report shirkers to the manager, and the managers take serious actions to punish the shirker, the shirker will care more about the peer pressure and the value of C will be higher.

4) Reduce CP. There are many ways that CP can be reduced. For example, if co-workers can report shirkers secretly, CP will be reduced.

In all, the situation is similar to the oligopoly situation. A successful oligopoly depends on potential

oligopoly rent, the number of members involved in the oligopoly, and the measurability of the behavior of the members. The key factor is very often the number of members. The same is true in the team incentive payment system.

In conclusion, most literatures of agency theory assume very general forms of utility functions. They correctly identified the major issues, but there are very little positive content. If concrete policy implications are to be derived, more specific forms of utility functions should be assumed. The LEN model assumed a specific utility function, and generated some interesting results. However, some important issues are ignored because of the simplified assumptions of the model: i) assuming that the principal's gross (random) wealth is a linear function of the agents' total effort X and the exogenous risk u (2-3) would ignore the effect of capital. Obviously, at least in the short run (when capital stock is fixed), for the same work effort, workers with high capital-labor ratio would generate more wealth. ii) Assuming the collective work effort of the team without agent i, X_{-i}, is a constant would ignore the possibility of cooperations between coworkers. iii) Assuming there is only one period would ignore the effect of long-term employment. In chapter 3, I introduce a more realistic agency model that builds on the LEN model.

CHAPTER 3

THEORETICAL MODELS

Since work effort is an abstract concept, to compare work effort quantitatively, I define work effort, x, as the equivalent of full effort working hours per working day. If nominal working hours is 8 hours per day, then x=8 means full work effort, and x=4 means that the work effort is so low such that the total work accomplished in 8 hours could have been accomplished in 4 hours if full work effort were supplied. And I define the full work effort as x_{max}. In this chapter:

1) I assume that the return (or the total contribution) of work effort Y, is composed of a nonrandom part, Px, and a random part ϵ; Px is a linear function of x, and ϵ is normally distributed with a zero mean and a finite variance. Y can be interpreted as value added and $Y=Px+\epsilon$. Y is distributed between the principal and agents. The effects of capital and price distortions are included by the introduction of P. A high P value is associated with a high capital-labor ratio and/or a high price. For the same work effort, the wealth generated would be higher if the worker has a higher capital-labor ratio and/or his product is favored by price distortions.

2) I assume that the agent has to supply a minimum work effort which is the level of effort elicited under available

weak monitoring techniques; assume that monitoring is exogenous and ineffective such that the minimum work effort is sufficiently below the full work effort; assume that the principal can affect the minimum work effort through affecting the exogenous monitoring, he can affect P through capital investment, but he cannot affect P through his effort in management, that is I assume that the production function is separable.

3) I assume that there are n agents and one principal; the team size n is small enough, so that there is no free-riding problem. And workers are identical so that $X=nx_i$, where X is the collective work effort of the team and x_i is the work effort of agent i, this assumption is relaxed latter.

4) I assume that both the principal and the agent are risk neutral; and there is only one period, these assumptions are relaxed latter.

5) I assume that agents are derive utility from money income and disutility from work effort. To preserve the strict concavity of the agents' utility function and simplify the analyses, following Cauley and Sandler (1991), I assume that the disutility is a quadratic function of work effort, this assumption is relaxed in appendix 2.

Five payment systems are discussed in this chapter: the fixed wage (FW) system; the profit-sharing (PS) system; the disutility of effort compensation (DEC) system; the extended PS (EPS) system; and the lump-sum profit turn-over (LST)

system. All incentive payment systems are compared with the FW system which as a starting point is not a incentive payment system.

3.1. The Fixed Wage System

The FW system is defined as follows: before real work is done, an explicit or implicit contract (agreement) between the principal and the agent is reached; in return to a wage payment the agent promised to offer a certain work effort level. The wage payment, even though subject to changes when new contracts are negotiated, is fixed during the period when the current contract is effective regardless of the real work effort. If the period of the "contract" is short, we call it short-term fixed wage (SFW) system; if the period is long, we call it long-term fixed wage (LFW) system. Under the SFW system, payment is dependent on work effort in principle; high work effort may lead to promotions. However, due to the poorly defined property rights, promotions in the Chinese state enterprises must be based on "black and white" evidence of work effort level, which is very often unavailable. So even though promotions are based on work effort and performance in principle, seniority is in fact the most important criterion, as long as a minimum work effort is supplied. As a result, payment is basically independent of work effort even in the SFW system. The SFW system in Chinese state enterprises also has the following problems:

1) Even in the SFW system, the interval between two wage adjustments is still too long, such that the information of work effort is biased. In principle, general wage adjustments should compensate total work effort between two wage adjustments. The principal is, however, "forgetful." The performance closer to wage adjustments may give the principal a "fresh impression" which may have a larger effect than the performance far before wage adjustments, especially when the interval between wage adjustments is several years. Unambiguous measurement of work effort is unlikely available in such a long period.

2) Wages are generally downward stick. When the probability of further promotions becomes rare, senior workers may reduce work effort and without being punished.

In the following discussion, I use FW for both LFW and SFW, and assume the period of the "contract" is long enough, such that payment is independent of work effort.

Under the FW system, there is no uncertainty with respect to the agents, for work effort x_i, the disutility of effort of agent i is x_i^2; the utility of money income of agent i is w_0 which is independent of work effort. The expected utility function of agent i is

$$EV_i=w_0-x_i^2 \qquad (3-1)$$

Maximizing EV_i with respect to x_i subject to x_i greater or equal to the minimum work effort, x_{i0}, the optimal effort of agent i under the FW system is x_{i0} (figure 1).

For effort level x_i, $Y=PX+\epsilon=nPx_i+\epsilon$, and the total profit of the principal is Y minus nw_0. Since we assume that the principal is risk neutral, the expected utility of the principal under the FW system is

$$EV_m=n(Px_i-w_0) \tag{3-2}$$

Since the optimal work effort is x_{i0}, the expected utility of the principal is $n(Px_{i0}-w_0)$. For given w_0, P, and n, the principal can only increase EV_m through strengthened monitoring, such that x_{i0} increases. If x_{i0} is also fixed, as I assumed, the principal is incapable of increasing EV_m.

3.2. The Profit Sharing System

Under the PS system, the principal takes the value of nPx_{i0} as a reference point, if the observed value Y is greater than nPx_{i0}, the difference between the two values is shared by the principal and agents. So for effort level x_i, the disutility of effort is the same as in the FW system; the utility from money income of agent i becomes $w_0+s(Y-nPx_{i0})/n$, the expected utility of agent i is

$$EV_i=w_0+sP(x_i-x_{i0})-x_i^2 \tag{3-3}$$

(see appendix 3 for a formal derivation of EV_i). Similarly we can show that the expected utility of the principal is

$$EV_m=(1-s)nPx_i-n(w_0-sPx_{i0}) \tag{3-4}$$

where x_i is the work effort of agent i, i=1,2,.....n; the other symbols are the same as defined before; w_0 is the utility generated from a subsistence wage or the wage level

under fixed wage and ineffective monitoring system when $x_i=x_{i0}$.

The PS system is a Pareto-improving system compared to the FW system, that is, neither the principal nor the agents can become worse off. After introducing the PS system, if workers still offer x_{i0}, the principal does not increase wage payments. If workers increase effort, the principal is better off. From the workers' point of view, if they calculate that the increased payments are not worth their effort, they can always reduce their effort to x_{i0}, and not be worse off. If they feel the increased payment is worth their effort, they increase effort and become better off. So both parties should welcome this system. In sum, the PS system is a Pareto improvement over the FW system.

3.2.1. Optimal Effort and Optimal Ratio of Share

In the PS system, s is selected by the principal either exogenously or endogenously. The agents take s as given and select the optimal work effort. If we assume agents are risk neutral, then the optimal effort level of agent i under exogenous s can be developed as follows

$$\max_{x_i} EV_i = w_0 + sP(x_i - x_{i0}) - x_i^2 \qquad S.T.\ x_{i0} \le x_i \le x_{imax} \qquad (3\text{-}5)$$

In case of interior solution, the first-order condition for (3-5) implies that

$$x_i' = sP/2 \quad (\text{for } x_{i0} < x_i' < x_{imax}) \qquad (3\text{-}6)$$

where x_i' is the conditional optimal effort, conditional on exogenous s; and x_{imax} is the full work effort. If P is too low or too high such that sP/2 is smaller than x_{i0} or greater than x_{imax}, then x_i' will equal to x_{i0} or x_{imax} accordingly (the corner solutions).

Proposition 1: Under the PS system, if agents expect all other agents in the same team act the same way as he does ($X=nx_i$), s is exogenously given, and P is not too small or too large (interior solution), then the conditional optimal work effort is positively related to s; i.e., increases in s induce greater work effort.

Agents obviously prefer a high s, in fact EV_i is maximized when s=1. If x_i' is smaller or equal to x_{i0}, by definition of the PS system, s=0. The principal's optimal s can be developed by maximizing EV_m with respect to s subject to: i) $x_i=x_i'=sP/2$; ii) $x_i'\leq x_{imax}$ or $s\leq 2x_{imax}/P$. The Lagrangian function Z appears in the form

$$Z=n(1-s)Px_i'-n(w_0-sPx_{i0})+\lambda\left(\frac{2x_{imax}}{P}-s\right) \qquad (3\text{-}7)$$

And the Kuhn-Tucker Conditions imply that

$$s^*= 1/2 + x_{i0}/P \quad \text{for } x_{i0}\leq x_i'\leq x_{imax} \qquad (3\text{-}8)$$

$$s^*= 2x_{imax}/P \quad \text{for } x_i'>x_{imax} \qquad (3\text{-}9)$$

where s* is the optimal s of the principal. (3-8) is the interior solution; (3-9) is one of the corner solution, another corner solution is s=0 when $x_i'=x_{i0}$.

I define the value of x_i' when $s=s^*$ as the unconditional optimal effort, x_i^*. In case of interior solution, we have

$$x_i^* = \frac{Ps^*}{2} = \frac{P}{4} + \frac{x_{i0}}{2} \tag{3-10}$$

If the P value is too high or too low, we may have corner solutions. Define P_{min} as the P value when $x_i^*=x_{i0}$ and P_{max} as the P value when $x_i^*=x_{imax}$, for $P<P_{min}$, $x_i^*=x_{i0}$; for $P>P_{max}$, $x_i^*=x_{max}$. Using (3-10), set $x_i^*=P_{min}/4+x_{i0}/2=x_{i0}$, I can solve

$$P_{min}=2x_{i0} \tag{3-11}$$

set $x_i^*=P_{max}/4+x_{i0}/2=x_{imax}$, we have

$$P_{max}=4(x_{imax}-x_{i0}/2) \tag{3-12}$$

In general, the optimal work effort under the PS system can be summarized as

$$x_i^* = \begin{cases} x_{i0} & if\ P \le P_{min}=2x_{i0} \\ (\frac{P}{4}+\frac{x_{i0}}{2}) & if\ P_{min}<P<P_{max} \\ x_{imax} & if\ P \ge Pmax \end{cases} \tag{3-13}$$

The optimal ratio of share of the principal under the PS system can be summarized as

$$s^* = \begin{cases} 0 & if\ P \le P_{min} \\ \frac{1}{2}+\frac{x_{i0}}{P} & if\ P_{min}<P<P_{max} \\ \frac{2x_{imax}}{P} & if\ P \ge P_{max} \end{cases} \tag{3-14}$$

Proposition 2: Under the PS system, if $X=nx_i$, s is not fixed by the principal (endogenous s), and P is not too small or too large (interior solution), then x_i^* is positively related to P and x_{i0}.

Proposition 3: Under the PS system, the optimal s is negatively related to P, and positively related to x_{i0} or x_{imax} depending on the P value.

Propositions 2 and 3 can be generated directly from (3-13) and (3-14). Proposition 2 can be seen more clearly in figure 2, where we can see that the optimal effort x_i* is the effort when the slope of the disutility of effort curve which is increasing as x_i increases equals the slope of sPx_i curve which is positively related to P and s. So as P or s increases, x_i* will increase too.

If s=s* is endogenously chosen, from (3-14) we can see that as x_{i0} increases, in case of interior solution, s* will increase, and x_i*=s*P/2. So increases in x_{i0} also induce increases in x_i*.

If $P>P_{max}$, the principal can hire new workers, so $P<P_{max}$ is more likely to be the case, because of labor stickiness in the Chinese situation, however, $P>P_{max}$ could also happen, I will discuss both cases but focus on the situation that $P<P_{max}$.

3.2.2. Potential Gains of the Principal and the Agents

Even though both the principal and the agents may benefit from switching the FW system into the PS system, the potential gains can be different under various situations. To analyze the incentives to accept the PS system, we need to look at the potential gains of both parties. The potential gains of the principal and agent i in case of interior solutions are

$$PGP=EV_m(s*,x_i*)-EV_m(x_{i0})=\frac{n}{2}(\frac{P^2}{4}+x_{i0}^2-Px_{i0}) \qquad (3\text{-}15)$$

since $P>P_{min}=2x_{i0}$ we have PGP>0 and

$$\frac{\partial PGP}{\partial P}=\frac{n}{2}(\frac{P}{2}-x_{i0})>0 \qquad (3\text{-}16)$$

$$\frac{\partial PGP}{\partial x_{i0}}=\frac{n}{2}(2x_{i0}-P)<0 \qquad (3\text{-}17)$$

$$PGW_i=EV_i(s*,x_i*)-EV_i(x_{i0})=\frac{1}{4}(\frac{P^2}{4}+x_{i0}^2-Px_{i0})>0 \qquad (3\text{-}18)$$

$$\frac{\partial PGW_i}{\partial P}=(\frac{P}{8}-\frac{x_{i0}}{4})>0 \qquad (3\text{-}19)$$

$$\frac{\partial PGW_i}{\partial x_{i0}}=\frac{1}{4}(2x_{i0}-P)<0 \qquad (3\text{-}20)$$

where PGP is the potential gain of the principal.

PGW_i is the potential gain of agent i.

If s is exogenous, I define $EV_i(x_i')-EV_i(x_{i0})$ as the conditional gain of agent i, CGW_i. And

$$\frac{\partial CGW_i}{\partial s}=P(\frac{sP}{2}-x_{i0})>0 \qquad for\ x_i'=\frac{sP}{2}>x_{i0} \qquad (3\text{-}21)$$

Proposition 4: Compared with the FW system, if P is not too small or too high, both the principal and agents gain in the PS system. The potential gain to both parties increases as P increases and/or as the minimum effort decreases. So the PS system should be welcome in enterprises with high productivity and ineffective monitoring.

3.2.3. Analysis of the Free-Riding Problem

The general features of the free-riding problem have been discussed in chapter 2, this model has, however, something new. This section focuses on the new features of the model.

Optimal effort x_i* is critically dependent on the assumption that $X=nx_i$. Under this assumption, each worker expects all other co-workers to behave the same way he does. If the team size is large, we may expect that only part of the co-workers' behavior is affected by agent i, because in a large team, agent i's effort is unlikely to be observed by all team members. All team members working together and observe each other's effort is likely to occur only in a small team. In general, I assume that there are n agents in a team, among the n agents m agents are working together with agent i. If agent i supply x_i, he expects that the m agents will also supply x_i, because they observe each others' work effort; and the collective work effort of the rest of the team is not affected by agent i. So the collective work effort of the team is

$$X=mx_i+(n-m)c \qquad m<n \qquad c=\text{constant} \tag{3-22}$$

the optimal work effort of agent i is generated as follows

$$\begin{aligned} \max_{x_i} EV_i &= w_0+\frac{sPX}{n}-\frac{sPX_0}{n}-x_i^2 \\ &= w_0+\frac{sPmx_i}{n}+\frac{sP(n-m)c}{n}-sPx_{i0}-x_i^2 \end{aligned} \tag{3-23}$$

the first order condition implies

$$x_{if}=sPm/2n \qquad i=1,2,\ldots\ldots\ldots\ldots n \tag{3-24}$$

where x_{if} is the optimal effort of agent i under free-rider considerations. If m=1, we have the Nash assumption and $x_{if}=sP/2n$; if m=n, $x_{if}=x_i^*=sP/2$.

Let's look at the extreme case that m=1 first. Under the Nash assumption, $x_{if}=sP/2n$, the principal's optimal s is

$$\max_s EV_m=(1-s)Pnx_i-n(w_0-sPx_{i0}) \quad S.T. \quad x_i=x_{if}=\frac{sP}{2n} \qquad (3\text{-}25)$$

$$F.O.C \quad \Rightarrow \quad s_f=\left\{\begin{array}{ll} \frac{1}{2}+\frac{nx_{i0}}{P} & if\ n\le n_{max} \\ 0 & if\ n>n_{max} \end{array}\right\} \qquad (3\text{-}26)$$

where n_{max} is the maximum team size above which $x_{if}<x_{i0}$, so $x_i=x_{i0}$ and $s_f=0$.

Setting $x_{if}=sP/2n$ equal to x_{i0}, I can calculate n_{max} as

$$n_{max}=sP/2x_{i0} \qquad (3\text{-}27)$$

If $s=s_f=1/2 + nx_{i0}/P$ (endogenous s) then

$$n_{max}=P/2x_{i0} \qquad (3\text{-}28)$$

Proposition 5: The maximum team size under which x_{if} is still greater than x_{i0} is positively related to the P value and negatively related to x_{i0}.

For example, if P=24, $x_{i0}=4$, $x_{imax}=8$, from (3-28) we have $n_{max}=3$. For n<3, however, the Nash assumption is very unlikely to hold.

If P=28, $x_{i0}=2$, $x_{imax}=8$, then $n_{max}=7$.

If P=30, $x_{i0}=1$, $x_{imax}=8$, then $n_{max}=15$.

From (3-12), we can see that the P values we arbitrarily selected are high enough to yield upper bound solutions for x_i in the non-free-riding model ($P=P_{max}$), still the n_{max} values are

very low. So, unless we assume the P value is very high, which is unlikely to hold in the Chinese situation, x_{if} is very likely smaller than x_{i0}, because actual n is very likely greater than n_{max}.

If P=28, x_{i0}=2, x_{imax}=8, then n_{max}=7. For n>7, $x_{if}<x_{i0}$, so that $x_i=x_{i0}$. If s is exogenously given (as in most Chinese cases), n_{max} is even smaller since in general s<1 (see 3-27 and 3-28). If actual n is large enough so that the Nash assumption holds, increasing incentive payment by offer a larger s or increase the P value will not increase work effort at all. To increase effort, we have to reduce the team size so that the Nash assumption no longer holds.

This result can be used to explain the Chinese situation. I assume there is a critical value of n, defined as n_c. Above n_c, the transaction cost is too high, so that the Nash assumption holds; below n_c, the transaction cost is small enough, so that agreement of co-workers will eliminate free-riding behavior. Assume, for the moment, that n_c=10, actual n is 20, s=0.2, x_{i0}=2, x_{imax}=8, P=56, then n_{max}=2.8. If the principal wants to increase x_i by increasing s to 0.5 and increasing P to 102, then n_{max} increases to 14. Since n is still greater than n_{max}, $x_i=x_{i0}$ will not change. In this case x_i will not increase until n is reduced below n_c or at least below n_{max}. In case of Chinese rural production team, with $n>>n_c$, only increases in the prices of agricultural products would unlikely increase work effort.

The free-riding behavior analyzed above is an extreme case that noncooperative agents involved in a prisoner's dilemma game. If we extend the one period model to a multi-period model, repeating games can yield cooperation according to the folk theorem (David, 1985). In fact, in the repeated game, agent i will find that if he reduces his work effort, his co-workers reduce their effort too. For a team size $n>n_{max}$, all noncooperative agents are only willing to supply x_{i0}. There are only two possible equilibria, either all agents supply x_{if} (x_{i0} if $n<n_{max}$), or all agents supply x_i* (the cooperative optimal work effort), and agents are better off if they all supply x_i*.

The potential gain from offering x_i* (the reward of cooperation) depends on whether $x_i=x_{i0}$ or $x_i=x_{if}$ in the first place. If $x_i=x_{i0}$ (which is likely as we discussed above), the potential gain of agent i is PGW_i (3-18); if $x_i=x_{if}$, the potential gain of agent i is

$$PGW_{if}=EV_i(x_i*)-EV_i(x_{if})$$
$$=\frac{P^2}{8}\left(\frac{1}{2}-\frac{1}{n}+\frac{1}{2n^2}\right)+\frac{x_{i0}^2}{2}(n-1)+\frac{Px_{i0}}{4}\left(\frac{1}{n}-1\right)>0 \qquad (3\text{-}29)$$

$$\frac{\partial PGW_{if}}{\partial n}=\frac{P^2}{8}\left(\frac{n-1}{n^3}\right)+\frac{x_{i0}}{2}\left(x_{i0}-\frac{P}{2n^2}\right)>0 \qquad (3\text{-}30)$$

for $n>1$ and $P>P_{min}$

Proposition 6: The reward of cooperation increases as team size increases, however, it cannot exceed PGW_i.

If there is no transaction cost, workers may reach agreements so that $x_i=x_i*$, and there is no free-riding behavior. If transaction costs are positive, effort level depends on the relative value of the per capita transaction cost and PGW_i or PGW_{if}. In general, reducing n reduces transaction cost, but does not affect PGW_i. So if keep reducing n, we can always solve the free-riding problem. If n cannot be reduced further for some other reasons, and transaction cost cannot be reduced either, we can increase the P value (price increase, or productivity increase through investment or introduce new technology) to increase PGW_i. If s is not equal to s*, we can increase the conditional gain of worker i, CGW_i, by narrowing the gap between s and s* (see 3-21). If CGW_i is greater than per capita transaction cost, we can eliminate free-riding behavior. Intuitively, if the consequence of all agents' defecting is too severe, agents would choose the cooperative strategy. Increasing CGW_i leads to a severer consequence of defecting, which may induce agents choose to cooperate.

In general, we have $1<m<n$, $x_{if}=sPm/2n$, and $n_{max}=sPm/2x_{i0}$, so n_{max} is positively related to m. As m increases, n_{max} increases, for a given n, it is more likely that $n<n_{max}$, and x_{if} increases toward x_i* (3-24).

The Chinese state enterprises can be divided into three categories, In the first category, m is very small and/or n is very large such that $n>n_{max}$. In this case, changing from the

FW system to the PS system does not affect work effort. In the second category, either m is very large or n is very small such that m=n. In this case, there is no free-riding problem. Thus changing from the FW system to the PS system has a significant effect on work effort. In the third category, m is somewhere between 1 and n, and $n<n_{max}$. Changing from the FW to the PS system has a positive, but less significant, effect on work effort than that of category 2. For a given n, the effect increases with m.

3.2.4. The Long-Run and the Socially Optimal Efforts

The results developed in this model can be used to explain a situation where $x_i<x_{imax}$. If $x_i<x_i'$, the optimal effort level under exogenous s, the reason might be that X is not equal to nx_i or we have a free-riding problem. If $x_i=x_i'$, but $x_i'<x_i*$, the reason might be that the principal made a mistake (or the government made a mistake because s is likely determined by government) so $s<s*$. If $x_i=x_i*$, but $x_i*<x_{imax}$, the model seems suggesting that it is rational or optimal to offer a below maximum effort level.

This conclusion is, however, still conditional on the one-period PS system. There are at least two reasons to doubt this conclusion.

1) The optimal effort, x_i*, is determined under the assumption that workers maximize only short-run (one-period) expected utility, EV_i. Up to now, all discussions are based

on a one-period model, in which the P value is independent of work effort in this period. Workers chose optimal work effort taking P as a constant. The P value is, however, positively related to capital stock, and the capital stock may in turn positively related to previous EV_m values which are positive functions of work efforts, either because imperfect capital market, so that the profits of the principal can only be invested in the same team (or consumed), or because the rate of return to capital is low elsewhere. If workers expect long-term employment, as in the case of Chinese state enterprises, they should maximize the present value of their life-time utility, instead of the one-period utility EV_i. The present value of their life-time utility, of course, depends on their time preference. To simplify our analysis, I assume workers are neutral with respect to time. This assumption is relaxed latter. And I assume, for the moment, there are only two periods, t_1 and t_2. The difference between a one-period and a two-period model is very obvious. In the two-period model, when a life-time employee determines his optimal work effort at the beginning of t_1, he may realize that the P value in t_2, P_2, will be affected by his work effort in t_1, because P_2 is a positive function of investment in the beginning of t_2, I_2, and I_2 is a positive function of EV_m in t_1, which is positively related to x_i in t_1. In the two-period PS model, when work effort is greater than x_{i0}, not only the extra profits generated is shared by the agents and the principal,

the increased work effort will also increase P_2, and agents are better off according to (3-19). Intuitively, we may expect that the long-run optimal work effort in t_1 is greater than the one-period optimal effort x_i*. To confirm this intuition, let's take an hypothetical example, with arbitrarily selected numbers, I want to show that the long-run optimal work effort is greater than the one-period optimal work effort. Using this example, I can also demonstrate the underlying mechanisms of the multi-period and one-period PS system in terms of returns of capital investment, and distribution of profit between the principal and the agents. It is interesting to realize that the returns of new investment is significantly swelled because of the switch of payment systems (from FW to PS) and the existence of below maximum work effort. Imagine there is a principal and a team of agents, assume team size is one for simplicity, assume the P value is a linear function of capital stock, or $P=KP_0$, where P_0 is independent of capital stock K, and assume the agent is neutral with respect to time. There are two periods, t_1 and t_2 in this model. Before t_1, the principal and the agent is under the FW system with a fixed wage payment w_0 and $x_{i0}=4$ (4 full effort working hours each day). The capital stock is 100 and $P_0=0.2$, so the P value is 20. Since the optimal work effort under the FW system is x_{i0}, the total payoff of the work effort is $Px_{i0}=80$. The difference between Px_{i0} and w_0 is the return to capital and management. Because there are no factor

markets, w_0 is determined arbitrarily by the principal (meet the subsistence requirements for example). Since our interest is not the level of w_0, I simply assume $w_0=Px_{i0}$. So if the FW system continuous, nothing will change in t_1 and t_2. At the beginning of t_1, however, the PS system is introduced. The agent is informed that 60% of the output above Px_{i0} will be paid to him as an incentive payment (s=0.6). Now we can compare two cases.

In case 1, employment is based on one period. In t_1, the agent do not know whether or not he is going to be employed in t_2. So he chooses the one-period optimal effort $x_i*=sP_1/2$ which is 6 in this case. .Relative to the FW system, his expected utility increased by 4. The utility of the principal increased by 16. This distribution seems unfair, however, it is the optimal choice of the agent under this system. Had the agent supplied an effort different from x_i*, his utility increase world be lower than 4. It's easy to verify than when $x_i=5$, his utility increase is 3; when $x_i=7$ his utility increase is also 3. These net gains are caused by the payment system change. Assume the principal invested all his net gain, 16 in this case, in the same agent. Total capital in t_2 is 116, and P_2 is 23.2. At the beginning of t_2, the agent is informed that he is employed in t_2. His optimal work effort in t_2 is $sP_2/2=6.96$; and his utility increase relative to the FW system is 8.76 in t_2; his total utility increase during the two periods is 12.76. The principal, on the other hand, increased

his utility by 40.3 in t_2, which can be divided into two parts, one is the increased value of Px_{i0}, $(P_2-P_1)x_{i0}=12.8$, the other is the increased value of $(1-s)(P_2x_i-P_2x_{i0})=27.47$. Alternatively, the return of capital investment in the FW system is much smaller than that of the PS system; the difference swells as x_{i0} decreases. If the principal do not invest in the same team, he would continuously gain 16 as in t_1. The difference between 40.3 and 16 is the return to his new investment. The total utility increase of the principal and the society during the two periods is 56.3 and 69.6 respectively.

In case 2, employment is based on 2 periods. And the agent knows that the principal will invest all his gains in the same team (required by law for example). In t_1, if the agent set $x_i=8>x_i*$, his utility increase is zero, the utility increase of the principal is 32. In t_2, however, the P value increased to $P_2=26.4$, the agent set $x_i=x_i*=7.92$. His utility increase is 15.3 which is also the total utility increase of the two periods. The principal, on the other hand, increased his utility by 67 in t_2, and increased his total utility by 99. The social total utility increase is 114.3. Comparing the two cases, we can see that both the agent and the principal are better off in case 2. The optimal work effort in the one-period model is no longer optimal in the two-period model.

Formally, assume there are T periods start from 1 to T during these periods agent's employment is guaranteed. Work effort in period j (j=1,2,.....T) is defined as x_j (the subscript i is dropped to avoid cluttering). In period j, ϕ_j percent of the EV_m in period j-1, $EV_{m,j-1}$, is re-invested in the same team. From (3-4) we know that $EV_{m,j-1}$ is a function of x_{j-1} and $EV_{m,j}'(x_j)>0$. Assume that the P value in j, P_j is a positive function of investment in j, I_j, and all previous investment. We have

$$I_j=\phi_j EV_{m,j-1}(x_{j-1},x_{j-2},\ldots\ldots,x_1) \qquad 0<\phi_j<1$$

$$P_j=P(I_j,I_{j-1},\ldots.I_0) \quad I_j'(EV_{m,j-1})=\phi_j>0 \quad P_j'(I)>0 \qquad (3\text{-}31)$$

$$j=1,2,\ldots\ldots\ldots\ldots.T$$

where I_j: the part of $EV_{m,j-1}$ which is re-invested in the same team.

P_j: the P value in period j which is previously assumed independent of work effort in the one-period model, but now is a function of all previous work effort.

It is easy to show that work effort in period j only affect the P values after j

$$\frac{\partial P_k}{\partial x_j}=0 \qquad \textit{for } k\le j$$

$$\frac{\partial P_k}{\partial x_j}=\sum_{i=0}^{k-j-1}\frac{\partial P_k}{\partial I_{k-i}}\phi_{k-i}\frac{\partial EV_{m,k-i-1}}{\partial x_j}>0 \quad \textit{for } k>j \qquad (3\text{-}32)$$

The optimal work effort in period j is generated by maximize the present value of a future utility stream, PVEV

$$\max_{x_j} PVEV = \sum_{j=1}^{T} \frac{1}{(1+\theta)^j} [w_0 + sP_j(x_j - x_0) - x_j^2] \qquad (3\text{-}33)$$

$$S.T. \quad (3\text{-}31) \quad (3\text{-}32)$$

$$F.O.C \Rightarrow x_{j.l}^* = \frac{sP_j + \sum_{k=j+1}^{T} (1+\theta)^{j-k} [s \frac{\partial P_k}{\partial x_j} (x_k - x_0)]}{2} \qquad (3\text{-}34)$$

where $x_{j.l}^*$ is the long-run optimal work effort in period j, and θ is the discount rate which reflect agents' time preference. From (3-34) we can see that as θ increases or the ϕ values decreases, the second term in the numerator decreases. If θ increase to infinity or all $\phi_k=0$ (k=j+1, j+2, ... T), the second term reduce to zero, and the long-run optimal efforts reduce to the one-period optimal effort, $x_j^*=sP_j/2$. Otherwise, the second term in the numerator is positive, and $x_{j,l}^*>x_j^*$ for all j<T; when j=T, $x_{T,l}^*=x_T^*=sP_T/2$.

Proposition 7: if workers expect the profit gained by the principal will be reinvested in the same team so that their productivity will be increased, and if they expect long-term employment, the long-run optimal effort will be greater than the one-period optimal effort; and $x_{j.l}^*$ is positively related to the value of ϕ.

Reinvest profit to the same team, of course, may not be good for the society as well as the principal, so if the principal can invest this money on other projects and earn

higher return, he will not care too much to increase workers effort in the long run, and the ϕ value will be low or even equal to zero, and the long-run and one-period optimal efforts will be equal. Only if the original team warrants extra investment, is there a positive ϕ value.

This model points out several directions to increase work effort:

a) If the value of ϕ is increased, $x_{j,l}$* increases. To increase ϕ, the principal's investment should be closely related to EV_m. The practice that firms turn over all the profit share of the principal to the state and investment is independently determined by the government in the Chinese state enterprises leads to zero ϕ, so that $x_{j,l}$*=x_j*. Very often the government official have strong incentive to control investment so that the ϕ value is small.

b) Even though workers may expect life-time employment, they may not expect the PS system to go on forever. Unless workers expect the PS system to last "long enough," they may not be willing to maximize PVEV.

c) If workers expect widespread official corruption or large nonproductive investment, so that the ϕ value is small, $x_{j,l}$* will be smaller.

In the above analysis, we assumed that reinvestment is used to increase the productivity of the same team. What if new workers is hired following this reinvestment? It is possible that the reinvestment is used in new projects so that

the P value of the original team is unaffected. In this case, the ϕ value is in fact zero. however, there are still two reasons to expect that the workers may benefit from the reinvestment. First, if the new workers hired are going to be the children or relatives of the original workers and the original workers are maximizing family utility (Byrd and Tidrick, 1987), they would also benefit from the reinvestment. Secondly, if the new employment quota is fixed by the government to ensure full employment policy, even though there is no new investment, the same new workers have to be hired any way, which means that the P value will decrease, in this case the original workers will benefit from the re-investment.

Now return to the two-period hypothetical example. Using (3-34) the long-run optimal efforts are 7.853 in t_1 and 7.854 in t_2, higher than the one-period optimal efforts (x_1*=6, x_2*=6.96). The total utility increase of the agent is 15.46 which is greater than that in case 1 (12.76) and case 2 (15.3). The principal's total utility increase is 95.6 which is less than that of case 2 (99). And the social utility increase (111.42) is also lower than that of case 2 (114.3). This implies that the long-run optimal work effort is not in general the socially optimal effort. In fact, the socially optimal work effort in t_1 is 10.

2) There is another reason to doubt the conclusion that it is optimal to supply x_i below x_{imax} when x_i*<x_{imax}. As discussed above, x_i* is only the optimal work effort of agent i, it is

not the socially optimal work effort. The social optimal effort, $x_{is}*$ can be derived by setting the marginal social utility equal to the marginal social disutility, or for an arbitrarily small effort, e, we have

$$Pe=(x_i+e)^2-x_i^2=2ex_i+e^2$$
$$x_{is}*=\frac{P}{2} \qquad when \quad e\to 0 \tag{3-35}$$

If we set s=1 in the PS system, then $x_i*=sP/2=P/2=x_{is}*$. The principal, of course, is unwilling to accept s=1, however, the principal may accept a new payment system which sets s=1 at the margin and the optimal effort equals the socially optimal effort. I return to this topic in the analysis of the LST system.

Proposition 8: under the PS system, the socially optimal s is one.

3.2.5. Incentive Compatibility Analysis

From (3-14) we can see that in order to determine the optimal s endogenously, the principal has to know the value of P and x_{i0}, but in fact very often the principal can only observe the value of PX or Px_i. Only the agents know the true value of P and x_{i0}. Now the question is that if the principal has no other ways to determine the value of P and x_{i0} from Px_{i0}, can he simply let the agents report the true value of P and x_{i0}. In other words, if workers know that the principal will set s=s*, is there any incentive for the workers to misrepresent the true value of P by reducing their effort

below the optimal effort, x_i*? Here we assume there is no free-riding behavior. From (3-21) we can see that as s increases, workers are better off, so workers do have an incentive to convince the principal that the P value is smaller than the true P value, since $s^*=1/2 + x_{i0}/P$ when $P<P_{max}$, and $s^*=2x_{imax}/P$ when $P>P_{max}$. If the principal believes that the P value is smaller than its true value, a higher s is chosen, and workers are better off. The optimal effort is, however, also positively related to the P value (3-13). If the principal believes that the P value is smaller than its true value, he will expect a lower effort than the optimal one. Since the principal can observe the value of Px_i, in order to convince the principal ex post that the P value is lower than the true P value, workers have to reduce their effort level below the optimal one. If $x_i<x_i^*$, workers may become worse off. The question is whether it's worth it to misrepresent the P value. Let's look at the case when $P<P_{max}$ first. In this case, $s^*=1/2 + x_{i0}/P$, if workers know the value of P and x_{i0}, but the principal can only observe Px_{i0}, the workers tell the principal that the P value is $P_c=qP$ ($q<1$), and x_{i0c} is x_{i0}/q, such that $P_cx_{i0c}=Px_{i0}$, if the principal has no other information about P and x_{i0}, and he believed the workers, then the principal will set $s=s_e$ and expect the optimal effort x_i*

$$s_e=\frac{1}{2}+\frac{x_{i0c}}{P_c}=\frac{1}{2}+\frac{x_{i0}}{q^2P}>s*$$
$$x_{ie}=\frac{P_c s_e}{2}=\frac{qP}{4}+\frac{x_{i0}}{2q} \qquad (3\text{-}36)$$
$$P_c x_{ie}=(\frac{qP}{4}+\frac{x_{i0}}{2q})\,qP=\frac{q^2P^2}{4}+\frac{Px_{i0}}{2}$$

To confirm the principal's expectation, workers have to offer an effort level x_{ic}, such that $Px_{ic}=P_c x_{ie}$

$$x_{ic}=\frac{P_c x_{ie}}{P}=(\frac{q^2P}{4}+\frac{x_{i0}}{2})<x_i* \qquad (3\text{-}37)$$

where s_e: The optimal s when estimated P is P_c

x_{ie}: The expected optimal effort when estimated P is P_c

$P_c x_{ie}$: The expected value of Px_i*

x_{ic}: The effort level which will confirm the principal's expectation

The net gain of agent i from misrepresenting the P value, NGC, is

$$\begin{aligned} NGC_i &= EV_i(x_{ic},s_e)-EV_i(x_i*,s*) \\ &= s_e P(x_{ic}-x_{i0})-x_{ic}^2-s*P(x_i*-x_{i0})+x_i*^2 \\ &= \frac{P^2}{8}(q^2-\frac{1}{2}-\frac{q^4}{2})+\frac{x_{i0}}{2}(q^2-1)(\frac{x_{i0}}{q^2}-\frac{P}{2}) \end{aligned} \qquad (3\text{-}38)$$

$$for \quad q<q_c=\sqrt{\frac{2x_{i0}}{P}} \qquad NGC_i<0$$

The first term of NGC_i is negative, the second term is negative if $q<q_c$. If the P value is close to P_{min}, then q_c is close to 1. Workers are worse off if they set $q<q_c$, so they are unlikely misrepresent the P value, at least they may not

set $q<q_c$. Since q_c is close to one, there is very little room to set $1>q>q_c$, and even they can set $1>q>q_c$, NGC_i is not necessarily positive, since the first term of NGC_i is negative. If the P value is close to P_{max}, we have:

1) If x_{i0} is very high, q_c is large, the second term may be positive for high q value, but if q is high, (q^2-1) is low, the second term will be small in absolute value, so even if workers may gain by misrepresent the P value a little bit (if they report a very low P value so that $q<q_c$, they will be worse off), the potential gain is very small.

2) If x_{i0} is very small, so q_c is low, choosing a low value of q ($q>q_c$), the second term will be positive, but the first term will be more negative, which will offset the positive second term. So in general, the incentive to misrepresent the P value when $P<P_{max}$ is very weak. If workers misrepresent the P value so that the principal sets $s=s_e>s^*$, then workers offer effort at $x_i=Ps_e/2$ instead of x_{ic}, workers will gain at the first period, at the same time, the true value of P is revealed, so the principal will change s_e to s^* in the second period. In the real Chinese situation, the potential productivity (production quota) is determined through negotiations between the principal and the agents. Since bargaining is costly in general, agents are unlikely misrepresent the P value through exhaustive bargaining, and supply suboptimal work effort to confirm the principal's expectation.

Proposition 9: if the P value is not very high, and the principal can adjust s period by period according to the observed value of Px_i, then the incentive of the agents to continuously misrepresent the P value by supply suboptimal work effort is very weak. They would not reduce their effort to confirm the principal's expectation of the P value ex post.

Secondly let's look at the case when $P>P_{max}$ (In fact, $P>>P_{max}$, since $P_c=qP$ must be still greater than P_{max}. If P is close to P_{max}, the room to reduce q is very small). In this case we have $s^*=2x_{imax}/P$. In order to convince the principal that the P value is in fact $P_c=qP$ ($q<1$), the effort level $x_{ic}=P_c x_{imax}/P=qx_{imax}$ has to be offered. If the principal is convinced that the true P value is P_c, he will set $s=s_e=2x_{imax}/P_c=s^*/q>s^*$.

The net gain to misrepresenting the P value is

$$\begin{aligned} NGC_i' &= EV_i(x_{ic}, s_e) - EV_i(x_{imax}, s^*) \\ &= x_{imax}[2x_{i0}(1-\frac{1}{q}) + x_{imax}(1-q^2)] \\ &\quad if\ \ x_{i0}\to 0 \quad then \quad NGC_i>0 \\ &\quad if\ \ x_{i0}\to x_{imax} \quad then \quad NGC_i<0 \end{aligned} \tag{3-39}$$

So if x_{i0} is small the incentive to misrepresent the P value is strong. For example, if $x_{i0}=2$, $x_{imax}=8$, and $q=0.8$, then $NGC_i'=15$. If workers do not reduce their effort to confirm the principal's expectation, they will gain more in the first period but will reveal the true P value. In the following periods the principal reduces s to s*.

The net gain in the first period for the above example is

$$EV_i(x_{imax}, s_e) - EV_i(x_{ic}, s_e) = 8.96$$

So workers will prefer to reduce their effort to confirm the principal's expectation and keep gain NGC_i in the following periods.

If the principal has no other information sources about the P value and/or x_{i0} value; if workers expect firm will set s=s*; and if x_{i0} is small relative to x_{imax}, then workers will have a strong incentive to misrepresent the true value of P by supplying suboptimal work effort. Of course, if the principal set s exogenously and independent of the P value, workers have no incentive to reduce their effort level. If workers are convinced that s is independent of their effort level, they will not supply suboptimal work effort, and the principal may use the observed value of Px_i to estimate the true P value. Since $x_i = sP/2$ or $Px_i = sP^2/2$, and Px_i is observed, we can calculate P as

$$P = \sqrt{\frac{2Px_i}{s}}$$

If the principal also has other information sources about the P value, he can use those information sources and set s=s* endogenously.

Proposition 10: If the P value is very high ($P >> P_{max}$) and unknown to the principal, and x_{i0} is small relative to x_{imax}, workers may supply suboptimal work effort to convince the principal that the P value is lower than the true P value, so

that a higher s might be offered. In this case, setting s exogenously will remove the workers' incentive to reduce their work effort.

If s is fixed for several periods, then workers may prefer misrepresent P value but not reduce work effort. In this case, even though the true P value is revealed in the second period, s cannot be changed for several periods. The effect of the once-and-for-all misrepresentation is much higher. And the principal might be worse off if he set s* according to the reported P and x_{i0}. In this case exogenous s might be selected.

So, workers always have incentive to misrepresent the P value, but they may or may not reduce work effort to confirm the principal's expectation. If s can be adjusted each period and workers have no incentive supplying suboptimal work effort to continuously misrepresent the P value, the principal can use the reported P value to calculate s* in the first period; if the agents have incentive reducing work effort to continuously misrepresent the p value continuously, or s is fixed for several periods, the principal may not use reported P to calculate the optimal s. He might prefer exogenous s.

3.2.6. Exogenous s and Potential Inefficiency

From above analysis, we can see that only under very rare situations should s be determined exogenously, fixed by government regulations for example. For one thing, P value

must be greater than P_{max}, but $P>P_{max}$ is inefficient, because in this case $x_i=x_{imax}$, new workers should be hired. Secondly, x_{i0} must be very low. And even under these rare situations, if there are many identical teams, workers may not reduce their work effort to confirm the principal's expectation. As long as workers in one team do not reduce their effort and reveal the true P value, the principal may change s for all teams in the next period. So if workers across teams are not fully cooperative, workers may misrepresent the P value once-and-for-all, but may not reduce their work effort. It seems that in most cases the principal should set s endogenously using the P value and x_{i0} reported by workers. In the real world, however, s is very often exogenously given to the firm, because of the following reasons:

1) The principal may not know which s is optimal. The optimal s in (3-14) is dependent on the assumption of the disutility of effort. Under different assumptions of the disutility of effort, the form of s* will change. So there is no uniform and simple equations to calculate s* (appendix 2).

2) The assumption that the principal can observe Px_i may not be hold, especially when the principal is a government organ, and the agent is a firm. Under the real profit retainment system of the Chinese state enterprises, the state government as the principal signs a contract with the enterprise in terms of s and the base of profits turnover (PX_0) which is very often the profit level of last year. In

this type of contract, the government is the principal and the firm represented by its manager is the agent. In the second level, the manager of the enterprise signs similar contracts with the shop leaders. Similar contracts might be signed all the way to individual workers if possible. The manager in this case becomes the principal and workers become agents. If the agents manage to misreport the value of PX by reporting some welfare expenditures as production costs (as practiced by many Chinese state enterprises), the principal loses the ability to check the truthfulness of the reported P value ex post. Now the agent can always misrepresent P without reducing their effort and still confirm the principal's expectation by simply reducing the reported value of PX. Even though work effort is not reduced by misrepresenting the P value, the principal becomes worse off. So the principal is unwilling to use the P value reported by the agents to calculate the optimal s. And if he has no other reliable information sources to calculate the P value, he is forced to select a fixed s.

If s is exogenous and in general unequal to s*, what is the potential loss to the principal? If the principal sets s below s*, social welfare is reduced since the socially optimal s is 1; workers are worse off since EV_i increases with s; the principal is worse off also since EV_m is maximized when s=s*. If the principal sets s above s*, social welfare is increased; workers are better off; but the principal is worse off. Since

s is determined by the principal, the principal will compare the losses under the two situations. In the following analysis, we assume that the principal knows the optimal s, but does not know the true P value. If the true value is P, but the estimated P value is P_e=P-dP, and P_e'=P+dP we have:

1) $P>P_{max}$. In this case, the conditional optimal effort level, x_i', and the expected x_{i0}, x_{i0e}, are

$$x_i'=\frac{P}{P_e}x_{imax}>x_{imax} \quad so \quad x_i=x_{imax}$$
$$x_{i0e}=\frac{x_{i0}P}{P_e} \quad s_e=\frac{2x_{imax}}{P_e}>s* \quad P_e x_{i0e}=Px_{i0} \tag{3-40}$$

The potential loss due to a larger than optimal s, PLL, is

$$PLL=EV_m(s*,x_{imax})-EV_m(s_e,x_i')$$
$$=2x_{imax}\frac{dP}{(P-dP)}(x_{imax}-x_{i0}) \tag{3-41}$$

If the estimated P value is P_e'=P+dP, the optimal effort level, x_i', and the estimated x_{i0}, x_{i0e}, are

$$x_i'=\frac{P}{P_e'}x_{imax}<x_{imax} \quad x_{i0e}'=\frac{x_{i0}P}{P_e'} \quad s_e'=\frac{2x_{imax}}{P_e'}<s* \tag{3-42}$$

The potential loss from a smaller than optimal s, PLS, is

$$PLS=EV_m(s*,x_{imax})-EV_m(s_e',x_i')$$
$$=x_{imax}\frac{dP}{P+dP}(P+2x_{i0}-2x_{imax}\frac{2P+dP}{P+dP}) \tag{3-43}$$
$$\approx x_{imax}\frac{dP}{P+dP}(P+2x_{i0}-4x_{imax})>0$$

Comparing PLL with PLS we see that if x_{i0} is not too close to x_{imax}, and P is not too high above P_{max}, we have PLL>PLS. The principal would rather overestimate P so that too small an s is offered than underestimate P and offer too high an s.

2) $P<P_{max}$. If the estimated P is $P_e=P-dP$, then we have

$$x_i'=\frac{Ps_e}{2}=\frac{P}{4}+\frac{x_{i0}P^2}{2P_e^2}>x_i* \quad x_{i0e}=\frac{x_{i0}P}{P_e} \quad s_e=\frac{1}{2}+\frac{Px_{i0}}{P_e^2}>s* \quad (3\text{-}44)$$

$$PLL=EV_m(s*,x_i*)-EV_m(s_e,x_i')=\frac{x_{i0}}{2}[1+\frac{P^2}{P_e^2}(\frac{P^2}{P_e^2}-2)] \quad (3\text{-}45)$$

Since $P/P_e>1$, PLL>0. As dP increases, PLL increases.

If the estimated P is $p_e'=P+dP$, then

$$x_i'=\frac{Ps_e'}{2}=\frac{P}{4}+\frac{x_{i0}P^2}{2P_e'^2}<x_i* \quad x_{i0e}=\frac{x_{i0}P}{P_e'} \quad s_e'=\frac{1}{2}+\frac{Px_{i0}}{P_e'^2}<s* \quad (3\text{-}46)$$

$$PLS=EV_m(s*,x_i*)-EV_m(s_e',x_i')=\frac{x_{i0}}{2}[1+\frac{P^2}{P_e'^2}(\frac{P^2}{P_e'^2}-2)] \quad (3\text{-}47)$$

Since $P/P_e'<1$, PLS>0, as dP increases, PLS increases too. In this case we still have PLL>PLS, for example, if P=20, dP=5, then $PLL=0.6x_{i0}^2/2$, $PLS=0.13x_{i0}^2/2$. So in general, we can conclude that it is very likely that PLL>PLS.

Proposition 11: If the principal has to set s exogenously, he prefers to offer a lower than optimal s rather than a higher than optimal s.

From proposition 1 and proposition 8 we know that work effort is positively related to s and the social optimal s is

1. The principal, however, prefers to select a lower s, which is inefficient. One way out of this track is to strengthen the audit of the PX value. If agents have no way to misreport the PX value, the incentive to misrepresent the P value no longer exists in most cases, and s could and should be determined endogenously.

3.2.7. Trade off Between Efficiency and Equality

The PS system has a special feature, when $P_{min}<P<P_{max}$ (the interior solution), EV_i increases with P (3-19). Even though teams with a higher advantage in P are penalized with a lower s*, which counteracts the effect of unequal endowments and leads to greater equality, changes in s* cannot fully offset the effect of unequal endowments, because even set s=s*, the agent with higher P is still better off relative to the agent with lower P. This result can be seen from (3-19). However, P is very often independent of work effort of agents. The teams which have higher capital-labor ratio, or which benefitted from the price distortion are better off. So we have an equality problem: if s=s*, teams with lower P are relatively worse off. The principal when dealing with multiple-teams, faces a trade-off between efficiency and equality. In this system, w_0 is fixed, x_{i0} is fixed, the only control variable is s, if s is unequal to s*, the principal is worse off. In practice, the principal often tries to equalize EV_i among teams, and adjust s accordingly, for teams with high

P value, $s<s^*$, and efficiency is lost. If the principal sets $s=s^*$ and ignore the equality problem, workers in relative disadvantage situations may feel unfairly treated, according to the efficiency wage theory, workers will take actions to punish the principal, which also lead to losses.

Proposition 12: Under the PS system, if the P value is not very high, the expected utility of agents is positively related to the P value. With unequal endowment of P, to keep horizontal equality, efficiency has to be sacrificed.

In fact, within an enterprise, inter-team relationships are often very important. Unless all teams increase work effort simultaneously, enterprise performance may not be improved significantly. In enterprises where each team is independent of other teams, horizontal equality is not very important. However, in enterprises where a few teams could become bottle-necks, horizontal equality is very important, and efficiency has to be sacrificed in the PS system. This is a drawback of the PS system. Another problem is that the optimal work effort x_i^* is not socially optimal. To solve these problems, I now turn to some other incentive payment systems.

3.3. The Disutility of Effort Compensation System

In the simple PS model as defined in the beginning of section 3.2, when $P<P_{max}$, x_i* is smaller than x_{imax}, and x_i* is not socially optimal, it is even not long-run optimal. For example, if $x_{i0}=2$, $x_{imax}=8$, then $P_{max}=28$ (3-12). If P=16, then $x_i*=P/4+x_{i0}/2=5$ (3-10), and the socially optimal effort is $x_{is}*=P/2=8$ (3-35). The PS system also has an equality problem, since EV_i increases with P. To keep EV_i values equal across teams, efficiency has to be sacrificed. In this section I introduce an alternative payment system, the disutility of effort compensation (DEC) system. The ideal DEC system is defined as follows: If agents supply x_{i0}, their payment would be w_0 as in the FW system; if they supply work effort above x_{i0}, but not greater than the principal desired value of x_i, x_{id}, their payment would be w_0 plus an incentive wage which fully compensate the disutility of extra work effort. The incentive wage is scheduled such that the expected utility of agent i is highest when he supplying x_{id}. For $x_i>x_{id}$, the incentive wage do not increase accordingly.

3.3.1. Advantages of the DEC System Over the PS System

The ideal DEC system can be modeled as

$$EV_i=w_0+c(x_i^2-x_{i0}^2)-x_i^2 \quad \text{for } x_i\leq x_{id} \tag{3-48}$$

$$EV_i=w_0+c(x_{id}^2-x_{i0}^2)-x_i^2 \quad \text{for } x_i>x_{id}$$

$$EV_m=n[Px_{id}-w_0-c(x_{id}^2-x_{i0}^2)] \tag{3-49}$$

where c is a constant, and c is greater than and close to 1; x_{id} is the principal desired value of x_i.

In this system, if agents only offer x_{i0}, the utilities of both parties are the same as in the FW system; if agents increase effort level to $x_{i0}<x_i\leq x_{id}$, the extra disutility of effort is $(x_i^2-x_{i0}^2)$ and the extra utility from the incentive payment is $c(x_i^2-x_{i0}^2)$ which is greater than the extra disutility of effort. The optimal effort is always equal to x_{id} (figure 3). The principal maximizes profit by setting c close to one; if c equals to one, agents are indifferent between all effort levels. By setting c>1, expected utility of agent i is maximized at x_{id}. The optimal x_{id}, $x_{id}*$, is determined by maximizing EV_m with respect to x_{id}. It's easy to verify that $x_{id}*=P/2c\approx P/2=x_{is}$ (the socially optimal work effort), since c is close to 1. Strictly speaking, there is also a trade-off between efficiency and equality; setting $x_{id}=x_{id}*$ is efficient, and the partial derivative of EV_i with respect to P is $(c-1)P/(2c^2)$, it's positive and increases with P. However, since c is close to one, the partial derivative is close to zero, the difference between utilities caused by unequal endowment in P is very small. Approximately, the equality problem in the PS system is solved.

Proposition 13: The advantages of the ideal DEC system over the PS system are: i) there is no serious horizontal equality problem in the DEC system; ii) the optimal effort in the DEC system equals to the principal desired work effort

which is very close to the socially optimal work effort. If we relax the assumption that agents are identical in terms of their disutility of work effort, the optimal work effort in the DEC system is not socially optimal in general (see appendix 4).

3.3.2. Disadvantages of the DEC System Over the PS System

Even though the ideal DEC system has some advantages over the PS system, it also has some disadvantages.

1) The most serious disadvantage of the DEC system is its dependence on the measurability of the effort level. In the PS system, the principal only controls s, even though the values of x_{i0} and P are not known by the principal, he can always set s exogenously, and work effort will be increased relative to the FW system. If the principal can observe the value of Px_{i0}, in most cases, he can simply let the agents report the value of P and x_{i0} (proposition 9), and set s endogenously. This improves the principal's utility and very likely improves the social welfare as well. In the DEC system, however, the principal has to know the value of x_i and x_{i0} before he can make the incentive payment. If effort level can be easily observed with only small errors, there will be no problem. Unfortunately, in many cases, effort levels cannot be observed accurately and/or without high costs. And if x_i can only be calculated from the observed value of Px_i, and there is no independent information about either the P

value or the effort level, the incentive of the agents to misrepresent the P value is very strong. If the principal can only observe the value of Px_i, or Px_{i0}, and asks the agents to report the P value and/or x_{i0}, the agents will report qP (q<1) and x_{i0}/q instead of P and x_{i0} and supply an effort level $x_{ic}=qx_{id}$, since $Px_{ic}=Pqx_{id}$, just confirm the principal's estimation of P and x_i*. The expected utility of agent i if he misrepresents P, EV_{ic}, and the expected utility if he does not misrepresent P, EV_i, are

$$EV_{ic}=w_0+c(x_{id}^2-\frac{x_{i0}^2}{q^2})-q^2x_{id}^2$$

$$EV_i=w_0+c(x_{id}^2-x_{i0}^2)-x_{id}^2 \qquad (3\text{-}50)$$

$$EV_{ic}>EV_i \qquad for \quad q>\frac{x_{i0}\sqrt{c}}{x_{id}}$$

So if x_{i0} is not too close to x_{id}, it is easy to find a q<1 so that agents are better off by misrepresenting the P value, and agents reduce their effort level below the optimal level. So in the DEC system, the principal cannot rely on the reported P or x_{i0} to calculate incentive payments.

The principal also cannot depend on co-workers to reveal the true effort level, as co-workers have no incentive to check the shirker. The cost of shirking is borne by the principal in the DEC system.

2) Another disadvantage of the DEC system is the weak relationship between the P value and agents' incentive payments. In the PS system,. for $P<P_{max}$, P value increase will

benefit both the principal and the agents. Agents also have incentives for innovations, technology improvement, and investment. In the DEC system, however, either the P value increase only benefits the principal so that workers have no incentive to increase the P value, or the agents reduce their work effort after the P value is increased. If the principal can observe effort levels or he has good knowledge of the P value, the utility of the agents in the DEC system is independent of the P value; if the principal has no knowledge of the P value and cannot observe effort levels, agents have incentive to increase the P value. They reduce work effort after the P value is increased. Suppose there is an innovation. After the innovation is adopted, the P value doubles. If the principal knows that the P value doubles after adopting this innovation, he expects the Px_i value, which is observable, also to double. The incentive payment does not change since work effort does not changed after the firm adopts this innovation. Only the principal benefits from adopting this innovation. Agents have no incentive to adopt this innovation. If, on the other hand, the principal does not know the P value, he can only guess the P value from Px_i. Agents reduce their work effort to convince the principal that the P value has increased less than two times after adopting this innovation. If the Px_i value has in fact increased 1.5 times, the principal cannot tell the difference between a) the P value has only increased by 1.5 times instead of 2 times,

and b) the P value has increased 2 times but the agents reduced their work effort accordingly. So the principal cannot reduce workers' incentive payment, especially in the Chinese state enterprises where any decrease of the incentive payment must be supported by black and white evidence that work effort has been reduced because of the reasons mentioned in chapter 2. Workers benefit from the increase in the P value because their work effort is lower for the same money payment. In this case workers have incentive to adopt this innovation and reduce their work effort afterwards. Either P does not increase or as P increases, x_i is reduced. Both results are inefficient from society's point of view.

3) Even if the mean value of P is known by the principal, the P value is certainly subject to random disturbances and new investment or technological improvements changes the P value. Thus the principal and agents have to continuously bargain about the true value of P. With the strong incentives of agents and the weak incentives of principals in the Chinese situation, the principal may accept a lower purported P value than the true P value, so that actual effort is smaller than the optimal effort. In the real Chinese situation, production quotas are very often used as the proxy of x_{id}, bonus is paid as long as the production quota is fulfilled. Production quotas are determined through negotiations between the principal and agents. In the DEC system, risks caused by uncertainty, such as power breaks and material shortages are

borne by the firm. When negotiating the production quota, the effects of uncertainty are very often exaggerated, as a result, production quotas are very low in general (Cauley and Sandler, 1991), and workers have no incentive to overfill production quotas.

Proposition 14: The disadvantages of the DEC system over the PS system are: a) the incentive to misrepresent the P value and reduce work effort below the optimal level is very strong regardless of the P value in the DEC system; b) agents in the DEC system either have no incentive to increase the P value or reduce work effort after the P value is increased; and c) in the DEC system, there is more rent-seeking activities which causes social losses.

3.4. The Extended Profit Sharing System

The PS system has an equality problem which is not serious in the DEC system, and the optimal effort in the PS system is unequal to the socially optimal effort. The DEC system also has its own problems. In this section, I introduce another payment system, the extended PS system. In the PS system, the base of profit for share between the principal and agents is $Px_i - Px_{i0}$. If the true values of P and x_{i0} are not known, the base for share is not affected, as Px_{i0} and Px_i can be observed by the principal according to our assumption. If $P > P_{min}$, changing from the FW system into the PS system causes work effort to increase, and both parties are

better off (proposition 4). In the extended profit sharing (EPS) system, the base of profit for share is not tied to $Px_i - Px_{i0}$; instead the system sets the base for share as $P(x_i - x_{i0e})$, where x_{i0e} is the principal's estimated value for x_{i0}, and in general x_{i0e} is unequal to x_{i0}. This is the only difference between the EPS system and the PS system; but it is an important difference. The principal can deliberately overestimate x_{i0} so that agents' payments are reduced. The conditional optimal effort under this system looks just the same as in the PS system, but the unconditional optimal effort in the EPS system is different from that of the PS system because the optimal s is different. The optimal s and unconditional optimal effort in the EPS system are closer to the socially optimal values. And the horizontal equality problem can be solved in the EPS system, because the principal has the ability to adjust the potential gains of the agent. These are the advantages of the EPS system over the PS system; the EPS system does not have the disadvantages of the DEC system. The disadvantage of the EPS system over the PS system is that the risk to reduce work effort to x_{i0} might be higher.

3.4.1. Optimal Effort, Ratio of Share, and Potential Gains

The EPS system can be modeled similarly as the PS system. With x_{i0} in (3-3) and (3-4) replaced by x_{i0e}. So we have:

$$EV_i = w_0 + sP(x_i - x_{i0e}) - x_i^2 \qquad (3\text{-}51)$$

$$EV_m = (1-s)nPx_i - n(w_0 - sPx_{i0e}) \qquad (3\text{-}52)$$

where Px_{i0e} is selected by the principal and is not equal to Px_{i0} in general.

The conditional optimal effort, in case of interior solution, can be developed as follows

$$\begin{aligned} &\max_{x_i} EV_i = w_0 + sPx_i - sPx_{i0e} - x_i^2 \\ &S.T.\ EV_i(x_i') > EV_i(x_{i0}) \\ &F.O.C. \Rightarrow x_i' = \frac{sP}{2} \end{aligned} \tag{3-53}$$

where x_i' is the conditional optimal effort under exogenous s.

The form of the conditional optimal effort is the same as in the PS system. The potential gains of the principal and agents are, however, different from that of the PS system. From (3-51) and (3-52) we can see that as x_{i0e} increases, the principal becomes better off, and agents become worse off. However, as x_{i0e} increases, it's more likely that the condition that $EV_i(x_i')>EV_i(x_{i0})$ does not hold. In this case workers reduce their effort to x_{i0}. For example, if $x_{i0}=2$, $x_{i0e}=6$, $P=20$, $s=0.7$, then $x_i'=Ps/2=7$. In the PS system, $EV_i(s,x_i')-EV_i(x_{i0})=Ps(7-2)-49+4=25>0$; in the EPS system, $EV_i(s,x_i',x_{i0e})-EV_i(x_{i0})=Ps(7-6)-49+4=-31<0$, so workers offer x_{i0}. In general, there is a critical value of x_{i0e}. Above this value, workers prefer to x_{i0}, and both the principal and agents are worse off. This critical value, defined as $x_{i0e}*$, can be developed as follows

$$\text{Set } EV_i(s,x_i')-EV_i(x_{i0})=Ps(x_i'-x_{i0e})-x_i'^2+x_{i0}^2=0$$

$$\text{Solve for } x_{i0e}*=Ps/4+x_{i0}^2/(Ps) \tag{3-54}$$

As P increases, Ps increases. For reasonable values of x_{i0} and Ps, x_{i0e}* also increases. As x_{i0} increases, x_{i0e}* is closer to x_{i0}. So if x_{i0} is low, and the P value is high, the principal can set x_{i0e}*>x_{i0e}>x_{i0}, EV_m will increase, and horizontal equity can be realized for teams with different P values. However, if x_{i0} is high, there is very little room for the principal to set x_{i0e}*>x_{i0e}>x_{i0}.

Proposition 15: In the EPS system, the principal has the ability to realize horizontal equality and extract some or even all of the potential gains of the agents by adjusting the profit base for share, especially when the minimum work effort is low and the P value is high.

If the principal sets x_{i0e} equal to x_{i0e}*, then the optimal s is one, which is the socially optimal s. We can develop this result as follows

$$\max_s EV_m = (1-s)\,Pnx_i - nw_0 + Pnsx_{ioe}$$

$$S.T. \quad x_i = x_i' = \frac{Ps}{2} \qquad x_{i0e} = x_{i0e}* = \frac{Ps}{4} + \frac{x_{i0}^2}{Ps} \tag{3-55}$$

$$F.O.C. \quad \Rightarrow \quad s*=1$$

The potential gain of the principal under the EPS system, PGPE, over the fixed wage system is

$$\begin{aligned} PGPE &= EV_m(x_i*, x_{i0e}*, S*) - EV_m(x_{i0}) \\ &= n\left(\frac{P^2}{4} + x_{i0}^2 - Px_{i0}\right) = 2PGP \end{aligned} \tag{3-56}$$

where PGP is the potential gain of the principal under the PS system (see 3-15).

The unconditional optimal effort and x_{i0e}* under endogenous s are

$$x_i^*=Ps^*/2=P/2 \quad \text{and} \quad x_{i0e}^*=P/4 + x_{i0}^2/P \qquad (3\text{-}57)$$

We can see that the optimal effort in this case is the socially optimal effort. And the principal will claim all potential gains relative to the FW system. The underlying mechanisms of the EPS system and the PS system can be compared using figure 4. The expected utility of the agent is composed of three items: i) sPx_i which is linearly related to x_i. For given P, the slope of sPx_i is positively related to s. ii) x_i^2, the disutility of work effort. iii) w_0-sPx_{i0} (in the PS system) or w_0-sPx_{i0e} (in the EPS system) which is independent of work effort, and it can only shift EV_i vertically. In the PS system, w_0-sPx_{i0} is not a control variable; while in the EPS system, w_0-sPx_{i0e} is a control variable. The principal can control x_{i0e} to affect EV_i. The optimal work effort is only dependent on s (for given P) in both systems, as long as x_{i0e} is not too high such that $EV_i(x_i')<EV_i(x_{i0})$. If $x_{i0e}>x_{i0}$, the agent is worse off, and the principal is better off in the EPS system, since the value of $sP(x_{i0e}-x_{i0})$ is transferred from the agent to the principal. If $x_{i0e}=x_{i0e}$*, then the principal extracts all potential gains relative to the FW system. Obviously, the principal prefers a higher work effort, which can be realized by setting a higher s, so at the margin, s=1.

Proposition 16: In the EPS system, the conditional optimal effort is the same as in the PS system, but the

unconditional optimal effort and optimal s are closer to the socially optimal values than that of the PS system. If agents are risk neutral or there is no uncertainty, at the extreme, the unconditional optimal effort and optimal s are the socially optimal values.

3.4.2. Risk and Uncertainty

From the above analysis, we can see that if workers are risk neutral or there is no uncertainty, the principal can always set x_{i0e} a little bit below x_{i0e}*, such that workers prefer to offer x_i* rather than x_{i0}. The optimal s and optimal work effort will be very close to the social optimal values, and the principal can extract more potential gains. All these are the advantages of the EPS system. These advantages, however, cannot be realized without cost which is the risk that the agents may reduce their effort all the way to x_{i0} and invalidate the incentive payment system completely. In fact, if we assume agents are risk averse, they will maximize certainty equivalent utility as follows

$$\max_{x_i} U_i(x_i, w_0, s) = w_0 + sP(x_i - x_{i0e}) - x_i^2 - \frac{\alpha s^2 \sigma^2}{2} \qquad (3\text{-}58)$$
$$S.T. \quad U_i(x_i') > U_i(x_{i0})$$

where U_i is the certainty equivalent utility of agent i (for a formal derivation of U_i, see appendix 3); the last term of U_i is a risk premium (RP) which is positively related to s^2, α, and σ^2. The value of α reflects the degree of risk

aversion; the value of σ^2 reflect the per capita uncertainty level.

In the above analysis, I have used EV_i ($EV_i=U_i+RP$) in place of U_i. I do so because I assumed that workers are risk neutral, which means RP=0, and $U_i=EV_i$. If workers are risk averse and there is uncertainty, then workers maximize U_i instead of EV_i when choosing optimal effort. However, if we assume that RP is independent of work effort, maximizing U_i or EV_i will have the same optimal effort in case of interior solution. The only effect of RP is that it may lead to a corner solution even under the condition that $P_{min}<P<P_{max}$. In this case, workers prefer x_{i0} to x_i', and the condition $P>P_{min}$ does not guarantee workers choosing $x_i>x_{i0}$ as in the case of risk neutral workers. This result can be seen more clearly in figure 7. From (3-18) we know that when $P>P_{min}$, the potential gain of agent i from choosing x_i* under the PS system, PGW_i, is greater than zero, which means $EV_i(x_i*)>EV_i(x_{i0})$. In figure 7 we assume $P>P_{min}$, so $EV_i(x_i*)$ is always greater than $EV_i(x_{i0})$, and $U_i(x_i*)$ may or may not greater than $U_i(x_{i0})$ which equals $EV_i(x_{i0})$, because RP shifts the EV_i curve down by a constant value. If PGW_i is larger than RP, agents prefer x_i* to x_{i0}; RP does not affect the choice of x_i*; and we can use EV_i in place of U_i. In the EPS system, however, if the principal sets $x_{i0e}=x_{i0e}*$, the potential gain of agents over the fixed wage system will be zero, and RP takes its maximum value, $\alpha\sigma^2/2$. Since $U_i(x_i*,s*,x_{i0e}*)<U_i(x_{i0})$, effort is reduced to x_{i0}.

To offset RP, agents must earn potential gains which are greater than RP; in other words, x_{i0e} must be smaller than $x_{i0e}*$. We assume, in general, $x_{i0e}=x_{i0e}*-c$, where c is a constant. In this case, the optimal s can be developed as

$$\max_s EV_m=(1-s)Pnx_i-n(w_0-sPx_{i0e})$$

$$S.T.\ x_i=x_i'=\frac{sP}{2}\quad x_{i0e}=x_{i0e}*-c\quad x_{i0e}*=\frac{sP}{4}+\frac{x_{i0}^2}{sP} \qquad (3\text{-}59)$$

$$F.O.C.\quad\Rightarrow\quad s*=1-\frac{2c}{p}$$

The potential gain of the principal over the fixed wage system, PGPE, is

$$PGPE=EV_m(x_i*)-EV_m(x_{i0})=n[\frac{(P-2c)^2}{4}+x_{i0}^2-Px_{i0}] \qquad (3\text{-}60)$$

PGPE decreases as c increases.

The potential gain of agent i over the fixed wage system, $PGWE_i$, is

$$PGPE_i=EV_i(x_i*)-EV_i(x_{i0})=c(P-2c) \qquad (3\text{-}61)$$

As long as $PGWE_i$ is greater than the risk premium, workers prefer offering x_i* to offering x_{i0}. $PGWE_i$ is at a maximum when c=P/4; for c<P/4, $PGWE_i$ increases as c increases.

The principal faces a trade-off between potential gain and risk when selecting the c value. If the c value is set low, PGPE is high, the risk that $x_i=x_{i0}$ and PGPE=0 is also high; if c value is set high, the risk is reduced, and PGPE is also reduced. There is a maximum value of c, defined as c_m. For $c>c_m$, PGPE is negative, and the principal prefers the fixed wage system.

We solve for c_m as follows

$$PGPE=n[\frac{(P-2c_m)^2}{4}+x_{i0}^2-Px_{i0}]=0$$
$$c_m=\frac{P}{2}-\sqrt{Px_{i0}-x_{i0}^2} \qquad (3\text{-}62)$$

There is another critical value of c, defined as c_c. For $c>c_c$, PGPE is smaller than PGP. We calculate c_c as follows:

$$PGPE=n[\frac{(P-2c_c)^2}{4}+x_{i0}^2-Px_{i0}]=PGP$$
$$c_c=\frac{P}{2}-\sqrt{PGP+Px_{i0}-x_{i0}^2} \qquad (3\text{-}63)$$
$$Where \quad PGP=\frac{1}{2}(\frac{P^2}{4}+x_{i0}^2-Px_{i0})$$

The principal will never set $c>c_m$, he may, however, set $c>c_c$ under certain conditions. To see why this could happen, let's relax the assumption that the risk premium is smaller than PGW_i. In the PS system (assume agents are risk averse), if RP is larger than PGW_i, agents offer x_{i0}, and PGP is reduced to zero, this result is not socially optimal if RP is still smaller than the sum of PGW_i and PGP, it is also not optimal for the principal. If the principal can transfer some of the PGP to workers such that the sum of the transfer and PGW_i is greater than RP, workers will offer x_i*, and both parties are better off. In the PS system, the only way that the principal can transfer PGP to PGW_i is offering an s higher than s*. With s>s*, the utility gain of agent i is $CGW_i=EV(x_i')-EV(x_{i0})$. In fact, if PGW_i ($CGW_i(s^*)$) equals RP(s*) at the first place, the agent is indifferent between supplying x_{i0} and supplying

x_i*. Increasing s to as* (a>1), CGW_i and RP increase to

$$CGW_i(s)=\frac{a^2P^2s*^2}{4}-as*Px_{i0}+x_{i0}^2$$

$$RP(s)=\frac{\alpha a^2s*^2\sigma^2}{2}=a^2RP(s*)$$

For P large enough, CGW_i is increasing with P; while RP is independent of P. There is a critical value of P, P_c, for $P>P_c$, $CGW_i(as*)>RP(as*)$, agent i is better off to supply x_i'. Setting $CGW_i(as*,P_c)=RP(as*)$, using $RP(s*)=CGW_i(s*)$, we can solve P_c as

$$P_c=\frac{(a^2-1)x_{i0}}{(a^2-a)s*} \tag{3-64}$$

The condition that $P>P_c$ is not difficult to meet, so as s increases, $CGW_i(s)$ increases faster than RP. The relative increase of $CGW_i(s)$ over RP(s) is, however, very small. If RP(s*) is much higher than $CGW_i(s*)$ at the first place, it will be impossible to have $CGW_i(s)>RP(s)$ even set s=1.

In the EPS system, however, we can set $x_{i0e}<x_{i0}$ to increase $PGWE_i$, in this case, RP does not increase. If s is exogenously set, RP(s) does not change; if s=s*=1-2c/P is endogenously given, setting $x_{i0e}=x_{i0e}*-c<x_{i0}$ means that a larger c is selected. As c increases, both s* and RP decrease. In fact, s* is not significantly affected since for reasonable values of P and c 2c/p is very likely a small number relative to one (3-59). The principal may set $c>c_c$, if $PGWE_i$ is smaller than RP at the first place. In general, the optimal s in the

EPS system is likely greater than the optimal s in the PS system, and we know that s=1 is the socially optimal s, so the socially potential gain is greater in the EPS system, and the EPS system can transfer PGPE to $PGWE_i$, to offset the agents' risk.

Proposition 17: Relative to the PS system, the EPS system has a higher ability to offset agents' risk.

The disadvantage of the EPS system is that the principal is more likely to make mistakes. Since RP is very hard to measure, the principal may underestimate RP, and set x_{i0e} too high. Even when workers are risk neutral, the principal may still set x_{i0e} too high, because in order to set $x_{i0e}=x_{i0e}*$, the principal has to estimate P and x_{i0}. If the estimated P and x_{i0} are incorrect, the principal may set $x_{i0e}>x_{i0e}*$. In this case, agents will offer x_{i0} only. In the PS system, however, the principal has no chance to make such mistakes, because Px_{i0} is observable; the only control variable is s. For given s, the utility gain, CGW_i is also fixed. To avoid workers reduce work effort to x_{i0}, the PS system installed an automatic fixed safety net which is not optimal with respect to the principal as well as the society. The EPS system offered the ability to adjust the safety net, but in the wrong hands, the ability will be harmful instead of helpful. The agents, on the other hand, may prefer the PS system, in which they are better off relative to the EPS system.

3.5. The Lump-Sum Profit Turn Over System

As discussed before, in general, the optimal effort level in the PS system and even in the EPS system are not equal to the socially optimal effort level. In the DEC system, the optimal effort can be controlled by the principal, the principal, however, may not know what is the socially optimal effort. The social optimal cannot be automatically realized. And the DEC system has some disadvantages mentioned in proposition 14. In fact, in the PS system or the EPS system, x_i* would be socially optimal only if s=1. The principal is, however, unwilling to accept s=1 (In the EPS system, if $x_{i0e}=x_{i0e}$*, the principal is willing to set s=1. However, if agents are risk averse, setting $x_{i0e}=x_{i0e}$* would lead to the corner solution that $x_i=x_{i0}$; when $x_{i0e}<x_{i0e}$*, s*<1, the principal is unwilling to set s=1. In order to set s=1 we need private ownership and self-employment which even in the West is not always true. The Chinese rural household production is approximately the case, if we ignore the effect of price scissors. So working effort in the rural household production is no longer a problem.

In the lump-sum profit turn over system, optimal work effort is automatically the socially optimal work effort.

3.5.1. Advantages of the LST System

The LST system can be modeled as follows:

$$EV_i=Px_i-LS-x_i^2 \qquad (3\text{-}65)$$

$$EV_m = nLS \tag{3-66}$$

$$\text{if } x_i = x_{i0} \text{ then } LS = Px_{i0} - w_0$$

where LS is a lump-sum profit turn over, other variables are the same as defined before. The optimal work effort, in case of interior solution can be developed as follows

$$\max_{x_i} EV_i = Px_i - LS - x_i^2 \quad S.T.\ EV_i(x_i*) > EV_i(x_{i0})$$

$$F.O.C. \Rightarrow x_i* = \frac{P}{2} = x_{is}* \tag{3-67}$$

$$For\ P_{\min} < P < P_{\max}* = 2x_{imax}$$

where $P_{min}=2x_{i0}$ (3-11) and $P_{max}*$ is defined as $2x_{imax}$. If $P \le P_{min}$, $x_i=x_{i0}$; if $P \ge P_{max}*$, $x_i=x_{imax}$ (the corner solutions). The optimal work effort in this system is the socially optimal work effort. To determine the value of LS, we need to define two critical values of LS, one is LS_{max}, another is LS_{min}. We define LS_{max} as the value of LS, for $LS > LS_{max}$, workers prefer the fixed wage system, and offer effort x_{i0}; we define LS_{min} as the value of LS, for $LS < LS_{min}$, the principal prefers the fixed wage system. To have both parties accept this system, LS must be somewhere between LS_{min} and LS_{max}. The two values are

$$EV_m(x_i*) = nLS_{\min} = EV_m(x_{i0}) = n(Px_{i0} - w_0)$$

$$\Rightarrow \quad LS_{\min} = Px_{i0} - w_0$$

$$EV_i(x_i*) = \frac{P^2}{4} - LS_{\max} = EV_i(x_{i0}) = w_0 - x_{i0}^2 \tag{3-68}$$

$$\Rightarrow \quad LS_{\max} = \left(\frac{P^2}{4} + x_{i0}^2 - w_0\right)$$

If we are comparing this system with the PS system, in order to have both parties better off relative to the PS

system we need two other critical values of LS, defined as LS_l and LS_h, and we can calculate the two values similarly as before

$$LS_l = \frac{P^2}{8} + \frac{Px_{io}}{2} + \frac{x_{io}^2}{2} - w_0$$
$$LS_h = \frac{3P^2}{16} + \frac{3x_{io}^2}{4} + \frac{Px_{io}}{4} - w_0 \qquad (3\text{-}69)$$

We need $LS_l<LS<LS_h$ to have both parties prefer the LST system to the PS system. Since $LS_h-LS_l=PGW_i>0$, there is room for both parties to be better off relative to the PS system. LS_h-LS_l is also the net social gain relative to the PS system. And since $LS_i>LS_{min}$ and $LS_h<LS_{max}$, even though the condition that $LS_l<LS<LS_h$ does not hold, the agents still offer x_i*, as long as we have $LS_{min}<LS<LS_{max}$. So in the LST system the principal can adjust the value of LS to solve the horizontal equality problem without loss efficiency. And the principal can also adjust the value of LS to offset risk premium.

Proposition 18: The advantages of the LST system are: a) the optimal work effort is automatically the socially optimal effort; b) the horizontal equality problem in the PS system is solved in this system; and c) the ability to offset a risk premium is high in this system.

3.5.2. Acceptability of the LST System

As mentioned before, when $LS_l<LS<LS_h$, both parties prefer the LST system to the PS system; when $LS_{min}<LS<LS_{max}$, both parties prefer the LST system to the FW system. In Chinese

state enterprises, however, the PS system is much more popular than the LST system (see chapter 4 for evidence). What are the reasons? In this section, we discuss the acceptability of the LST system, and focus on the acceptability to the principal for two reasons. One of the reasons is that the principal generally has more power to make decisions; another is that the principal is more likely to be the loser. In fact, the real payment systems are selected by either the principal or mutual agreement. When the payment systems are selected solely by the principal, the agents can only react passively by changing their work effort accordingly. The principal may start a new payment system as long as it's acceptable to him. When the system must be mutually agreed upon before being adopted, the principal is more likely to be the loser. Only the agents know the true P value and true x_{i0}. They agree to introduce the new system only when they become better off under the new system. The principal may, however, accept a new system which he believes would make him better off which in fact makes him worse off. Besides, the agents have stronger incentive to bargain because the benefit generated from a better deal will go directly into their pocket. The principal may, however, have a weaker incentive to bargain because even though the benefit generated from a better deal is positively related to his utility, it does not directly go into his pocket, which makes a lot difference.

It's easy to have the principal prefer the LST system to

the FW system as long as the P value is fixed, since LS_{min} is observable (see 3-68). However, to have the principal prefer the LST system to the PS system, we need LS greater than LS_t which is a function of P and x_{i0}. If the principal has no reliable information of the P value, we come back to our old question---can the principal use the P value reported by agents to calculate the LS_t value? The answer is similar to that from the PS system. If the principal can only observe the value of Px_{i0}, LS is fixed for all following periods, agents report qP (q<1) as the true P value, and x_{i0}/q as the true x_{i0}, then if the principal believes the agents, the agents are better off and the principal is worse off. To see how could this happen, let's assume that the principal sets $LS=aLS_t$, where a>1. The potential gain to agent i by misrepresenting the P value is

$$\begin{aligned} PGC_i &= LS(P, x_{i0}) - LS(qP, \frac{x_{i0}}{q}) \\ &= a(\frac{P^2}{8} - \frac{q^2P^2}{8} + \frac{x_{i0}^2}{2} - \frac{x_{i0}^2}{2q^2}) \\ &= a[\frac{P^2}{8}(1-q^2) - \frac{x_{i0}^2}{2q^2}(1-q^2)] \\ &\quad for\ P > \frac{2x_{i0}}{q} \quad PGC > 0 \end{aligned} \tag{3-70}$$

If $P>P_{min}/q$, we have $PGC_i>0$. Since the condition that $P>P_{min}/q$ is not very difficult to meet, agents are likely to have incentive to misrepresent the true value of P. In this case, however, the optimal work effort is still socially optimal. The principal may become worse off, unless

independent information about the P value can be obtained. Here, the problem is a distribution problem, not an efficiency problem. Since the payment system is selected by the principal, however, he may not select the LST system. The key to the agents wanting the LST system and the principal not wanting it lies in the agents having superior information about effort and productivity and therefore being able to dupe the principal.

This system can be improved so that the principal is willing to accept the LST system even though he does not know the P value. In fact, if LS is not fixed for several periods, but is subject to changes in each period based on the observed Px_i value, and agents do not continuously misrepresent the P value by reducing their work effort, the principal knows the true P value in the second period, and adjusts the LS value accordingly, then misrepresent the P value in the first period is not a very serious problem preventing the principal from accepting the LST system. Now, the question is that are the agents going to reduce their work effort and continuously confirm the principal's expectation that the true P value is qP ($q<1$). To answer this question, we assume that the principal can observe the value of Px_i. To convince the principal that the P value is in fact qP ($q<1$), the agents have to reduce their effort below their optimal effort. We define the estimated P value of the principal as P_e which is also the P value reported by the agents. If $P_e=qP<P_{max}*=2x_{imax}$,

to confirm the principal's expectation, workers effort level has to be x_{ic}, and $Px_{ic}=P_e x_{ie}$, where x_{ie} is the estimated optimal effort level, and $x_{ie}=P_e/2=qP/2$, so we have $x_{ic}=q^2P/2=q^2x_i*$. We can calculate the net gain of agent i from misrepresenting the P value when $x_i=x_{ic}$, $NGCL_i$, as

$$\begin{aligned} NGCL_i &= EV_i(x_{ic}, P_e) - EV_i(x_i*, P) \\ &= P(x_{ic}-x_i*) + PGC + (x_i*^2 - x_{ic}^2) \\ &= \frac{P^2}{8}[(4-a)q^2 - 2q^4 - (2-a)] + \frac{ax_{i0}^2(q^2-1)}{2q^2} \end{aligned} \tag{3-71}$$

where $NGCL_i$ is the net gain to agent i from misrepresenting the P value and reducing effort to confirm the expectation of the principal in the LST system.

$$\begin{aligned} \frac{\partial NGCLi}{\partial a} &= (1-q^2)\left(\frac{P^2}{8} - \frac{x_{i0}^2}{2q^2}\right) \\ for \quad q &> \frac{2x_{i0}}{P} \Rightarrow \frac{\partial PGCC}{\partial a} > 0 \end{aligned} \tag{3-72}$$

When a=1 we have

$$NGCL_i = \frac{P^2}{8}(3q^2 - 2q^4 - 1) + \frac{x_{i0}^2(q^2-1)}{2q^2} \tag{3-73}$$

From (3-73) we can see that the second term is negative, the first term is also negative if $q<0.7$, so agents will not misrepresent the P value too much. If $q>0.7$, the first term is positive but the value is very small. If x_{i0} is small enough so that the second term is very small in absolute value, we may have $NGCL_i>0$, but the net gain of

misrepresenting is very small, and q must be significantly greater than 0.7. If x_{i0} is large enough, we will have $NGCL_i<0$ all the time, and agents have no incentive to misrepresent the P value continuously. There is a critical value of x_{i0}, defined as x_{i0c}, for $x_{i0}>x_{i0c}$, $NGCL_i<0$ all the time. x_{i0c} can be calculated as

$$x_{i0c}=\sqrt{\frac{2q^2A}{1-q^2}} \quad where \ A=\frac{P^2}{8}(3q^2-2q^4-1) \qquad (3\text{-}74)$$

The maximum $NGCL_i$ is obtained when $x_{i0}=0$ and q=0.86 (the first term is maximized when q=0.86). As the P value increases, $NGCL_i$ increases. If $P=P_{max}*/q=2x_{imax}/q$, we have the maximum $NGCL_i$ equal to $0.0845x^2_{imax}$. If $P_e>P_{max}*$, then $x_{ie}>x_{imax}$, new workers should be employed. Because in the LST system the principal has only weak control of the production process, agents are given more freedom to manage their production. The condition that $x_i<x_{imax}$ does not hold any more, so we can extend the result developed when $P_e<P_{max}*$, by substituting x_i* for x_{imax}.

Proposition 19: In the LST system, if the principal can observe the value of Px_i, if the value of LS depends on the expected P value and expected x_{i0}, and LS subject to readjustment in each period, and if x_{i0} is high, then agents have no incentive to reduce their effort below the optimal level; if x_{i0} is low, agents have only weak incentive to misrepresent the P value by reducing their effort below the optimal level.

Even in the case that $NGCL_i>0$, agents still face a trade-off; they can report $P_e=qP$, but do not reduce their effort to confirm the principal's expectation, so next period, the principal will expect the true P value by observing Px_i*, and change LS accordingly. In this case, the agents get a once-and-for-all gain from misrepresenting the P value, which is equal to PGC_i (see 3-69). Or the agents can chose to misrepresent continuously by reducing their effort to confirm the principal's expectation, and gain $NGCL_i$ in each period. If agents expect there are n periods to gain $NGCL_i$, then if $PGC_i>nNGCL_i$; agents will not reduce effort, if $PGC_i<nNGCL_i$, they will chose misrepresenting continuously. By increasing the value of a in the LS, PGC_i increases proportionately, $NGCL_i$ increases less than proportionately, so the condition that $PGC_i>nNGCL_i$ is more likely to be hold. If the principal can use other information sources to estimate the P value, the value of n decreases, and agents are more likely to chose not to reduce effort.

For example, if P=20, q=0.8, a=1, then $P_e=16$, $x_i*=10$, $x_{ic}=6.4$, $x_{i0c}=4.24$.

$x_{i0}=2$, $NGCL_i=3.915$ $PGC_i=16.87$ $PGC_i/NGCL_i=4.3$

$x_{i0}=4$, $NGCL_i=0.62$ $PGC_i=13.5$ $PGC_i/NGCL_i$PGCC=21.7

$x_{i0}=6$, $NGCL_i=-5.005$

So, as long as the principal can adjust the LS value in each period and x_{i0} is not too low, the principal will prefer the LST system. The agents are, however, not willing to

accept the LST system with LS adjusting each year, especially when the principal is the state government and the agent is the manager of an enterprise. The agents are afraid that the principal may set LS too high in the following periods when the true P value has been revealed. If agents reduce work effort to x_{i0}, the principal may use the revealed value of P to criticize the agents for their low work effort at the first place. And the opportunity cost to accept a LST system with LS adjusted each year is very high, when it's possible to change into a LST system with LS fixed for several years. As a result, most Chinese state enterprises currently under the LST system have the LS values fixed for at least several periods. In this case, the principal cannot use the reported P value to calculate LS_t, and has to set LS exogenously. Two problems follow when LS is fixed for several periods. First, the P value may increase over time because of technology improvement, so that the LS value based on the current P value and greater than LS_t now may become smaller than LS_t when the P value increases. To solve this problem, the LS value is subject to a predetermined increase each year. This is an extra source of uncertainty. If agents are risk averse, they will not be willing to accept the LST system when the LS plus the risk premium of uncertain technology improvement is so high that their utility in the LST system is lower than that in the PS system. So, the predetermined increase of LS is very often too low to offset the effect of P value increases

over time. Second, even if the P value is fixed over time, a mutual agreement about the P value has to be reached based on bargaining. When only the agents know the true P value, and the incentive of the agents to bargain is stronger, the P value is very often underestimated, and the LS set too low. The principal may prefer the PS or EPS system with s exogenously selected, so that he can claim at least a share of the increased profits.

The LST system is, however, still the socially optimal system with respect to effort level. And the government officials may overestimate the effect of misrepresenting the P value. As discussed above, agents in this system have higher income relative to the PS system, if the principal sets $LS=LS_l$. The difference of income between the two systems comes from two sources: a) by misrepresenting the P value, agents can reduce the LS value; b) the optimal effort level is higher in the LST system. The government officials observed that the income level in the LST system is higher, but they often ignore the fact that the effort level in this system is also higher and interpret the high income as the result of misrepresenting the P value, so the distribution problem is often over-estimated.

If we relax the assumption that the principal's effort is independent of labor productivity or the P value, then under the LST system with fixed LS value over time, the principal will have no incentive to increase the P value. This will

cause social losses. To solve this problem, the LS value has to be adjusted over time. As long as the LS value can be adjusted, the principal may accept the LST system.

3.5. Concluding Remarks

Five payment systems have been discussed in this chapter based on a simple principal-agent model. We start by assuming that both the principal and the agents are risk neutral; the team size is small enough such that there is no free-riding behavior (or the team size is one for simplicity); the expected social return of work effort is Px_i; the social cost of work effort is x_i^2 which is borne by the agent in all payment systems. Under these assumptions, it's easy to verify that the socially optimal work effort is P/2 (when marginal return equals marginal cost). Intuitively, this model is similar to a pie distribution. The "pie" (Px_i) is made by the agent, so the size of the pie is determined by the agent (for given P). Since the cost to make the pie (x_i^2) is borne by the agent, the utility generated from the volume of the pie received by the agent must be greater than the disutility of the cost, otherwise the agent will not make the pie at the first place. Since the "tools" used to make the pie belong to the principal, he has the privilege to determine how the pie is distributed. If the principal is also the agent, distribution of the pie is irrelevant, and the optimal size of the pie is socially optimal.

In the FW system, the agent's payment is independent of his performance; work effort is only affected by monitoring. If monitoring is ineffective, either due to asymmetric information of work effort, or due to poorly defined property rights, such that only a minimum work effort (x_{i0}), which is significantly below the socially optimal work, is guaranteed, then the optimal work effort is x_{i0}. Using the pie distribution analogy we can see that since the volume received by the agent is fixed, and the cost to make the pie increases with the size of the pie, the agent is better off to make a pie as small as possible.

To increase work effort, four incentive payment systems are introduced. The agents' optimal work efforts under all payment systems are given in figure 6. In general, the optimal efforts under the incentive payment systems are greater than x_{i0}. In the PS system, the "pie" is shared by the principal and the agent. In terms of the agent, the return of work effort is only sPx_i, so the optimal work effort is $sP/2$ which is below the socially optimal work effort, because the optimal s of the principal is smaller than one. In this system, the pie is cut twice. In the first division, the whole pie is divided into three parts. The volumes of part 1 and part 2 are fixed and received by the principal and the agent accordingly. Part 3 is whatever left, and it is shared by the principal and the agent in the second division. The share ratio s is determined by the principal. The principal,

when determines the value of s, faces two choices: he can select a smaller share of a larger pie, or a larger share of a smaller pie. With a larger s, the principal's share is smaller, the pie is, however, larger, because the optimal x_i increases with s. The principal's optimal s maximizes the "volume of the pie he received," and it is smaller than one. In this system, s is the only control variable.

In the EPS system, the pie is still cut twice, in the first division, however, the volume of part 1 (received by the principal) is no longer fixed. The principal has two control variables. He can select a higher s, such that the agent is willing to make a larger pie; then cut part 1 larger, such that the loss in part 3 (due to a smaller share) is fully or over compensated. So the optimal s in the EPS system is higher than that of the PS system, and the optimal work effort in the EPS system is closer to the socially optimal work effort.

In the LST system, the pie is divided into two parts. However, instead of part 2 (received by the agent) is fixed as in the FW system, part 1 is fixed in the LST system. The principal can cut a large part for himself, such that the total volume received by him is larger than that in other payment systems, so he is willing to accept this system. Since the marginal return of work effort is P, the optimal work effort is P/2, which is automatically the socially optimal work effort.

In the DEC system, if work effort is observable, the incentive payment can be scheduled such that the utility of the agent is maximized when supplying the principal desired work effort which is approximately the socially optimal work effort. However, work effort is very often not observable, or there is no good proxy of work effort. In this case, the DEC system cannot generate socially optimal results.

These results are, however, based on the assumption that the team size is small enough so that there is no free-riding problem. If the team size is too large so that agent i treats all other agents' effort as a constant, the optimal effort under the free-riding considerations may be smaller than the minimum work effort. Because the marginal cost of x_i is borne by agent i; while the marginal return of x_i is shared by not only the principal, but also all co-workers in the team, the marginal return of x_i in terms of agent i is reduced significantly as the team size increases. In this case, the free-riding problem will invalidate all incentive payment systems. The free-riding behavior is an extreme case that noncooperative agents involved in a prisoner's dilemma game. If we extend the one period model to a multi-period model, repeating games can yield cooperation. In the repeated game, there are only two possible equilibria; either all agents supply x_{if} (the optimal work effort under free-riding behavior), or all agents supply x_i* (the optimal work effort without free-riding behavior). Agents are better off if they

supply x_i*. If the transaction cost of cooperation is lower than the reward of cooperation, rational agents would cooperate. To solve the free-riding problem, we can either increase the reward of cooperation, for example, increase the ratio of incentive wage to total wage, or reduce the transaction cost. The best way to reduce transaction cost is reducing the team size. If the team size cannot be reduced any more, allowing production teams to be organized voluntarily may also significantly reduce the transaction cost.

The optimal efforts under all but the DEC system are positively related to the P value. The relationship between the optimal efforts and the P value is given in figure 6. Under the PS system, agents' utility is positively related to the P value; this causes an equality problem. The teams with a higher P value are better off relative to the teams with a lower P value. This equality problem can be solved in all other incentive payment systems.

If we assume agents are risk averse and there is uncertainty, then agents maximize U_i (the certainty equivalent utility) instead of EV_i (the expected utility of a risk neutral agent). The difference between U_i and EV_i is the risk premium which we assume is independent of work effort. The risk premium does not affect the optimal effort, but does affect agents' choice between the optimal effort and the minimum effort (see figure 7). Agents must earn a net gain

which is greater than the risk premium under the incentive payment systems relative to the FW system. The ability to offset risk premium is higher in the LST system or the EPS system than that of the PS system.

At this point, we cannot say which incentive payment system is always better than others. We can say, however, that under certain conditions a specific incentive payment system is the best. And it seems safe to say that all incentive payment systems as we modeled are better than the FW system.

CHAPTER 4

EMPIRICAL EVIDENCE AND POLICY IMPLICATION

The theoretical propositions developed in chapter 3 are based on certain assumptions. In order to explain the real Chinese situation and generate some policy implications, empirical evidence must be provided. For example, propositions in chapter 3 suggest that if the manager of an enterprise is maximizing profit and the team size is small enough so that free-riding is not a serious problem, then under the profit sharing system, work effort, return from the effort, and firm profit are higher than under the fixed wage system. However, if the manager is not maximizing profit, or the team size is too large so that free-riding is a serious problem, then the incentive payment systems do not improve work effort over the fixed wage system. To see the real effectiveness of the payment system reform on work effort, empirical tests have to be conducted. Ideally, I would like to test the following:

1) Other things being equal, which payment system discussed in chapter 3 generates the highest work effort?

2) Other things being equal, how is work effort affected by the control variable in all payment systems?

3) Other things being equal, what is the relationship between team size and work effort (effect of free-riding)?

4) Other things being equal, what is the relationship between uncertainty and work effort (effect of risk aversion)?

Even though comparing all payment systems is interesting and important for policy makers, it is infeasible at this stage because of the limitations on data availability. For one thing, I need to have a sample of diversified enterprises under different payment systems. Secondly, I need to know the specific payment system of each sample enterprises. All these information are not available in the data set used in this study. These information may, however, be available in the future through a better designed survey.

A problem with this type of empirical study is the unobservability of work effort. This problem can be solved by using proxies of work effort. One of the commonly used proxy is labor productivity represented by per capita value added (assume homogeneous workers). Using proxies to represent work effort, however, has to control the variables which also have nonrandom effect on the dependent variable. For example, using per capita value added as dependent variable, and dummy variables of payment systems as independent variables, we might expect that variables such as capital-labor ratio, economics of scale, human capital, price distortion, managerial ability, and environmental conditions all have nonrandom effects on per capita value added. Standard econometrics shows that omitting relevant variables leads to biased estimations of the coefficient. In empirical studies,

all other effects except the control variables are represented by a random error term which is only approximately random at best. So strictly speaking, all real empirical studies are biased. As more and more relevant variables are controlled, the bias may become less and less. Basically, proxies are imperfect measures of relevant variables. Using proxies may reduce the quality of the regressions. This is, however, unavoidable given the data unavailability.

The optimal payment system may change under different conditions. For example, if agents are less risk averse and/or uncertainty is low, the LST system might Pareto-dominate the other systems; if agents are very risk averse and uncertainty is high, the PS system might be the optimal system. When empirical comparison of all payment systems is infeasible, analysis of each of the payment systems become more important. Unfortunately, not all payment systems can be evaluated. In this chapter, only the FW system, the PS system with fixed s, and the DEC system are tested.

A direct test of the effect of risk aversion is very difficult. In this chapter, an indirect test of this effect is conducted. The results of this test can, however, only be treated as weak empirical evidence, because of the distortion caused by omitted variables.

The firm level data used in this chapter are from the World Bank and collected by the Chinese State Statistical Bureau. The sample covers 396 state industrial enterprises.

The data set includes two separate years, 1980 and 1987. In 1980, the Chinese state enterprises were basically under the fixed wage system. Incentive payments took the form of a bonus which was generally independent of enterprise performance. In 1987, the payment system was changed significantly. Not only did the percentage of incentive payment increase, but also the nature of the incentive payment also changed. The incentive payment was dependent, in principle and in reality, on the performance of the firm.

The original sample includes 176 variables which include the annual gross output value in 1980 price, the annual gross output value in current price, the value added in current price, total annual wage, total annual bonus, number of workers, nominal working hours, shut down hours, shut down hours due to power shortage, total profits and taxes, net fixed assets, total sales revenue, sales revenue generated by self-sales, the retained profit ratio, the total welfare funds, and the total reward funds. The availability of the data varies across variables; data items for some variables are missing in some enterprises. The following general features of the sample are based on 348 enterprises; enterprises with incomplete data are excluded. From 1980 to 1987, significant changes of the payment system took place in this sample. In terms of the mean, the percentage of bonus on total wages increased from 14.16% in 1980 to 24.16% in 1987. Table 1 compares average values for some key variables between

1987 and 1980. From 1980 to 1987, the total labor force in this sample increased 27.7%; the per worker annual wage in current price increased 87.7%, from 939 to 1763 yuan; the per worker value added in current price increased 44.6%, from 8652 to 12510 yuan; the per worker profits increased 29.8%, from 7576 to 9835 yuan; the capital labor ratio increased 44.8%. The per worker wage increase is 21.36% of the per worker value added increase, and 36.47% of the per worker profit increase.

In this chapter, model predictions for three payment systems are tested. The FW system is tested using 1980 data; the PS system is tested using 1987 data; and the DEC system is tested using both 1980 and 1987 data. The indirect test of the effect of risk aversion is conducted using 1987 data.

4.1. Work Effort Under the Fixed Wage System

The first hypothesis is that work effort is independent of fixed wages in Chinese state enterprises. A direct test of this hypothesis is infeasible due to limitations on data availability. Since work effort is unobservable, per worker value added and per worker profit are used as proxies of work effort. In 1980 Chinese state enterprises were approximately under the FW system, so average wage of enterprise in 1980 is used as a proxy of fixed wage. The actual hypothesis to test becomes that per worker value added and per worker profit are independent of average wages in 1980. To generate this hypothesis, I assumed that workers are identical and the

minimum work effort, x_{i0}, is only dependent of monitoring, for given monitoring, x_{i0} is fixed. Under the Chinese FW system, average wage of a firm is affected by three factors:

1) Wage difference across regions. Wages of state employees differ from region to region. The rationale behind this difference is that general price levels differ from region to region (different transportation cost is the main reason of the price difference). To equalize real wages, wages in high price regions should be higher. The categorization of regions is unfair because relative price levels change over time, and the categorization is fixed and very rough. This potential unfairness is, however, unlikely to cause resentment, because the difference is small in magnitude, and the unfairness of this difference is not very clear. Most workers even do not know the price levels and wage differences between regions. Workers care more about their relative wage with respect to their peers than that of other regions. So even though I relax the assumption that minimum work effort only depends on monitoring, and introducing the efficiency wage theory (Katz, 1988; Korzec, 1988; Krueger & Summers,1988) that the perception of fairness affects minimum work effort, the wage difference among regions is still unlikely to cause work effort difference.

2) Wage difference across industries. Wage differences across industries are very complicated. The rationale behind this difference is that working conditions are different among

industries, and differences in working conditions should be compensated accordingly. The fairness of the compensations is questionable because working conditions change over time and wage differences across industries are fixed. For example, wages of the same grade are in general higher in heavy industries than in light industries because traditionally heavy industries need more physical strength. With increased utilization of machines, jobs in heavy industries may not need more physical strength any more. The unfairness of industrial wage differences is, however, also unlikely to cause serious resentment for the same reasons discussed in wage difference across regions.

3) Wage difference of average tenure. Under the eight-grade wage system, promotion to a higher grade is based on work effort and performance (in principle); in fact, promotion is based on seniority. If firm A has more older workers than firm B, other things being equal, the average wage of firm A is greater than firm B. If we assume that older workers have higher skills and that minimum work effort is the same for all workers, we should expect a positive relationship between average wage and per worker value added. This is questionable because seniority may not always be positively related to skill, especially for simple jobs. During the very early stage, experience may be positively related to skill, after a certain period of time the relationship between seniority and skill is unclear. Using per worker value added or per worker

profit as dependent variable, average wage as independent variable, the relationship between labor productivity and average wage can be tested. If there is a positive relationship between wage and productivity, there are three alternative explanations: i) labor productivity is positively related to average tenure; ii) average wage in 1980 is not a good proxy of fixed wage; and iii) work effort is positively related to fixed wage (the efficiency wage theory). The empirical result cannot distinguish these alternative explanations. On the other hand, if the relationship between wage and productivity is insignificant, not only the hypothesis that productivity is independent of average tenure is not rejected, but also the original hypothesis (work effort is independent of fixed wages) is not rejected. According to standard economic theory, labor productivity is also a function of economies of scale, and capital labor ratio, so they are included in the regression equation. I use the net fixed assets divided by the number of workers to represent the capital labor ratio, KL. Since there is no direct measurement of enterprise scale, following the traditional method, gross output value, G, is used to represent enterprise scale.

A source of distortions is omitted variables; price distortion, managerial ability, human capital, and environmental conditions are all affecting productivity. In Chinese state enterprises, price distortion is especially likely to have non-random effect on value added or profit.

However, it is reasonable to assume that price distortions are smaller within an industry than across industries. So regression by industry is likely to reduce the effect of price distortions. Regression by industry is also useful to reduce the effect of omitted firm-specific variables. In this section, 10 three-digit industries (among a total of 33 three-digit industries based on the Chinese industrial classification) which have large enough sample sizes are regressed independently. The total sample includes 10 two-digit industries. Regressions by two-digit industry are also conducted.

Following a conventional method by assuming that the relationship between per worker value added and the independent variables is log-linear (Wang, 1990), the regression equation is

$$\log(VA) = a + b\log(W) + c\log(G) + d\log(KL) + \epsilon \qquad (4\text{-}1)$$

where VA is per worker value added, W is average wage, G is gross output value, KL is capital labor ratio, a is a constant, b, c, d are coefficients, and ϵ is the error term which is assumed to be a normally distributed random variable with zero mean and finite variance.

The results of the regressions (OLS) using 1980 data are given in table 2 and 3, where table 2 is the regressions by three-digit industry; and table 3 is the regressions by two-digit industry. Replacing the dependent variable by per capita profits, similar regressions are conducted and the

results are given in table 4 and 5 respectively. The results of these regressions demonstrated a consistent picture. Among the 10 three-digit industries, only in one industry (the machine industry) the wage coefficient is significant at the 10 percent level (table 2); among the 9 two-digit industries (the mining industry is excluded because its sample size is too small to conduct an independent test), also only in one industry the wage coefficient is significant at the 10 percent level (table 3). Using per worker profits as the dependent variable, the results are similar: only one two-digit industry survived the 10 percent confidence level test (table 5); there is one three-digit industry which also passed the 10 percent confidence level test; the sign of the wage coefficient is, however, negative (table 4). In most industries, the coefficients of scale are significant, which might be interpreted as that economics of scale is important in this sample. To see the overall wage effect of the total sample, 10 dummy variables, each representing a two-digit industry, are added into (4-1). Five of the industry dummies are significant at the 10 percent level. The result of the regression is

$$\begin{aligned} va = & -3.17 + \underset{(1.21)}{0.31w} - \underset{(-0.69)}{0.03kl} + \underset{(14.35)}{0.42g} - \underset{(-5.22)}{0.88D1} + \underset{(6.49)}{0.62D2} \\ & + \underset{(3.30)}{0.61D5} - \underset{(-4.09)}{0.56D8} - \underset{(-5.44)}{0.53D9} \end{aligned} \qquad (4\text{-}2)$$

R^2=0.57 DF=363 T-ratios are in parentheses

where va, w, kl, and g are logVA, logW, logKL, and logG respectively. The insignificant industry dummies are excluded. D1 is the dummy for mining industry, D2 for food industry, D5 for power plants and oil refinery, D8 for steel industry, and D9 for machine, tools, ship, and automobile repair. Price distortions are most serious in these industries. In China, the prices of raw materials and agricultural products were and still are below their market values (Wiemer, 1990). Because most of the enterprises in the sample mining industry are coal mines, the negative sign of the coefficient of D1 may reflect the fact that coal price is set too low. The positive signs of the coefficients of D2 and D5 are also consistent with our expectation, because the raw materials used in food industry (the agricultural products) and in power plant and oil refinery (coal and crude oil) are very cheap, the outputs of these industries are, however, receiving relatively high prices. As a raw material, steel has a price which is also below its market value, which is reflected by a negative coefficient on D8. The reason for the negative coefficient on D9 is unclear. We know that many of the enterprises in this industry are unprofitable for many reasons besides price distortion. A detailed analysis of these reasons is, however, beyond this study. The main conclusion we are interested in is that after controlling for all relevant variables the wage coefficient is still insignificant at the 10 percent level. It seems reasonable to

conclude that under the FW system productivity and profitability (both are proxies of work effort) are independent of fixed wages in the sample Chinese state enterprises.

There is a problem with these results. By definition, value added equals profit plus wage and welfare expenditure, if higher wage does not lead to higher value added, then either profit is lower, or welfare expenditure is lower. In fact, enterprises can choose, to a certain degree, the division between wage and welfare expenditure, so one explanation that both value added and profit are independent of wage is that enterprises with higher wage spent less in welfare. Another explanation is that the wage variance is too small, such that the relationship between wage and value added (and profit) is econometricly insignificant (in fact the wage variance in 1987 is about 8 times greater than that of 1980). If this is the case, the none hypothesis is still not rejected.

4.2. Work Effort Under the Profit Sharing System

In chapter 3, it is shown that changing from the FW system to the PS system increases work effort, and if s is exogenously given, work effort is positively related to s. These propositions, however, depends on the following assumptions:

1) Monitoring is ineffective such that the minimum work effort is below the optimal work effort.

2) the free-riding problem is not too serious such that the optimal work effort under free-riding consideration (3-24) is below the minimum work effort.

3) Agents are not so risk averse and uncertainty is not so high such that the utility of agents when they supply x_i* is below their utility when they supply x_{i0}.

Since the wage system reform in China is very complicated, a detailed discussion of this reform is conducted later. At this point, suffice it to say that the wage system reform is basically changing from the FW system to the PS system with an exogenous s. And we are interested in whether or not the wage system reform is associated with work effort increases. To answer this question, statistical regressions are needed.

After an exhaustive search of the literature, Weitzman (1990) found sixteen formal econometric studies on the relation between profit sharing and productivity. These studies vary greatly in data sources, methodologies, and attempts to control for biases. "Salient problems in the estimation of the effect of profit sharing on productivity include all the standard econometric problems of production function estimation, the potential endogeneity of profit sharing, and omitted variable biases due to the unobservable character of managerial quality and other firm-specific variables" (Weitzman, 1990). Generally speaking, these

studies use either value added or sales per employee (in logarithm form) as dependent variables; profit sharing is measured as a dummy variable, profit share per employee, profit share as a percentage of compensation, and/or percent of employees covered by profit sharing. The specification is most often based on a Cobb-Douglas production function, several studies also use the more general constant elasticity of substitution (CES) and translog functions. In the studies that try to control for endogeneity, the most common approach is instrumental variables. The use of instrumental variables does not, however, change the results greatly relative to the ordinary least squares (OLS) specifications. The general picture that emerges from the econometric studies is that profit sharing and productivity are positively related.

The methodology and specifications used in these studies are also useful for our purpose. I use value added and profits as proxies of work effort and control for capital labor ratio and economic scale. The only problem is that there is no data about profit share. However, if we can assume that average fixed wages are the same across firms, or the difference of fixed wages across firms is very small such that we can approximately assume fixed wages are the same, then the difference of average wage in 1987, when most of state enterprises have adopted the PS system, is caused by differences of profit sharing. We can use average wage as a proxy of profit sharing. In fact, in 1980, the variance of

average wage in our sample is only 16000; by 1987, the variance increased to 125664; about 8 times as large as that in 1980. Adjusting for the mean of average wage increase, variance of average wages in 1987 is still more than doubled (the average wage of the sample in current price increased 87.7% from 1980 to 1987. If each worker's wage were increased the same percentage, the variance of average wages in 1987 would have been 3.52 (1.877^2) times as large as that in 1980. The actual variance in 1987 is, however, about 8 times as large as that in 1980). Assume that the increased variance is mainly caused by differences in incentive payments. Using average wage in 1987 as independent variable, we can test the relationship between productivity and profit sharing. Since regression by industry can only reduce the effect of price distortions, the estimated coefficients may still be biased. However, if we can assume regressions of the FW system and the PS system are biased in the same direction, the relative significance of the wage coefficients and even the estimated coefficients still reveal some information about the effect of wage system reform. The question is that do we expect the relative values of the omitted variables changed nonrandomly between 1980 and 1987. Among the omitted variables, there is no clear reason to expect managerial ability, human capital, and environmental conditions changed nonrandomly between 1980 and 1987. Price distortions is possible to change nonrandomly because of the price adjustments and the price reform in

general. As a result of the price adjustments, the relative planned prices were changed to reduce price distortions; the effect of these price adjustments, however, was very limited as a matter of fact. The price reform which attempted to change the price determination mechanism from planned pricing to market pricing was also not very successful to reduce the price distortions (Wiemer,1990). As a result, price distortions were still very serious in 1987, especially in state enterprises. The sample data set includes an item which gives the increased (decreased) revenue caused by price changes of inputs and outputs. The values of this items are, however, very small in general. It seems reasonable to assume that the regressions using 1980 data and 1987 data are biased in the same direction.

The regression equation is the same as (4-1), only the meaning of W has changed. Using per worker value added as dependent variable and 1987 data, the results of the regressions by three-digit industry and two-digit industry are given in table 6 and 7 respectively. Using per worker profit as proxy of work effort, the results are given in table 8 and 9 respectively. The results of these regressions demonstrated a consistent picture of the relationship between profit sharing and productivity. The wage coefficients are significant at the 5 percent level in 7 of the 10 three-digit industries (table 6 and 8) and 6 of the 9 two-digit industries (table 7 and 9).

Adding the dummy variables for the 10 two-digit industries into the regression equation, the result of the regression is

$$va = -1.04 + 0.95w + 0.17kl + 0.29g - 1.05D1 + 0.52D2$$
$$(5.91) \quad (3.72) \quad (10.40) \quad (-6.67) \quad (5.84)$$
$$+ 0.29D5 - 0.42D8 - 0.20D9 \qquad (4\text{-}3)$$
$$(1.99) \quad (-3.29) \quad (-2.19)$$

$R^2=0.55$ DF=383 T-ratios are in parentheses

where va, w, kl, and g are logVA, logW, logKL, and logG respectively. The insignificant industry dummies are excluded. Comparing this result with the result in section 4.1, two interesting points can be made:

1) The wage coefficient which is insignificant in the regression using 1980 data becomes significant.

2) The significance and the sign of the coefficients of industry dummies are the same in both regressions, which supports our expectation that price distortions are toward the same direction in both 1980 and 1987.

There may be reason to believe that error terms associated with very large firms have larger variances than error terms associated with smaller firms. To test for the existence of heteroscedasticity, I conducted the Park-Glejser test (Pindyck and Rubinfeld, 1981). I regressed the natural logarithm of the square of the residuals calculated from (4-3) on the natural logarithm of gross output value, G (a proxy of scale),

obtaining the following results (t statistics in parentheses):

$$\log\epsilon_i^2 = -1.64 + 0.08 \log(G_i) \qquad (4\text{-}4)$$

$$(-2.71) \qquad (1.19)$$

Since the estimated coefficient of G is insignificant at the 10 percent level, I conclude that heteroscedasticity is not present.

To eliminate the firm specific effects, we can take advantage of the fact that the data set is a panel data, and use the difference between 1987 and 1980 data to estimate the relationship between wage increases and value added increases. One problem with this approach is that many firms have negative growth in some of the variables used in the regression. Since the logarithm of a negative value is not defined, I cannot use the specification that the relationship between the dependent and the independent variables is log-linear. Assuming the relationship between the dependent and the independent variables is linear, the result of the regression is

$$DVA = -0.18 + 5.36DW - 0.00009DKL + 0.00003DG$$

$$(-1.6) \quad (3.99) \quad (-5.65) \quad (7.32)$$

R^2=0.19 DF=369 T-ratios are in parentheses

where DVA is the difference between 1987 and 1980 per worker value added, DW is the difference between 1987 and 1980 average wages, DKL is the difference between 1987 and 1980 capital-labor ratios, and DG is the difference between 1987 and 1980 gross output values.

Using the number of workers as the proxy of enterprise scale, the regression result is

$$DVA = -0.30 + 6.82DW + 0.0000006DKL + 0.195DL$$

$$(-2.5) \quad (4.88) \quad (0.05) \quad (3.23)$$

R^2=0.096 DF=369 T-ratios are in parentheses

where DL is the difference between 1987 and 1980 total workers, other variables are the same as defined before.

These results also can not reject the none hypothesis that work effort is positively related to incentive wage. The low R^2 might be the result of poor specification.

It seems reasonable to conclude that labor productivity is positively related to profit sharing in 1987 in the sample enterprises. The causality between productivity and profit sharing, however, is not clear from the regressions because "regression techniques assume but do not prove causality" (Pindyck and Rubinfeld, 1981). The positive relationship between labor productivity and incentive wage in 1987 can be interpreted differently. In one hand, we can say that because of the wage system reform, workers realized that increased work effort which in general improves firm performance leads to higher money income, and respond to this expectation, they increase their work effort accordingly, which in turn increases labor productivity and incentive wage payment, so generated the positive relationship between labor productivity and incentive wage. In this case, the causality is from wage to labor productivity. On the other hand, enterprises can

increase profits or productivity through other channels instead of increased work effort. Since under the new wage system, enterprises with increased profits will have increased wage payment, a positive relationship between the two variables could be observed without increasing work effort. In this case, the causality is from labor productivity to wage payments.

The real situation is very complicated; there are basically two forces pushing in the opposite directions. In one hand, the state government, under the pressure to keep horizontal equality, struggles continuously to make sure money payment is only positively related to work effort. To reach this target, very often the profit level of the previous year is selected as a reference point, only the profits above this reference point is shared by workers. Workers in a very profitable firm may not receive any incentive payment unless profit is increased. The potential profit growth rate is also adjusted. A commonly accepted method to adjust potential profit growth rate is using the average profit growth rate of several previous years as a reference point. Only the current profit above previous level plus the average growth can be shared by workers. On the other hand, the state enterprises, under the pressure of workers to increase real payment regardless of work effort, struggles continuously to increase workers' real income. Facing rather strict government

regulations toward money payment, enterprises can increase workers' real income through 2 channels:

1) Workers' real income can be increased through in-kind payment or subsidized low price consumer goods supply. Firms can avoid government regulations by including the cost of these payments into production cost. The more profitable a firm is, the easier to include these payments in production cost and still have a reasonable profit level. As a result of this opportunity, in-kind payments are very popular in China (Walder, 1987; 1989). As a matter of fact, in-kind payments or welfare level is higher in profitable firms regardless of work effort. This cannot, however, explain the positive relationship between money wages and profitability, since in-kind payments are not included in money wages.

2) Workers' real income can be increased through increased money payment. To increase money payment, however, the growth rate of profits should be greater than before. Profits also can be increased through two channels. One is increased work effort; another is increased management effort which may or may not contribute to social gains. Managers may take advantage of the price distortions or other weakness of the planned economy, for example, reducing the production of in-plan low price products and increasing the production of high price products, or reducing input price and/or increasing output price through bribery activities. The government encourages the management effort which leads to social gain

and struggles continuously to prevent the effort which do not lead to social gain. In principle, incentive payment should only be positively related to profit increases caused by increased work effort. Very often, however, the government authorities cannot exactly distinguish profit increases caused by increased work effort from profit increases caused by increased management effort; they, of course, are not totally blind of the differences either. As a result, incentive payments are positively related to both work effort and management effort. The degree that payments related to work effort varies from place to place and from time to time based on the equilibrium of the two opposite forces.

Under the current payment system, even though there is no upper limit for the incentive payment in principle, in reality, there is always an upper boundary about the workers' money income to keep horizontal equality. For some enterprises, it's very easy to increase profits without suffering the disutility of higher work effort and still reach the upper boundary of wage income. In these enterprises, the causality is likely from labor productivity to money payments. However, not all enterprises are this lucky, for many enterprises, without increasing work effort would not increase labor productivity to such a high level, so in order to increase money payments, work effort has to be increased after other less painful potentials to increase labor productivity or profits have been used up. In general, the causality goes

both directions, the regressions cannot, however, tell us to what degree the wage rates changes caused the labor productivity changes and vice versa.

4.3. Effect of Bonus on Work Effort

Under the ideal DEC system, the principal can control workers' optimal work effort by truncating the incentive payments (figure 3). The feasibility of this system, however, depends on the measurability of work effort. If work effort is easy to observe, the principal can compensate workers with a bonus according to observed work effort. In this case, work effort is positively related to bonus. If, on the other hand, work effort is unobservable, there is no way to pay bonus based on work effort. In this case, work effort is independent of bonus payment. Work effort, of course, is determined by many factors. Some of the factors can be observed with "black and white" evidence, for example, nominal working hours; other factors cannot be observed with black and white evidence, for example, intensity of work or investment in certain human capital. In Chinese state enterprises, because of the poorly defined property rights, to pay bonus according to work effort, not only work effort should be observable, but also it should be observed with black and white evidence. In practice, bonus is fixed to a certain percentage of total fixed wage. If black and white evidence is available that work effort is reduced, for example, if a

worker come late or leave early for, say, three times per month or a serious accident in production (an evidence of careless working) is caused by this worker or the production quota is not fulfilled, his bonus will be canceled. So bonus payments do give workers an incentive to work carefully, work on time, and supply a minimum work effort such that the production quota is fulfilled, which in general lead to higher productivity. The effect of a bonus on productivity is, however, limited, because some important factors of work effort, such as intensity of work or investment in certain human capital, is not affected by bonus. In this section, empirical tests using 1980 and 1987 data are conducted to see the effect of bonus on labor productivity. The regression equation is similar to Equation 4-1, only per worker bonus is added as an independent variable. The none hypothesis is that per worker value added is positively related to per worker bonus; the effect of bonus on labor productivity is decreasing over time due to the formation of worker-worker and worker-supervisor coalitions (Lee and Mark, 1990). For a smaller sample (sample size is given in table 10), the results of the regressions are given in table 10; and the following conclusions can be drawn:

1) Bonus coefficients in both 1980 and 1987 regressions are significant at the 5 percent level and positive.

2) Wage coefficient in 1980 is still insignificant after controlling for bonus; and wage coefficient in 1987 is still significant.

3) The signs and significance of industry dummies are similar to that in section 4.1 and section 4.2.

Since the relationship between the dependent variable and the independent variables is log-linear, the effect of wage (or bonus) on value added is not only determined by the estimated coefficient on wage (or bonus), but also determined by the ratio of value added and wage (or bonus). The effects of wage and bonus on value added are given by the following partial derivatives. In 1987

$$\frac{\partial(VA)}{\partial W}=0.557\frac{VA}{W} \qquad \frac{\partial(VA)}{\partial B}=0.238\frac{VA}{B} \qquad (4\text{-}5)$$

where VA is per worker value added, W is average wage, and B is average bonus. The sample means of VA, W, and B are 11514 yuan/year, 1767 yuan/year, and 427 yuan/year (these numbers are little bit different from numbers in table 1, because the sample size is a little bit different). In terms of the sample means, the partial derivatives are calculated as

$$\frac{\partial(VA)}{\partial W}=3.63 \qquad \frac{\partial(VA)}{\partial B}=6.41 \qquad (4\text{-}6)$$

We can interpret these numbers as follows: in 1987, other things being equal, if wage increased one yuan, value added would increase 3.63 yuan; if bonus increased one yuan, value added would increase 6.41 yuan.

In 1980, the sample means of VA, B, and W are 8358 yuan/year, 137 yuan/year, and 946 yuan/year respectively. The partial derivative of VA with respect to B is 13.22; the partial derivative of VA with respect to W is 0.6 (the coefficient of wage is insignificant in 1980)

The reason that bonuses have higher effect on labor productivity than profit sharing might be that bonuses are dependent on individual performance, so that there is no free-riding problem; while the profit sharing system has the free-riding problem which reduces the effect of profit sharing on work effort.

As discussed earlier, the traditional bonus system is in principle a disutility of effort compensation system. Bonus payment is in principle based on work effort level which is measured either by the principal or the face-to-face group discussions. One problem with this system is the formation of the worker-supervisor coalition which will tend to degenerate the incentive payment into a general wage supplement (Cauley and Sandler, 1991). In 1980, when the bonus system was resumed just after the Culture Revolution, the worker-supervisor coalition had not yet taken form. The government had paid a lot attention to make sure that bonus was paid according to work effort, so the effect of bonus on labor productivity was higher. Over time, the worker-supervisor coalition developed, and the bonus started to degenerate into a general wage supplement which, according to our theory,

cannot affect work effort; thus, over time the effect of bonuses was decreasing.

Using the average wage in 1987 as a proxy for profit sharing may cause distortions. If we assume that fixed wages in 1987 are proportional to fixed wages in 1980, and the difference is caused by inflation, then we can use fixed wages in 1980 times the inflation rate as a proxy for fixed wages in 1987, and use the difference between actual wages in 1987 and fixed wages (proxy) in 1987 to represent the profit sharing, WPS. Since wage adjustment is in general delayed from inflation, the official inflation rate between 1979 and 1986, which is 33%, is used to make the adjustment. Replacing average wages by WPS in the previous regression, the result of the regressions is

$$\begin{aligned} va = &- 1.55 + 0.15wps + 0.15kl + 0.31g + 0.26b \\ &\quad\quad (3.99) \quad (2.95) \quad (10.63) \quad (4.78) \\ &- 0.93D1 + 0.46D2 + 0.43D5 - 0.37D8 \\ &\quad (-6.06) \quad (5.02) \quad (2.64) \quad (-2.86) \end{aligned} \tag{4-7}$$

R^2=0.57 DF=334 T-ratios are in parentheses

where wps is log(WPS), other variables are the same as defined before. The sample means of VA, WPS, and B are 11659, 505, and 423 respectively. The partial derivatives are

$$\frac{\partial VA}{\partial WPS}=3.46 \qquad \frac{\partial VA}{\partial B}=7.17 \tag{4-8}$$

Comparing (4-6) with (4-8) we see that the values of the partial derivatives in the two regressions are similar; the

partial derivative of VA with respect to profit sharing decreased from 3.63 to 3.46, and the partial derivative of VA with respect to bonus increased from 6.41 to 7.17. Both profit sharing and bonus are significant at the 5 percent level. So the conclusions generated from the earlier regression still hold.

4.4. Test of Workers' Attitude Toward Risk

In Chapter 3, we have shown that the optimal payment system is dependent on the assumptions about workers' attitude toward risk. With different assumptions about workers' risk attitude, the optimal payment system may change. To make policy related suggestions, it is very useful to know workers' risk attitude. The degree of risk aversion is, of course, determined by the curvature of the utility of income curve. If a worker's utility is linear with respect to income, then the worker is risk neutral; if his utility of income is concave, then he is risk averse. The larger the curvature of the utility of income curve, the higher is his degree of risk aversion (Laffont, 1989). There are many factors affecting the curvature of workers' utility of income curve, one of the most important factors which affecting this curvature is workers' income level, because the curvature changes as income level changes. The curvature of utility of income curve is likely very large around the subsistence income level. The income level of Chinese state workers is, however, not located

in this area. In fact, the basic subsistence needs of the Chinese state workers are covered by their enterprises and/or the state government through heavily subsidized housing, medical care, education, food, and a basic wage which is fixed. The wage system reform is only relevant to the above subsistence level income. It seems reasonable to expect that Chinese state workers are not very much risk averse. This expectation, of course, should be tested empirically.

Unfortunately, direct test of workers' risk attitude is very difficult or even impossible; an indirect test may, however, still improve our knowledge about this topic. In chapter 3, it is shown that if workers are risk neutral, then uncertainty of firm profits will not affect work effort at all; if workers are risk averse, a risk premium is needed to compensate uncertainty. The higher the uncertainty, the larger the risk premium. If the risk premium is so high such that agents' utility when supply x_i* is smaller than the utility when supply x_{i0}, work effort will be reduced to x_{i0} (see figure 7). So if workers are significantly risk averse, uncertainty will be negatively related to work effort. Because both work effort and uncertainty are not directly observable, proxies have to be used to conduct the regression. For work effort, I use per capita value added as its proxy as before. For uncertainty, I use the ratio of the sales revenue generated by self-sales to the total sales revenue, SSR, as its proxy. As a state enterprise, uncertainty comes basically

from two sources, one is the uncertainty of input supply, which will affect production in real term; another is the uncertainty of output sales, which will affect firm profits in money term. If a state enterprise can always sell its products to the state government at a fixed price, there will be no uncertainty of output sales; if, however, the enterprise is responsible to sell its own products in the market, the uncertainty level will be high. So the self-sales ratio, SSR, may reflect the uncertainty level. Problem to use SSR as the proxy of uncertainty comes from the ignorance of uncertainty of input supply, so the result generated from this regression should be treated as a weak evidence. The regression equation is specified as

$$\log VA = a + b\log W + c\log G + d\log KL + e\log B + f\log SSR + \Sigma g_i D_i + \epsilon \qquad (4\text{-}9)$$

where D_is are the industry dummies, SSR is the self-sales ratio, other variables are the same as defined before. The none hypothesis is f=0.

Using 1987 data, the result of the regression is

$$\log VD = -0.79 + 0.58\log W + 0.27\log G + 0.17\log KL + 0.21\log B \qquad (4\text{-}10)$$

$$(-1.67) \quad (2.76) \quad (8.47) \quad (3.08) \quad (3.28)$$

$$-0.02\log SSR - 0.91D1 + 0.46D2 + 0.28D5 - 0.39D8 - 0.18D9$$

$$(-0.58) \quad (-5.52) \quad (4.84) \quad (1.98) \quad (-2.98) \quad (-1.85)$$

$R^2=0.55$ DF=329 T-ratios are in parenthesis.

The insignificant industry dummies are excluded. Since the coefficient of logSSR is insignificant at the 10% level, the none hypothesis is not rejected; uncertainty is

insignificant to explain labor productivity in this regression. This result offered an indirect evidence that Chinese state workers are not very much risk averse under the current wage system.

4.5. Concluding Remarks

This chapter has examined the effects of fixed wage, profit sharing, bonus, and uncertainty on work effort in a sample of 396 Chinese state enterprises, finding that fixed wage and uncertainty have no significant effect on work effort. While bonus has a significant positive effect on labor productivity, the effect was decreasing over time. Profit sharing also has a significant positive effect on labor productivity. Under the profit sharing system (1987), if average wage increases one yuan, per worker value added would increase 3.63 yuan; if average bonus increases one yuan, per worker value added would increase 6.41 yuan. These figures may exaggerate the effects of wage reforms on work effort, because the causality between incentive wages and labor productivity goes both directions. Unfortunately, the regressions cannot tell us to what degree wage changes caused labor productivity changes and vice versa. The regression results should be treated as qualitative instead of quantitative evidences.

CHAPTER 5

REAL SITUATIONS AND POLICY IMPLICATIONS

By 1977, for reasons discussed in chapter 2, the effect of nonmaterial incentives became weaker and weaker; increasing work effort through increased monitoring was also very difficult. The political climate of the early to mid 1970's also inhibited managers from improving monitoring and work discipline. Any efforts toward strengthened monitoring would be entitled "capitalist restoration" or "revisionist theory of the primacy of productive forces." Even though the political climate was changed by 1977, managers had not fully woken up from the nightmare yet. The wage system at that time was a fixed wage system; bonuses had been abolished in 1966. Work effort was low not only because workers were under the fixed wage system, but also because of the low general wage level which according to the efficiency wage theory will reduce work effort further. From 1963 to 1977, wage increases were virtually frozen; in 1977 the average real wage was lower than in 1952 (Walder, 1987). The low general wage level can also be seen through the national distribution analysis. In 1978, individual disposable income was only 49.1% of total GNP, the state and enterprises took 16.7% and 31.7% total GNP respectively; by 1988, the ratios changed to 62%, 12.9%, and 25% respectively (Zhang, 1991). Behind this low wage policy

was a theory that "only when the large river (the state) is full, the small rivers (individuals) may have water." Since the small rivers are in general upper reaches of the large river, the causality of the theory is wrong. Unfortunately, even today many policy makers still insist on this theory in the distribution of the extra profits. The slogan that "state get large share, enterprise get middle share, and individuals get small share" is still very popular in the Chinese literature. General dissatisfaction about the stagnant incomes accumulated to a high level. Work effort was affected seriously. Chinese policy makers realized that wage reforms must be conducted.

5.1. Wage Reforms and the Associated Problems

There were two main problems in the Chinese wage system when policy makers began to design the wage reforms in 1977. The first problem was that the general wage level was too low; the second problem was that worker's payments were independent of their performance. The Chinese policy makers wished to kill two birds with one stone by allowing 40% of the workforce who were working harder than others to increase their wage by one grade. This wage increase, attempted to increase work effort, led to serious contention among workers and reduced work effort. For one thing, all workers believed that they deserve a wage rise after 15 years of wage frozen. Since the low wage policy was instituted in 1957, a large backlog of

pay-related grievances had accumulated. Despite large nominal wage increases, this backlog, coupled with inflation, prevented real living standard from return to the 1956 level until 1984 (Walder, 1987). Secondly, the common method used to determine who should belong to the lucky 40% was face-to-face small group discussions in which vague criteria of performance were applied plus the final decision of the leadership group. Workers' performance was very difficult to evaluate not only because there often were no clear-cut criteria for its measurement, but also because the period to which the performance need to be evaluated was too long. In principal, the 40% should be the workers who on average worked harder than the rest of the workforce at least since the last wage adjustment, which was 15 years ago, not those who worked harder the day or the month right before the wage adjustment. The performance over a long period was, however, difficult to evaluate. So the evaluations often became conflict-ridden. The final decision of the leadership group may cause extra dissatisfactions also because the leaders may use their power to benefit themselves and/or their friends. The contention became more serious also because of workers' expectation that wages would not be readjusted again for a long time. As a result, the lucky 40% believed that they deserved the wage rise and did not increase work effort; those not chosen to receive rises believed that they were treated unfairly and reduced their work effort accordingly. The Chinese policy

makers realized that the two birds cannot be killed with this stone. To reduce the contention over wage matters, later wage adjustments became general wage increases. Since 1976, state industrial wages have increased several times. From 1977 to 1984, the average industrial wage rose 40%, adjusting for the official inflation rate. These wage rises increased the real per capita disposable income among state workers, but the second problem was still unsolved. Workers' payment was still independent of their work effort.

To solve the second problem, Chinese policy makers turned to incentive pay, mainly bonuses. In 1978, the year in which bonuses were revived, incentive pay (bonuses and piece rates) were only 3.1% of the total wage. By 1984, incentive pay increased to 24% (Statistical Yearbook of China, 1985). Compared with the wage rise of 40% of the workforce, using bonus to link payment with performance has the following advantages:

1) Using the face-to-face small group discussions to evaluate work effort was much easier in the case of bonus, since co-workers only need to evaluate the work effort of one month instead of several years.

2) Since each group discussion only affects bonus income for one month, workers do not take it too seriously, so contention among workers was much weaker.

3) Since the bonus was determined on a monthly bases, workers have to work hard continuously in order to receive

high bonus. The problem under regular partial wage rise system that workers may increase work effort right before wage adjustment and relax after wage rises can be solved.

Even though the bonus system has these improvements, some problems still exists. For one thing, the bonus distribution is in principle based on work effort which is very often hard to observe. The small group oral assessment also has many problems which have been discussed in chapter 2. The Chinese policy makers also realized the problems of group oral assessment. By 1981, many enterprises, in response to government insistence, reportedly had abandoned the traditional small group oral assessment and were trying to use more precise quantitative formulas to evaluate workers' performance (Shirk, 1981); such as attendance record, safety record, and the completion of production quota. This approach may improve the bonus system if more precise formulas are available. There was, however, another intrinsic problem with the bonus system; the total bonus fund was fixed or at least independent of enterprise performance. Prior to 1979, enterprise bonus funds were set at 5% of the total wage bill. Coalition of workers developed, bonuses were paid relatively equally, and rotating the high bonuses to different workers each month. Obviously, if all workers compete for high bonuses by increase their work effort as the policy maker's wish, the disutility of effort would increase, but the total bonus fund was fixed, workers would be worse off. So they

would rather distribute bonus equally and not compete by increasing work effort. This problem was also understood by policy makers, and the next step of the wage reform was trying to link the size of bonus funds to the performance of the enterprise as a whole.

The profit retention system was introduced to achieve this target. Under this system, the enterprise retains a certain percentage of its after-tax profits for investment in plant capacity, renovations, and additional bonuses. By 1984, the vast majority of factories had adopted this system (Walder, 1987). This is a profit sharing system by nature, the share percentage was, however, very low at the early stages. According to a survey of 403 state enterprises, the ratio of profit retention was 16.54% in 1980, and 34.17% in 1987 (Du, 1990). Under this system, enterprise bonus funds grew much faster than productivity and profit mainly because of the "soft budget constrain." Enterprises began using a number of legal and illegal means to inflate the wages of workers, including evading accounting regulations. To stop the bonus inflation, a harsh bonus tax was levied on all bonus funds exceeding one-third of the annual wage bill. The tax rate can be as high as 300%. As a result, most enterprises were usually able to pay right up to the four months limit and virtually nothing above it, regardless of enterprise performance. Since bonus payment was virtually limited from above, enterprises turned to regular wage rise which do not

have to pay tax. A new payment system which "link up total wage funds with enterprise performance" or LWP system in short and being called "wage-efficiency link up" (Gong Xiao Gua Gou) in the Chinese literature was introduced in 1985. By the end of 1987, 60% of the large and medium-sized state enterprises were under the LWP system (Xin, 1989); by 1988, 49.2% state industrial enterprises were permitted by the central government to adopt the LWP system, plus the enterprises permitted by the local governments, about 70% enterprises were under the LWP system, accounting about 80% of the total workforce. There were various forms of LWP system. For example, in railway, coal mining, and transportation sectors, total wage bill is linked up with real output or amount of work; in construction sector, total wages are linked up with the output value; and in manufacturing sector, total wages are linked up with the profits and taxes remitted or realized (Pan, 1990). Using real output or output value to represent enterprise performance may lead to ignorance of quality and profit targets; while using profits and taxes to represent performance may also has problems because of price distortions and a lack of fair competition conditions. However, how to represent performance is only a technical problem which should and can be solved according to specific situations. The LWP system is also criticized for the way it links wages with performance. The ratio of profit and tax increase to total wage increase was set by the government as 1:0.75. The real

ratios were above this level. The ratio was 1:0.73 in 1985; 1:1.015 in 1986; 1:1.2 in 1987; and 1:1.04 in 1988 (Pan, 1990). Link up two increase rates by a fixed ratio (for example if the ratio is 0.5, then profits increase 40%, wages increase 20%) will be irrational if the base of profits and taxes is much smaller relative to total wage bill and the potential to increase profits is high. The increased profits and taxes in value may be smaller than the increased wage bill. This problem is, however, easy to correct (link up wage increase and profit increase by values instead of ratios as in the PS system for example) and as long as the base of profits and taxes is significantly greater than the wage bill, the problem is not serious. This system is still a profit sharing system with exogenous factor of share. Along with the LWP system which linked up the workers' income to firm performance is the "contracted system" (Cheng Bao Zhi) which determine the distribution between enterprise and the state. Under the contracted system, the manager of the enterprise sign a contract with the state in terms of the profit turn over (profit sharing or lump-sum), new investment in plant and equipment, the period of the contract, and some other agreements according to specific situations. The contracted system is a combination of the profit sharing system and the lump-sum profit turn over system, with only a few enterprises under the LST system. The contracted system was adopted in some enterprises more than a decade ago and commonly adopted

since 1987. So mainly the current wage system is a combination of the contracted system, the LWP system, and the bonus system.

5.2. Analysis of the Current Wage System and Policy Implications

There are different opinions about the current wage system adopted by Chinese state enterprises. The main attack against wage reforms is that when individuals become better off because of the wage reforms, the government become relatively (if not absolutely) worse off. Even though this is only a distribution problem, after all, there is net social gains after the introduction of the incentive payment system, if this argument is valid, the willingness of the government to extend the wage reforms would be affected. This argument is based on the division of national income. Under the Chinese economic system, national income is distributed among government, enterprises, and individuals. And the share of government revenue in national income decreased. In 1987, the first year when the contracted system was commonly adopted, the share of government revenue in national income decreased from 28% to 24.1%; in 1988, the share decreased further to 19.3% (Xun, 1990). Inflation also contributed to the pressure on government revenues. In chapter 3, I argued that in a properly designed incentive payment system, both the principal and the agents are better off (proposition 4). It's rational

to have an optimal s greater than 0.5 (in this case, the agents' income increases faster than that of the principal), because the cost of extra work effort is borne by the agents. The only way that the principal may become absolutely worse off is that the base of profit turn over (PX_0) is set too low. In fact, if the P value is increasing over time (due to technology improvement), and PX_0 is fixed or adjusted less faster than actual PX_0 increases, the principal may become worse off. To attract all enterprises to accept the PS system, the base of profit turn over was set very low. The base of profit turn over in 1987 was set 0.9% lower than the actual profit turn over in 1986, the base in 1988 was 3.6% lower than the actual profit turn over in 1987 (Xun, 1990). In chapter 3, I argued that the willingness of the agent to accept the incentive payment system depends on his attitude toward risk and the uncertainty level. In the empirical analysis of the sample enterprises, we see that the degree of risk aversion in the Chinese state enterprise is not very high. So our first policy suggestion is that the government should not force all enterprises to accept the PS system by reducing its share of profits.

Another related attack is that the share of individual income in national income increased too fast. Under the traditional Chinese economy, the accumulation (saving) rate was determined by the government, and it was very high (above 30%). Many Chinese policy makers believed that individual

income (mainly wages) is used to "consume" rather than to "invest". When the total Chinese savings deposit reached 700 billion yuan, a panic among policy makers developed. This savings deposit was called a "tiger". And many policy makers believe that "this tiger will break its cage eventually, and swallow the market." The results would be either super inflation or rationing. Under the planning economy, policy makers were accustomed to feel that the economy was under their control. They have not accumulated enough experience to control the economy using market force. They are worried that the current wage system caused uncontrolled increase in individual income that outstripped increases in productivity and profitability. Xin (1989) pointed out that from 1981 to 1988, national income increased 212.7%, while social purchasing power (individual income) increased 271%, (under price control) the gap between aggregate demand and aggregate supply increased continuously. Supply can only meet 92.9% of the demand. This argument is highly debatable because individual income is not equal to aggregate demand. In fact, in 1991 China had a widespread deflation rather than inflation. Still if Chinese policy makers believe that individual real income increased too fast, and it is caused by wage reforms, the will be reluctant to extend the wage reforms. However, the rapid growth of individual income was not caused by the introduction of the incentive payment systems. Individual real income can be increased through

legal and illegal means. The legal means involve various forms of bargaining between enterprises and government agencies in terms of the mandatory production plans, tax breaks and terms of credit, and prices for products. The illegal means include evading accounting regulations; many firms keep two sets of books: one for the upper levels, and the other for internal use (Walder, 1987). If we take a closer look at the increase in individual real income, we can see that most of the increase is caused by non-wage income, such as the in-kind income, or enterprise subsidized low price consumer goods. In fact, from 1981 to 1988, the average annual growth rate of national income was 9.9% in comparable price, while average annual total wage growth rate was only 7.2%. The ratio of total wage to national income was rather stable at the 0.2 level (Xin, 1989). The inflation of individual real income is mainly caused by non-wage incomes. In fact, the wage increase of the enterprises under the LWP system was below the increase of their economic performance. In 1985 and 1986, wage increases of the enterprises under the LWP system were 15% and 9.1% respectively; while the increases of profit and taxes remitted were 20.4% and 9.4% respectively. A survey of 1302 enterprises under the LWP system in 1987 shows that total wage increase was only 5.2%, below the 6.4% increase rate of profit and taxes remitted (Xin, 1989). The inflation of individual real income is not caused by the introduction of the incentive payment systems.

Because cash disbursements are still under administrative control, most enterprises turned to distributions in kind through welfare expenditures, housing construction, and other strategies to distribute income to their workers. One widely-used technique is to charge a lower price when selling goods and receive some consumer goods also in lower prices from the buyer. For example, a steel company may sell cheap steel to a fishing company, and the fishing company will sell cheap fish to the steel company. Workers' real income is increased because they can buy cheaper goods. Some Chinese economists discovered 40-50 different strategies to circumvent state regulations and expend the funds available for workers' welfare consumption. The actual bank expenditures on wages were some 42% higher than the official figures reported by the State Statistical Bureau (Walder, 1989). So individual real income would still be inflated by distributions in kind, even under the fixed wage system. We should not blame the wage reform which if does not make things better at least does not make things worse.

If enterprises can effectively evading accounting regulations, or reduce the reported profits by transfer some of the potential profits into in-kind payments. Workers' real income is not closely related to their performance. A less profitable enterprise (because less work effort is supplied) may still have high real income by evading the accounting regulations more than others. Increased real income from

evading accounting regulations is certainly less painful than increasing work effort. The opportunity to increase real income without increased work effort may make the incentive payment system less effective. So our second policy suggestion is that the government should "open the front door--reduce the control on cash disbursements; and close the side doors--strengthen accounting regulations, forbid the in-kind payments or enterprise subsidized consumer goods." Because the cash disbursements are under strict control, and the inflation rate is high, workers' real income without these "side door" income (in-kind payments for example) is too low relative to the workers in private sector or collective sector where controls over cash disbursements are much weaker. It is very common that university professors earn much less than peddlers. The government agencies also realized this fact, and they are very sympathetic to the workers, as most of the government agencies also earn some "side door" incomes. So evading accounting regulations is not treated as a serious crime as long as the money is not taken by the manager himself, this is why the "side door" payments cannot be stopped effectively. To close the "side doors," we have to open the "front door" at the same time. The illegal means to transfer state wealth to individual income were traditionally controlled by political mass movements, now the government promised that there will be no political movements any more,

but economic laws should be established to control the illegal behaviors of the enterprises and individuals.

Another attack of the wage reforms is that the contracted system caused the short-term behavior of the enterprises. The manager is only maximizing the short-term profits, which lead to reduced investment in plant and equipment, machine overuse, and resource waste. From the Fifth Five-Year Plan (1967-79) to the Sixth (1981-85) productive capital construction projects (investment in plant and equipment) increased by only 6%, whereas non-productive projects (investment in housing and other worker amenities) increased by 129%--70% of the latter for housing." (Reynolds, 1987). To control the short-term behavior of the enterprises, Chinese government relied on administrative interference, such as to put long-term development targets into the contracts. However, if enterprises have no incentive for long-term development, the administrative interference is very often ineffective; the so called "expenditure on technical innovations" might in fact be something else.

The short-term behavior is associated with the contracted system which is explained in section 5.1. There are two main reasons for the short-term behavior of the managers. One is that the period of the contract is too short. If the manager expects that he will not be the manager after this short contracted period, why should he concern the long-term performance of the firm any way. If, on the other hand, the

manager expects he will continuously be the manager in the future, he will have some concern for the long-term performance of the firm. In fact, the enterprises under long-term contracts or expect that they will continuously under the contracted system have demonstrated no short-term behavior. For example, the Capital Steel Company has been under the contracted system for more than ten years, since 1981, its annual profit growth rate has been 20%. According to the original contract, 40% of the retained profit should be used as productive investment; the enterprise itself increased the percentage to 60%. During the ten years, total retained profits is 2.78 billion yuan, most of them are used to develop production, only 25.8% is used in welfare and rewards (Yang, 1990). So, to control the short-term behavior, policy makers should convince enterprise managers that the contracted system will last for a long time. The terms of the contract should be renewed according to changing environmental conditions, but the contracted system should not change. A market for managers should be developed, such that managers will concern their reputations in the long-term.

The second reason which caused the short-term behavior is the pressure from the workers. As Wolder (1989) demonstrated, the Chinese manager "being the director of an enterprise is like being a mayor in a second respect--he is acutely concerned with public opinion. The citizens under his jurisdiction evaluate him according to the success of his

tenure by their standards." A 1985 survey of 900 enterprises found that 80% of employees believed that the manager represents the interests of the factory or its workers; 8% believed the manager represents the interests of the state (Yang, 1987). Managers may leave their jobs after the contracted period, but most Chinese state workers expect long-term employment in the same enterprise. Why should state workers only be concerned with their short-term interests and sacrifice their long-term interests by putting pressures on the managers to maximize short-term income? For one thing, workers are not sure how long the contracted system may last; if the wage system is going to return to the fixed wage system very soon, they would be better off to maximize the short-term interests. Secondly, most workers expect that the profit sharing system is only share profits but not share losses. If in the future, enterprise performance deteriorated, workers expect the state will bear the full responsibility. So in order to control the short-term behavior, policy makers should also convince workers that the contracted system is not going to change in the near future, and workers will also bear the cost of deteriorated enterprise performance. The short-term behavior is not inherently a problem of the incentive payment system. We can solve this problem by improving the current contracted system.

Some economists blamed that the current incentive wage system caused an horizontal inequality problem (Xun, 1990;

Pan, 1990). From Chapter 3 we can see that under the profit sharing system with an exogenous share ratio, workers in the enterprises which have higher potential to increase profits, either because their original work effort is low or because the P value is high, will be relatively better off. The difference of income between enterprises, especially when the difference is caused by unequal objective conditions, will cause the strong resent of workers with relative low incomes. To reduce the horizontal inequality, the state has to offer all kinds of preferential treatment, or tax cut, or subsidies, or even give tacit consent to the illegal behaviors to the enterprises in relative disadvantageous conditions. As a result, all workers' real income increased regardless of their performances. The incentive payment system may loss its potential effectiveness. This problem can be solved by introducing the extended profit sharing system or the LST system with the lump-sum profit remitted adjusted according to the objective conditions of the enterprises. This is not a easy job, especially in the early stage when the principal has no experience to manage the incentive payment system. As time goes by, however, horizontal inequality can be reduced as the value of lump-sum profit turn over is adjusted toward the optimal level. In this case we also need to "open the front door, and close the side doors," give tacit consent to illegal behaviors of enterprises, such as evading the accounting regulation, is a harmful practice.

Another problem which goes to the other extreme is the so-called "whipping the fast ox" problem. Even if the profits of an enterprise increased significantly through increased work effort or management effort, the government superiors may believe that the enterprise received too much favorable treatment, and that they do not deserve such treatment in the future. In chapter 3, I argued that to have both the principal and the agents prefer the LST system to the FW system, the value of LS must be greater than LS_{min}, and smaller than LS_{max} (3-68). If $LS>LS_{max}$, the enterprise may found that it's not worthwhile to offer such a high effort level. However, the manager may not simply reduce work effort and return to the previous situation, because the superiors may use the fact that profits can be high as an evidence that the manager is not doing a good job. To avoid this awkward situation, the manager either keep two different sets of books and build the "slush funds" so that the superiors do not know the real profit level, or not use all the potentials to increase profits, profits will be increased so that the manager is able to pay the regulated maximum bonuses, otherwise the workers will be very unhappy. They will, however, not use all the potentials to improve enterprise performance. To solve this problem, the base of profit turn over should be set rationally, and based on the principle of mutual benefits. Again this is not a easy job, but experience may make things better.

Many economists believe that the current wage reform failed to achieve its full potential to improve enterprise performance or work effort (Lee, 1990; Xiao, 1990;). There are several reasons which tend to undermine the potential effectiveness of the incentive payment systems introduced in the Chinese wage reforms. From chapter 3, we can see that if the team size is too large, incentive payment systems may not improve work effort because of the free-riding problem. This is just the case for Chinese state enterprises. Even though under the ideal contracted system, contracts should be signed to individual level, that is, the enterprises sign contracts with the state, the workshops sign contracts with the enterprise, the production teams sign contracts with the workshop, and workers sign contracts with the team, in reality, contracts may stop at the workshop level or the team level because the lack of internal accounting. "Especially in the early years, many managers had virtually no experience in managing incentive systems. In a number of plants there were no existing quotas, and basic records and shop statistics were in the process of being restored" (Walder, 1987). Shop directors and group leaders have commonly side-stepped this problem by paying out incentive wages relatively equally, which leads to a larger team size.

Very often limitations on cash disbursements caused the enterprises to turn to distributions in-kind or welfare expenditures. In this case, the team size is all the workers

in the enterprise, which is high enough to cause free-riding behavior. So stope in-kind distribution and other "side-door" payments will alleviate the free-riding problem From my regressions, we infer that the effect of the profit sharing on work effort is significantly reduced by free-riding behaviors. So, I suggest that the team size to which the incentive payment is based upon should be reduced further through:

1) Strengthen the internal accounting system. This might be costly and difficult at the early stage when managers have no experience in managing the incentive system. However, instead of side-step the difficulties by practice "egalitarianism," the managers should pay the cost and gain their experience.

2) The state should loose the control over cash disbursements and strengthen the control over in kind or welfare distributions.

3) If the team size cannot be reduced further, we still can reduce the effect of free-riding problem by increase peer pressure of co-workers which can be realized through the following: a) Increasing the ratio of incentive payment to total payment, such that the potential benefit of agent j by put pressure on shirker i increases (see 2-21). b) Developing a secret shirker report system, applying severe sanctions on shirkers being reported by co-workers, and developing a social norm which against shirkers, such that the potential disutility of agent j by put pressure on shirker i decreases.

We should not go to the extremes by either ignore the material incentives or ignore the non-material incentives. c) Reducing the transaction cost of co-workers to reach agreement. For example, allowing production teams to be organized voluntarily.

The second reason which might reduce the effectiveness of the wage reform is that most Chinese state enterprises are under the PS system instead of the LST system. Or even though the enterprise is under the LST system with the state, the workshops or production teams are still under the PS system. While the optimal work effort under the LST system is greater than that of the PS system in general (proposition 18). So we suggest that more Chinese state enterprises should adopt the LST system with the lump-sum profit turn over adjusted each year.

Uncertainty and risk averse agents may also undermine the potential effectiveness of the incentive payment systems. The regression of our sample enterprises dose not support the idea that uncertainty has significant effect on work effort. This, however, may be the result that uncertainty of workers' income is low, because the ratio of base wage and non-wage income to total income is still very high. If we increase the percentage of incentive payment, things might change. However, uncertainty of workers' income can be reduced by build up a enterprise wage reserve fund which serves the function of a buffer. Extra income caused by uncertain

factors should be saved for the "rainy days".

If agents expect that they can increase real income through legal and/or illegal means without increase work effort, the incentive payment system will loss its effectiveness. So the principal should strictly implement the terms of contracts. More specific and fair economic laws should be developed, and a judicial system to implement the economic laws should also be developed.

Even though the current wage reform failed to achieve its full potential, it dose not mean that it is totally ineffective. For one thing, since the effect of non-material incentives has been reducing, if there were no wage reforms, work effort would have been reduced and so would the labor productivity. The fact that labor productivity was not reduced is an evidence that the wage reform is effective. The regressions of our sample also demonstrated a positive relationship between wage reform and labor productivity. In fact, the wage reforms, especially the introduction of the contracted system and the LWP system, did show positive effect on enterprise performance. From 1978 to 1986, the average annual growth rate of total output value of the state in-budget industrial enterprises was 7.5%, while the value in 1987 and 1988 when the contracted system was commonly adopted was 11.1%. The increased profits and taxes in 1987 and 1988 of the same enterprises was 3.7 billion yuan greater than the total increase from 1979 to 1986. The increase of realized

profits of enterprises was even bigger (Zhu, 1990). It seems safe to say that the current wage reforms effectively increased work effort, but they still have many defects which need to be corrected. We should improve the current wage system by deepening the wage reform instead of returning to the original wage system.

CHAPTER 6

CONCLUSION

The work effort of Chinese state employees is motivated through non-material incentives, monitoring, and material incentives. In this study, I argued that while non-material incentives were very effective in motivating work effort during the 1950's and even the early 1960's. Over time, due to increased official corruptions and perquisites, and the stagnation of living standard growth, the effect of moral encouragement decreased. In the market economy, with perfect information about job performance and sufficient power of the employer to fire or punish shirkers, the FW system is often sufficient to generate full work effort through effective monitoring. However, when perfect observation of actions and outcomes is impossible, complete contracting between a principal and agents is infeasible. With asymmetric information about the agent's effort level, three methods can be used to improve work effort: i) direct acquisition of information by the uninformed party-- monitoring arrangements; ii) rearranging the pattern of allocation so that informational asymmetries have a smaller impact---the optimal assignment problems; and iii) design of compensation rules when individual actions are not observed and basic incentives are in conflict (MacDonald, 1984). The principal compares the benefits and costs of all methods, then determines the optimal

method or the combination of methods. Harris and Raviv (1979) show that if the agent is risk averse, then full observability of the agent's action will lead to Pareto improvement, which suggests potential gains to monitoring. Holmstrom (1979) also points out that only under a special condition monitoring is useless. In general there is positive demand for monitoring arrangements. Unfortunately, in Chinese state enterprises, not only is information often imperfect, but also the power of the "supervisor" to punish shirkers is very limited (Lee, 1990) as long as a minimum work effort, x_{i0}, is supplied. I argued that the minimum work effort is significantly below full work effort. Since the ineffectiveness of monitoring is mainly caused by the limited authority of Chinese principal, which is a political problem, rather than asymmetric information, direct acquisition of information is not useful in the FW system. A difference between my study and the norm of agency theory is that standard agency theory treats monitoring as an endogenous variable, but in this study monitoring is treated as an exogenous variable. And the so called optimal work efforts in this study are in fact second-best solutions.

Most literatures of agency theory assume very general forms of utility functions. They correctly identified the major issues, but there are very little positive content. If concrete policy implications are to be derived, more specific forms of utility functions should be assumed. Using the

concepts of agency theory, modelling alternative payment systems of Chinese state enterprises, analyzing these specific models, and generating concrete policy implications are the main contributions of this study. The innovation of this study is the explicit introduction of the P value and the boundary work efforts, x_{imax} and x_{i0}. Their significance with respect to the Chinese situation can be seen when specific solutions are discussed. This study also relaxed some assumptions typically adopted by the norm of agency theory literature but violated in the Chinese case, and analyzed the implications of the alternative assumptions. As in the agency theory, three basic choices are made in this study:

i) Choosing optimal information systems which are used to measure agents' performance. For example, the principal can use work effort or proxies of work effort, the result of individual work effort, or the result of collective work effort to measure agents' performance. ii) Choosing the incentive payment system which generates Pareto-optimal work effort. iii) For each incentive payment systems, find out the value of optimal control variables.

If the result of individual work effort, Px_i, is easy to observe, this study shows that the LST system is potentially the socially optimal system, because the agent claims all the marginal return of his work effort. While this result is also given in the Western agency theory (Harris and Raviv, 1979), the contribution of this study lies on the discussion of the

acceptability of the LST system. It is shown that there are two critical values of the lump-sum paid by the agent, LS_{min} and LS_{max}, if LS is set between them and agents are risk neutral or uncertainty is low, both the principal and the agent would prefer the LST system to the FW system. There are two other critical values of LS, LS_l and LS_h, between them both parties would prefer the LST system to the PS system. The principal may, however, not willing to select the LST system, because LS_l is not known automatically and LS_l is increasing over time as the P value, which is a positive function of capital, technology level, and relative prices, increases. The principal may set LS below LS_l, in this case, he could have been better off if he had selected the PS system. Allowing LS to be adjusted period by period may solve this problem.

Even when P is stable over time, to determine the LS value, either by the principal alone, or through negotiations between the principal and the agent, the P value has to be estimated, because only the agent know the true P value. Since the principal can observe the value of Px_i, to convince the principal ex post that the P value is lower than the true P value, the agent has to reduce his work effort below the optimal level, which reduces his income. The agent has to compare the benefit from a lower LS with the cost of reduced income caused by suboptimal work effort. The question is whether it is worth it to misrepresent P by reducing work

effort below the optimal level. It is shown that if LS is fixed over time, the agent has incentive to misrepresent P by supply suboptimal work effort. In this case, the LST system is not socially optimal. The study also shown that if LS is adjusted over time, the agent has no incentive to misrepresent P by reducing work effort (Proposition 19).

If I relax the assumption that the P value is independent of the principal's management effort , then under the LST system with fixed LS over time, the principal has no incentive to increase the P value. This causes social losses. To solve this problem, LS has to be adjusted over time, such that the management effort offered in the current period is compensated in the following period through increased LS. As long as LS can be adjusted, the principal may accept the LST system.

In this study, I assumed that the principal maximizes the profits of the firm, this is an oversimplification of the Chinese case. If LS is determined through negotiations, the agent has stronger incentive to bargain because the benefit generated from a better deal goes directly into his pocket. The principal may, however, have a weaker incentive to bargain because even though the benefit generated from a better deal is positively related to his utility, it does not directly go into his pocket, which makes a lot difference. Besides, if the principal is the manager, since the promotion of managers is based on both the approval of supervisory state organs and the election of the Workers Congress (Hong and

Lansbury, 1987), managers must consider both the interests of the state and the interest of the workers. Unless black and white evidences exist, the manager is more willing to give up in the negotiations. If uncertainty level is high, the manager is unwilling to suffer continuous negotiations. In conclusion, the principal is willing to accept the LST system if uncertainty is low and LS is adjusted overtime.

In the LST system, risks are borne by the agent. If the agent is risk averse and uncertainty is high, he may demand a large risk premium to offset the risk. Setting LS below LS_h no longer guarantees that the agents would prefer the LST system to the PS system; if the risk premium is very high, the agents may not willing to accept the LST system even when LS is below LS_{min}. In this case the principal prefers the FW system to the LST system, and there is no LS which is mutually acceptable.

In Chinese state enterprises, payment systems include multi-level structures. In the top level, the state government is the principal, the managers of the enterprises are the agents. The government officials may overestimate the effect of misrepresenting P. The agent in the LST system has higher income relative to the PS system, if the principal sets $LS<LS_h$. The difference of the agent's income between the two systems comes from two sources: i) by misrepresenting the P value, the agent can reduce the LS value; and ii) the optimal effort level is higher in the LST system. The government

official observed that the income level in the LST system is higher, but he often ignores the fact that the effort level in this system is also higher and interprets the high income as the result of misrepresenting the P value, and increases LS accordingly. In the top level, the assumption that the principal cannot force the agent to supply work effort above x_{i0} may not hold. If LS is adjusted over time, and after several periods the LS value is set above LS_{max}, the agent would prefer the FW system and reduce work effort to x_{i0}. because the agent is the manager of an enterprise, he may find that it is difficult to reduce effort due to the pressure from the state government. This phenomenon is commonly called "whipping the fast ox". So unless the LS value is fixed over time, the manager is very often reluctant to reveal the true information about the potential productivity of the enterprise by increase work effort to the optimal level. This problem is artificial, improved creditability of the principal may help to solve this problem. The real enemy of this system is uncertainty. However, uncertainty of workers' income can be reduced by building up an enterprise wage reserve fund which serves the function of a buffer; extra income caused by uncertain factors can be saved for the "rainy days". This may reduce the uncertainty level and the risk premium demanded by the agents.

Another problem with the LST system is that some agents produce only intermediate products, so they cannot actually

"pay" the principal a lump-sum. In this case, the agent may pay the principal a lump-sum in real terms, or a variety of the LST system may be used. For example, if a project can be completed in 10 days under the FW system when the minimum work effort is supplied, but can be completed in 5 days if full work effort is supplied. Now the principal makes an offer that 8 days' salary will be paid for this project (quality requirements have to be satisfied) regardless of how long it takes to complete this project. If the agent accept this offer and complete this project in 5 days, his income per day increased and his work effort increased also. If the utility generated from the extra income is greater than the disutility of the extra work effort, the agent will accept this offer, and both parties are better off. If the principal is too greedy and only offered 6 days' salary, the agent might unwilling to accept this offer and continuously supply the minimum work effort. Since the lump-sum payment is determined project by project, which is equivalent to LS adjusted over time, the principal is willing to accept this system. And since the agent can always reduce work effort to the minimum level, he is also willing to accept this system, as long as the uncertainty and/or his degree of risk aversion is not too high. However, if the agent believes that the principal might use the information that this type of project can be completed in 5 days to push the agent increasing his minimum work effort, which is more likely in the market economy where

property rights are well defined, he will be reluctant to accept this offer. If the agent is convinced that the principal cannot or will not use the information to push the agent increasing his minimum work effort, he may willing to accept the system. In China, this is likely true when the agent is the worker and the principal is the manager, but is unlikely true when the agent is the manager and the principal is the state government.

The above discussions can be used to explain why the LST system is not very popular in Chinese state enterprises even though it generates the socially optimal work effort. When the socially optimal work effort is unobtainable, this study shows that the profit sharing system Pareto dominates the FW system. The PS system introduced in this study is slightly different from the traditional profit sharing system. Because only the profit above that of the FW system is shared, both parties prefer the PS system to the FW system. However, because only part of the marginal return of work effort is claimed by the agent, while the total marginal cost is borne by the agent, the optimal work effort is below the socially optimal work effort. The solution in this system is second-best. While these results have been given in the agency theory literatures, the contribution of this study lies on the discussions of some interesting issues which are not discussed in typical profit sharing analyses with the help of the

explicit introduction of the P value and the boundary work efforts.

In the PS system, the principal can either fix the share ratio, s, or select the optimal s, s*, which generates maximum firm profits. Setting s=s* is efficient (second-best), and the agent with higher P is relatively better off. Even though the agent with a higher advantage in P is penalized with a lower s*, which counteracts the effect of unequal endowments and leads to greater equality, changes in s* cannot fully offset the effect of unequal endowments (3-19). Since P is very often independent of work effort of the agent (the agent who work with higher capital-labor ratio, or whose products are favored by the price distortions is better off), we have an equality problem. The principal when dealing with multiple-agents, faces a trade-off between efficiency and equality. In this system, the only control variable is s. In practice, the principal often tries to equalize utilities among agents, and adjusts s accordingly, for the agent with high P value, s<s*, and efficiency is lost. If the principal sets s=s* and ignores the equality problem, workers in relative disadvantage situations may feel unfairly treated, according to the efficiency wage theory, workers will take actions to punish the principal, which also lead to losses. This is a drawback of the PS system.

Incentive compatibility is also an interesting issue in the PS system. From (3-14) we can see that to determine s*,

the principal has to know the value of P and x_{i0}, but in fact very often the principal can only observe the value of Px_{i0} and Px_i; only the agent knows the true value of P and x_{i0}. Now the question is that if the principal has no other ways to determine the value of P and x_{i0} from Px_{i0}, can he simply use the observed Px_{i0} to calculate the P value. In other words, if workers know that the principal will set s=s*, is there any incentive for the workers to misrepresent the true P by reducing their effort below the optimal effort. From (3-21) we can see that as s increases, workers are better off, so workers do have an incentive to convince the principal that the P value is smaller than the true P value, since s* is negatively related to P. If the principal believes that the P value is smaller than its true value, a higher s is chosen, and workers are better off. The optimal effort is, however, positively related to the P value (3-13). If the principal believes that the P value is smaller than its true value, he will expect a lower effort than the optimal one. Since the principal can observe the value of Px_i, in order to convince the principal ex post that the P value is lower than the true P value, workers have to reduce their work effort below the optimal level. Workers may become worse off when supply suboptimal work effort. The question is whether worth it to misrepresent the P value by supply suboptimal work effort. This study shows that if the P value is not very high, and the principal can adjust s period by period according to the

observed value of Px_i, then the incentive of the agents to continuously misrepresent the P value is very weak. They would not reduce their effort to confirm the principal's estimation of the P value ex post (Proposition 9). However, if the P value is very high and x_{i0} is small relative to x_{imax}, workers may supply suboptimal work effort to convince the principal that the P value is lower than the true P value, so that a higher s might be offered. In this case, setting s exogenously will remove the workers' incentive to reduce their work effort (Proposition 10). This result suggests that s should be fixed if P is very high. Incentive compatibility analysis is especially interesting in the Chinese case because the workers tend to be harder bargainers and the managers tend to be weaker bargainers. If the workers have incentive to misrepresent, either the principal is likely to be a loser, or he is unwilling to accept the system.

From above analyses, we can see that under certain conditions fixed s is justified. If s is exogenous and in general unequal to s*, what is the potential loss to the principal? If the principal sets s below s*, social welfare is reduced since the socially optimal s is 1; workers are worse off since their utilities increase with s; the principal is worse off also since his profit is maximized when s=s*. If the principal sets s above s*, social welfare is increased; workers are better off; but the principal is worse off. Since s is determined by the principal, the principal will compare

the losses under the two situations. The study shows that if the principal has to set s exogenously, he prefers to offer a lower than optimal s rather than a higher than optimal s. From proposition 1 and proposition 8 we know that work effort is positively related to s and the social optimal s is 1. The principal, however, prefers to select a lower s, which is inefficient. One way out of this track is to strengthen the audit of the Px_i value. If agents have no way to misreport the Px_i value, the incentive to misrepresent the P value no longer exists in most cases, and s could and should be determined endogenously.

In the one-period PS model, the agent chooses optimal work effort taking P as a constant. The P value is, however, positively related to capital stock, and the capital stock may in turn positively related to previous firm profits which are positive functions of previous work efforts, either because imperfect capital market, so that the profits of the principal can only be invested in the same firm (or consumed), or because the rate of return to capital is low elsewhere. If workers expect long-term employment, as in the case of Chinese state enterprises, they should maximize the present value of their life-time utility, instead of the one-period utility. Intuitively, in the multi-period PS model, when work effort is greater than x_{i0} in period one, not only the extra profits generated is shared by the agent and the principal, the increased work effort also increases the P values in the

following periods as long as the principal invests at lest ϕ percent of this profit share in the same firm, and the life-time worker is better off since the agent's utility is positively related to P. Formally, this study shows that the optimal work efforts of an life-time worker in all periods except the last period are greater than that of an worker employed period by period, as long as ϕ is not zero. And the long-run optimal work effort is positively related to ϕ and negatively related to the workers time preference. Reinvest profit to the same team, of course, may not be good for the society as well as the principal, so if the principal can invest this money on other projects and earn higher return, he will not care too much to increase workers effort in the long run, and the ϕ value will be low or even equal to zero. Only if the original team warrants extra investment, is there a positive ϕ value.

While the above results seems supports life-time employment, the negative effects of life-time employment are not discussed. At this point, I cannot make any suggestion regarding which type of employment should be adopted. However, these results do have policy implications: i) If the value of ϕ is increased, the optimal work effort increases. To increase ϕ, the principal's investment should be closely related to the firm profit. The practice that firms turn over all the profit share of the principal to the state and investment is independently determined by the government in

the Chinese state enterprises leads to zero ϕ, so that the long-run optimal work effort reduced to the one-period optimal work effort. If the firm deserves extra investment, letting the firm using its profits to invest will increase work effort. ii) Even though workers may expect life-time employment, they may not expect the PS system to go on forever (the principal may switch back to the FW system). Unless workers expect the PS system to last "long enough," they may not be willing to maximize their long-run utility. To solve this problem, the principal should not only make a commitment, but also improve his creditability. iii) If workers expect widespread official corruption or large nonproductive investment, so that the ϕ value is small, their work effort will be smaller.

The last issue discussed in the PS system is that if the P value is very high, the optimal work effort in the PS system is the same as in the LST system. For example, if the full work effort is 8, the P value is 40, the share ratio is 0.5, and the minimum work effort is zero, then the socially optimal work effort is P/2=20, the optimal work effort in the PS system is sP/2=10. Since both are greater than the full work effort, we end up with the some corner solution x=8. In the market economy, the assumption that P is very high and nominal working time is fixed are unlikely hold. If the optimal work effort is significantly greater than the full work effort, the employer will hire new workers. As the capital-labor ratio

decreases, the P value also decreases; and as nominal working time increases, the full work effort (work effort is defined as the equivalent of full work effort working hours) also increases. However, in the Chinese state sector, with underdeveloped factor markets, very high P value and fixed nominal working time is as lest possible, if not very common.

In the PS system, the risk is also shared by both parties. If the agent is risk averse and uncertainty is high, he demands a risk premium, RP, to offset the risk. Switching from the FW system to the PS system, both parties gain (proposition 4), if RP is greater than the gain of the agent, the agent supplies the minimum work effort, x_{i0}, and the gain of the principal is also reduced to zero, this result is not Pareto optimal if RP is still smaller than the total gain of both parties. If the principal can transfer some of his gain to the agent such that the sum of the transfer and the gain of the agent is greater than RP, the agent will supply the optimal work effort, and both parties are better off. In the PS system, the only control variable is s, as s increases, RP also increases. This study shows that as s increases, the gain of the agent grows faster than the increase of RP. However, the relative increase is very small. If RP is very high, it might be impossible to have the gain greater than RP even set s=1. So the PS system is inflexible to offset risk.

The PS system has three main problems: i) the optimal work effort is below the socially optimal work effort; ii)

there is an equality problem; and iii) it is inflexible to offset risk. These drawbacks can be improved in the extended profit sharing, EPS, system, which is also a contribution of this study. In the PS system, the profit (or output) in the FW system, Px_{i0}, is selected as a reference point, realized profit above Px_{i0} is shared. The only control variable is s. In the EPS system, the principal can also control the reference point. To realize horizontal equality, the principal can adjust the reference point rather than s, such that efficiency is not sacrificed. The principal can also adjust the reference point to offset risk. Since the principal can adjust the reference point, he is willing to offer a higher s, such that the optimal work effort is greater than the optimal work effort in the PS system. In general, the study shows: i) in the EPS system, the principal has the ability to realize horizontal equality and extract some or even all of the potential gains of the agent by adjusting the reference point, especially when the minimum work effort is low and P is high (Proposition 15); ii) relative to the PS system, the optimal work effort in the EPS system is closer to the socially optimal level. If the agent is risk neutral, at the extreme, the optimal work effort is the socially optimal one (Proposition 16); and iii) relative to the PS system, the EPS system has a higher ability to offset the agent's risk (Proposition 17).

The disadvantage of the EPS system is that the principal is more likely to make mistakes. Since RP is very hard to measure, the principal may underestimate RP and set the reference point too high. In this case, the agent only supplies x_{i0}. In the PS system, the principal has no chance to make such mistakes. The PS system installed an automatic fixed safety net which is not optimal with respect to both parties. The EPS system offered the ability to adjust the safety net, but in the wrong hands, the ability will be harmful.

If production technologies are dependent, i.e., there are externalities, and/or there is economies of scale, compensating individual performance may not be efficient. For example, workers may unwilling to help or teach each other; some workers may hide the tools which are commonly used for their own convenience. Compensating collective performance may solve these problems. And some times only the result of collective work effort of a production team is observable. In this case, free-riding behavior may undermine the effectiveness of all incentive systems. The free-riding problem has been discussed in the agency theory literatures. The contribution of this study is extend the one-period model to a multi-period model; through analyzing the transaction cost of cooperations of coworkers and the potential reward of cooperations, points out some directions to solve the free-riding problem. It is shown that in a repeated game, worker

i will find that if he reduces his work effort, his co-workers reduce their effort too. When the optimal work effort under free-riding consideration is below the minimum work effort, x_{i0}, all workers are only willing to supply x_{i0}. There are two possible equilibria, either all workers supply x_{i0}, or all workers supply the Pareto optimal work effort. Obviously, workers are better off if they all choose the cooperative strategy. There is a potential gain between the cooperative and the noncooperative equilibria. If there is no transaction cost, or the potential gain is greater than the transaction cost, worker will reach agreements that no one should behave like a free-rider. There are three ways to solve the free-riding problem; one is breaking the result of collective work effort into the result of individual work effort through strengthened internal accounting, which is costly and may not be efficient if there are externalities; the second is reducing the team size and/or allowing production teams to be organized voluntarily, and letting more workers in the same team work together and observe each other's work effort such that the transaction cost is smaller; and the third is increasing the potential gain by increasing the ratio of incentive payment to fixed wage.

If good proxies of work effort exist, the DEC system may improve work effort. However, collusion between the managers and the workers may undermine the effectiveness of this system; the optimal work effort in this system is not socially

optimal; and workers have no incentive to investing in human capital and adopting advanced technology. Because work effort is not directly observable, the result of work effort is very often used as a proxy for work effort. However, the result of work effort, Px_i, is composed of two factors, x_i and P. To determine actual work effort, the value of P has to be estimated. With asymmetric information on P, workers have strong incentives to misrepresent P. On the other hand, managers must consider both the interests of the state and the interests of the workers. As a result, the P value is often underestimated. In this system, risks caused by uncertainty, such as power breaks and material shortages are borne by the firm. When negotiating the p values, the effects of uncertainty are often exaggerated. Workers can reduce their work effort and still receive the bonus payment. When good proxies of work effort exist, there is no need to estimate the P value, collusion between managers and workers is unlikely to develop. However, because only work effort is compensated, workers have no incentive to increase P by adopting advanced technology and investing in human capital. If work effort is observable, and workers are identical in terms of P and the disutility of work effort, the manager can select the socially optimal work effort level as a reference point; when workers supply this work effort, a large enough bonus is paid such that the utility form money income is greater than the disutility from work effort. Since workers are not identical

and the manager can select only one reference point, in general the optimal work effort in this system is not socially optimal.

The theoretical modelling of this study serves the function of a "map". It identified the destination and alternative "roads", even though the quality of the roads are mentioned (some are "freeways", some are "highways", and some are driveways"), the best road to the destination depends on the specific location of the "traveller." To use the map, the traveller has to identify his "location" in terms of the uncertainty level, the degree of risk aversion, the P value, the political constraints, the minimum work effort, the transaction cost of cooperation, and the nature of the production technology (are there externalities and economies of scale, are there good proxies of work effort, and how expensive to measure individual performance). For each specific location, suggestions which lead to Pareto improvements (second-best) are made in this study.

The empirical studies and the discussions of real Chinese situations are aimed to test the "map" and identify the location of Chinese state enterprises. Due to the limitations of the data, this study only examined the effects of fixed wage, profit sharing, and bonus on work effort in a sample of 396 Chinese state enterprises, finding that fixed wage has no significant effect on work effort. While bonus has a significant positive effect on labor productivity, the effect

was decreasing over time. Profit sharing also has a significant positive effect on labor productivity. Under the profit sharing system (1987), if average wage increases one yuan, per worker value added would increase 3.63 yuan; if average bonus increases one yuan, per worker value added would increase 6.41 yuan. These figures may exaggerate the effects of wage reforms on work effort, because the causality between incentive wages and labor productivity goes both directions. Unfortunately, the regressions cannot tell us to what degree wage changes caused labor productivity changes and vice versa. The regression results should be treated as qualitative instead of quantitative evidences.

Based on the analyses of real Chinese situations, in respond to the concerns over Chinese wage reforms, this study also suggests: i) the government should not attract all enterprises to accept the PS system by reducing its share of profits; ii) the government should open the "front door"--reduce the control over cash disbursements; and close the "side doors"--strengthen accounting regulations, forbid the in-kind payments and/or enterprise subsidized consumer goods; iii) to control the short-term behavior, policy makers should convince enterprise managers that the contracted system will last for a long time; the terms of the contract should be renewed according to changing environmental conditions, but the contracted system should not change. And a market for managers should be developed, such that managers will concern

their reputations in the long-term. Policy makers should also convince workers that the contracted system is not going to change in the near future, and workers will also bear the cost of deteriorated enterprise performance; iv) Chinese state enterprises should adopt the LST system with the lump-sum profit turn over adjusted each year when conditions permit; and v) if agents expect that they can increase real income through legal and/or illegal means without increasing work effort, the incentive payment system may loss its effectiveness. So the principal should strictly implement the terms of contracts. More specific and fair economic laws should be developed, and a judicial system to implement the economic laws should also be developed.

Even though the current wage reform failed to achieve its full potential, it dose not mean that it is totally ineffective. For one thing, since the effect of non-material incentives has been reducing, if there were no wage reforms, work effort would have been reduced and so would the labor productivity. In fact, the wage reforms did show positive effect on enterprise performance. From 1978 to 1986, the average annual growth rate of total output value of the state in-budget industrial enterprises was 7.5%, while the value in 1987 and 1988 when the contracted system was commonly adopted was 11.1%. The increased profits and taxes in 1987 and 1988 of the same enterprises was 3.7 billion yuan greater than the total increase from 1979 to 1986 (Zhu, 1990). It seems safe

to say that the current wage reforms effectively increased work effort, but they still have many defects which need to be corrected. We should improve the current wage system by deepening the wage reform instead of returning to the original wage system.

APPENDICES

Appendix 1. Tables and Figures

Table 1

Average Values of Key Variables

In 1980 and 1987

Variable	Unit	1987	1980	growth
Total labor force	Thousand	1288.2	1008.6	27.7%
Average wage	Yuan/year	1763	939	87.7%
Per worker value added	Yuan/year	12510	8652	44.6%
Per worker profit	Yuan/year	9835	7576	29.8%
Per worker bonus	Yuan/year	426	133	220%
Capital labor ratio	Yuan/worker	17,100	11,800	44.8%

Note: all values are the mean values of the sample and in current prices.

Table 2

Effect of Fixed Wage on Value Added per Worker

(by three-digit industry)

	Constant	logW	logKL	logG	R^2	N
FP	-1.146 (-0.337)	0.783 (0.614)	0.149 (0.549)	0.34 (2.54)	0.34	20
BR	0.107 (0.028)	0.394 (0.252)	0.151 (0.715)	0.103 (0.713)	0.12	16
CI	-3.512 (-1,636)	-0.930 (-1.654)	-0.135 (-0.399)	0.282 (1.493)	0.48	15
TE	-0.484 (-0.034)	0.919 (1.181)	0.308 (1.699)	0.329 (2.692)	0.63	29
PA	-6.639 (-1.958)	-0.477 (-0.511)	-0.439 (-1.455)	0.581 (2.834)	0.51	16
PH	-6.372 (-2.67)	-0.210 (-0.334)	-0.121 (-0.718)	0.663 (4.611)	0.67	19
RU	-5.241 (-2.327)	-0.200 (-0.237)	0.218 (1.104)	0.580 (6.598)	0.77	18
CM	-2.967 (-1.104)	0.625 (0.665)	-0.008 (-0.050)	0.473 (3.864)	0.58	21
MA	-0.549 (-0.214)	1.928 (2.102)*	-0.212 (-1.017)	0.495 (5.456)	0.65	28
AS	1.525 (0.527)	1.657 (1.682)	-0.140 (-1.144)	0.142 (1.343)	0.45	16

Note: T-ratios are in parentheses (the t-ratios of the wage coefficient followed by * is significant at the 10 percent level), and N is sample size (same for all tables below). FP is for Food Processing; BR for Brewery; CI for Cigarettes; TE for Textile; PA for Paper; PH for Pharmacy; RU for Rubber; CM for Construction Materials; MA for Machine; AS for Automobile and Ship.

Table 3

Effect of Fixed Wage on Value Added per Worker

(by two-digit industry)

	Constant	logW	logKL	logG	R^2	N
TT	-4.621 (-5.333)	-0.131 (-0.446)	-0.028 (-0.520)	0.463 (13.800)	0.39	372
FD	-4.440 (-2.155)	0.009 (0.013)	0.013 (0.095)	0.571 (6.454)	0.45	58
TC	-3.698 (-2.440)	0.003 (0.007)	0.102 (0.910)	0.413 (5.569)	0.63	43
PT	-3.692 (-1.857)	0.170 (0.282)	-0.436 (-2.684)	0.403 (3.903)	0.40	41
EN	-1.779 (-0.329)	0.893 (0.457)	0.290 (1.021)	0.467 (2.690)	0.73	15
CH	-4.031 (-2.388)	0.210 (0.390)	-0.104 (-0.937)	0.505 (6.839)	0.52	67
CM	-2.967 (-1.104)	0.625 (0.665)	-0.008 (-0.050)	0.473 (3.864)	0.58	21
ST	1.390 (0.425)	1.750 (1.589)	-0.050 (-0.199)	0.225 (2.338)	0.52	26
ET	-0.736 (-0.176)	1.372 (1.004)	0.108 (0.439)	0.436 (3.087)	0.56	29
TM	-0.649 (-0.349)	1.264 (1.979)*	-0.159 (-1.552)	0.311 (4.539)	0.44	56

Note: TT is Total Sample; FD includes Food Processing, Brewery, and Cigarettes; TC includes Textile, Clothing, and Leather; PT includes Timber Mill, Paper, and Writing Materials; EN includes Power Plant, Oil Refinery, and Cocking; CH includes Paint, Pharmacy, Nylon, Rubber, and Plastic; CM includes Cement, Glass, and Construction Materials; ST includes Steel Rolling and Smeltery; TM includes Tools, Machines, Automobile, and Ship; ET includes TV, Radio, Meter, and Electronics.

Table 4

Effect of Fixed Wage on Profits per Worker

(by three-digit industry)

	Constant	logW	logKL	logG	R^2	N
FP	-2.089 (-0.507)	0.755 (0.489)	0.184 (0.556)	0.422 (2.620)	0.34	20
BR	-0.347 (-0.073)	0.326 (0.168)	0.176 (0.682)	0.119 (0.675)	0.11	15
CI	-4.730 (-1.967)	-1.193 (-1.892)*	-0.158 (-0.417)	0.337 (1.591)	0.52	15
TE	-1.145 (-0.353)	0.925 (1.056)	0.361 (1.763)	0.397 (2.889)	0.67	27
PA	-1.681 (-0.409)	0.922 (0.887)	-0.326 (-0.957)	0.357 (1.450)	0.46	15
PH	-9.711 (-2.671)	-0.583 (-0.609)	-0.095 (-0.368)	0.933 (4.255)	0.63	18
RU	-7.072 (-2.970)	-0.478 (-0.536)	0.205 (0.984)	0.698 (7.509)	0.81	18
CM	-4.696 (-1.076)	0.637 (0.417)	0.024 (0.091)	0.663 (3.335)	0.50	21
MA	-1.467 (-0.343)	2.358 (1.501)	-0.587 (-1.503)	0.646 (3.941)	0.49	26
AS	2.011 (0.486)	2.275 (1.615)	-0.216 (-1.239)	0.178 (1.179)	0.42	16

Note: FP is for Food Processing; BR for Brewery; CI for Cigarettes; TE for Textile; PA for Paper; PH for Pharmacy; RU for Rubber; CM for Construction Materials; MA for Machine; AS for Automobile and Ship.

Table 5

Effect of Fixed Wage on Profits per Worker

(by two-digit industry)

	Constant	logW	logKL	logG	R^2	N
TT	-5.682 (-5.245)	-0.189 (-0.517)	-0.030 (-0.457)	0.540 (12.745)	0.36	359
FD	-5.817 (-2.453)	-0.231 (-0.277)	0.028 (0.182)	0.648 (6.377)	0.45	57
TC	-6.622 (-3.612)	-0.547 (-1.032)	0.072 (0.530)	0.573 (6.413)	0.67	41
PT	-3.546 (-1.392)	0.526 (0.702)	-0.380 (-1.799)	0.469 (3.548)	0.39	40
EN	-2.225 (-0.369)	0.974 (0.447)	0.319 (1.008)	0.521 (2.687)	0.73	15
CH	-4.911 (-2.140)	0.409 (0.565)	-0.106 (-0.710)	0.642 (6.433)	0.51	66
CM	-4.696 (-1.076)	0.637 (0.417)	0.024 (0.095)	0.663 (3.335)	0.50	21
ST	3.386 (0.829)	2.696 (1.976)*	-0.196 (-0.637)	0.211 (1.712)	0.48	25
TM	-0.578 (-0.211)	1.591 (1.687)	-0.266 (-1.727)	0.325 (3.058)	0.30	54
ET	-8.282 (-1.435)	-0.761 (-0.392)	0.133 (0.426)	0.703 (3.966)	0.55	27

Note: TT is Total Sample; FD includes Food Processing, Brewery, and Cigarettes; TC includes Textile, Clothing, and Leather; PT includes Timber Mill, Paper, and Writing Materials; EN includes Power Plant, Oil Refinery, and Cocking; CH includes Paint, Pharmacy, Nylon, Rubber, and Plastic; CM includes Cement, Glass, and Construction Materials; ST includes Steel Rolling and Smeltery; TM includes Tools, Machines, Automobile, and Ship; ET includes TV, Radio, Meter, and Electronics.

Table 6

Effect of Wage on Value Added per Worker Under Incentive Payment System (by three-digit industry)

	Constant	logW	logKL	logG	R^2	N
FP	1.25 (0.96)	1.29 (2.79)*	0.29 (1.63)	0.09 (0.85)	0.48	20
BR	0.32 (0.10)	0.95 (1.08)	0.66 (2.19)	0.14 (0.61)	0.55	17
CI	0.13 (0.11)	1.04 (2.74)*	-0.05 (-0.19)	0.37 (5.02)	0.90	15
TE	2.50 (2.48)	1.65 (5.10)*	0.39 (4.41)	0.02 (0.26)	0.70	28
PA	-2.82 (-1.44)	0.93 (1.24)	0.20 (0.74)	0.50 (2.83)	0.50	15
PH	-0.12 (-0.07)	0.81 (1.98)*	0.02 (0.12)	0.18 (1.30)	0.37	18
RU	-1.58 (-1.40)	1.09 (2.28)*	0.50 (2.89)	0.41 (4.73)	0.86	18
CM	-1.41 (-1.50)	1.24 (3.82)*	0.05 (0.38)	0.38 (4.58)	0.77	21
MA	0.25 (0.25)	1.21 (3.71)*	0.07 (0.25)	0.17 (1.89)	0.56	28
AS	1.37 (0.51)	1.31 (1.44)	0.00 (0.00)	1.37 (0.14)	0.22	16

Note: FP is for Food Processing; BR for Brewery; CI for Cigarettes; TE for Textile; PA for Paper; PH for Pharmacy; RU for Rubber; CM for Construction Materials; MA for Machine; AS for Automobile and Ship.

Table 7

Effect of Wage on Value Added per Worker Under Incentive Payment System (by two-digit industry)

	Constant	logW	logKL	logG	R^2	N
TT	-1.46 (-2.96)	0.82 (4.46)*	0.23 (4.34)	0.31 (10.15)	0.43	373
FD	-4.39 (-3.11)	0.68 (1.28)	0.17 (1.01)	0.68 (7.73)	0.59	59
TC	0.88 (0.94)	1.15 (4.00)*	0.37 (3.53)	0.11 (1.50)	0.68	41
PT	-0.78 (-0.67)	1.34 (2.98)*	0.01 (0.08)	0.36 (3.65)	0.44	42
EN	-3.10 (-0.65)	0.96 (0.53)	0.16 (0.37)	0.56 (2.24)	0.66	15
CH	-0.92 (-1.18)	0.86 (3.26)*	0.19 (2.09)	0.27 (4.95)	0.55	67
CM	-1.41 (-1.50)	1.24 (3.82)*	0.05 (0.38)	0.38 (4.58)	0.77	21
ST	0.77 (0.75)	1.34 (2.89)*	-0.04 (-0.38)	0.14 (3.46)	0.60	27
TM	1.19 (1.35)	1.31 (4.47)*	0.07 (0.94)	0.07 (1.23)	0.40	56
ET	-0.57 (-0.26)	1.02 (1.49)	0.37 (1.45)	0.26 (2.14)	0.64	29

Note: TT is Total Sample; FD includes Food Processing, Brewery, and Cigarettes; TC includes Textile, Clothing, and Leather; PT includes Timber Mill, Paper, and Writing Materials; EN includes Power Plant, Oil Refinery, and Cocking; CH includes Paint, Pharmacy, Nylon, Rubber, and Plastic; CM includes Cement, Glass, and Construction Materials; ST includes Steel Rolling and Smeltery; TM includes Tools, Machines, Automobile, and Ship; ET includes TV, Radio, Meter, and Electronics.

Table 8

Effect of Wage on Profits per Worker Under

Incentive Payment System (by three-digit industry)

	Constant	logW	logKL	logG	R^2	N
FP	1.81 (0.68)	1.69 (1.99)*	0.48 (1.34)	0.05 (0.23)	0.28	20
BR	3.67 (0.97)	1.82 (1.97)*	0.21 (0.52)	-0.09 (-0.34)	0.38	16
CR	-0.04 (-0.03)	1.13 (2.81)*	-0.09 (-0.34)	0.40 (5.12)	0.90	15
TE	2.94 (1.92)	2.38 (4.86)*	0.51 (3.75)	0.07 (0.60)	0.67	28
PA	-7.60 (-1.43)	1.34 (0.66)	-0.11 (-0.14)	1.11 (2.32)	0.39	13
PH	-0.06 (-0.02)	0.86 (1.44)	0.02 (0.07)	0.14 (0.78)	0.24	18
RU	-2.57 (-2.47)	1.43 (3.23)*	0.21 (1.14)	0.54 (6.93)	0.88	17
CM	-3.03 (-1.91)	1.49 (2.57)*	-0.01 (-0.03)	0.57 (4.05)	0.67	20
MA	-1.97 (-0.75)	1.51 (1.99)*	0.09 (0.12)	0.40 (1.74)	0.33	28
AS	-1.94 (-0.44)	1.10 (0.74)	0.02 (0.11)	0.27 (1.02)	0.25	15

Note: FP is for Food Processing; BR for Brewery; CI for Cigarettes; TE for Textile; PA for Paper; PH for Pharmacy; RU for Rubber; CM for Construction Materials; MA for Machine; AS for Automobile and Ship.

Table 9

Effect of Wage on Profits per Worker Under Incentive Payment System (by two-digit industry)

	Constant	logW	logKL	logG	R^2	N
TT	-2.26 (-3.22)	1.11 (4.26)*	0.25 (3.33)	0.41 (9.30)	0.38	358
FD	-5.17 (-2.79)	0.72 (1.04)	0.31 (1.39)	0.74 (6.35)	0.51	57
TC	0.69 (0.49)	1.67 (3.90)*	0.38 (2.38)	0.18 (1.65)	0.60	40
PT	-2.96 (-1.20)	1.76 (1.96)*	-0.17 (-0.59)	0.67 (3.26)	0.31	40
EN	-0.58 (-0.06)	3.36 (0.98)	0.03 (0.03)	0.67 (1.42)	0.53	15
CH	-1.77 (-1.69)	0.99 (2.83)*	0.12 (0.94)	0.35 (4.68)	0.47	65
CM	-3.03 (-1.91)	1.49 (2.57)*	-0.01 (-0.03)	0.57 (4.05)	0.67	20
ST	1.36 (0.74)	2.37 (2.86)*	-0.18 (-0.93)	0.22 (3.04)	0.55	27
TM	-0.46 (-0.25)	1.47 (2.41)*	0.11 (0.75)	0.22 (1.74)	0.25	55
ET	-2.10 (-0.74)	1.06 (1.14)	0.21 (0.63)	0.39 (2.40)	0.60	28

Note: TT is Total Sample; FD includes Food Processing, Brewery, and Cigarettes; TC includes Textile, Clothing, and Leather; PT includes Timber Mill, Paper, and Writing Materials; EN includes Power Plant, Oil Refinery, and Cocking; CH includes Paint, Pharmacy, Nylon, Rubber, and Plastic; CM includes Cement, Glass, and Construction Materials; ST includes Steel Rolling and Smeltery; TM includes Tools, Machines, Automobile, and Ship; ET includes TV, Radio, Meter, and Electronics.

Table 10

Effect of Bonus on Value Added per Worker

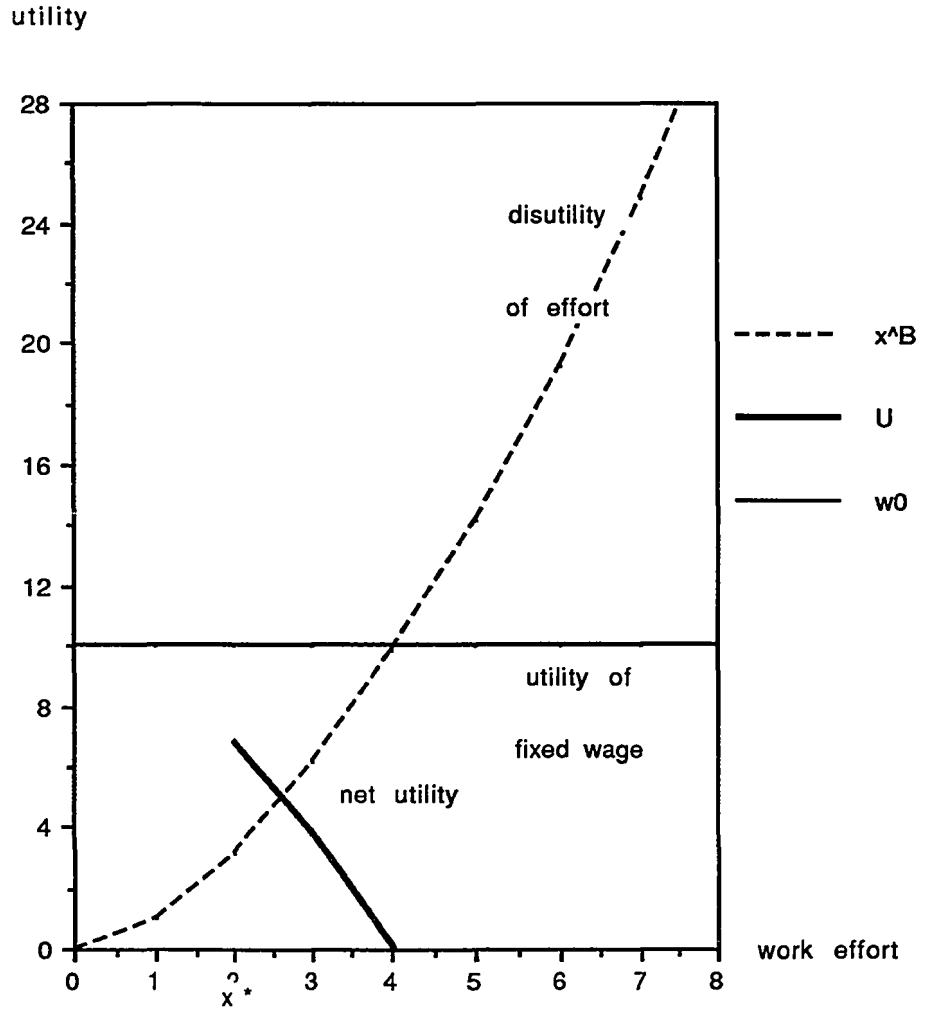
Under Fixed Wage and Incentive Wage Systems

1980	Coefficient	T-ratios	1987	Coefficient	T-ratios
logW	0.068	0.223	logW	0.557	2.756*
logKL	-0.064	-1.243	logKL	0.151	3.127*
logG	0.040	12.787*	logG	0.272	9.214*
logB	0.217	3.083*	logB	0.238	3.835*
D1	-0.874	-4.870*	D1	-0.981	-6.168*
D2	0.576	5.754*	D2	0.507	5.571*
D5	0.622	3.295*	D5	0.313	2.059*
D8	-0.539	-3.839*	D8	-0.405	-3.192*
D9	-0.479	-4.508*	D9	-0.194	-2.013*
CON.	-2.661	-3.246*	CON.	-0.764	-1.668

Note. R^2=0.58, DF=326 in 1980 regression; R^2=0.56, DF=356 in 1987 regression. W is for average wage, KL for capital labor ratio, G for gross output value, B for per capita bonus, CON for constant. The insignificant industry dummies are excluded. D1 is a dummy variable for mining industry, D2 for food industry, D5 for power plant and oil refinery, D8 for steel industry, D9 for machine, tools, ship, and automobile repair.

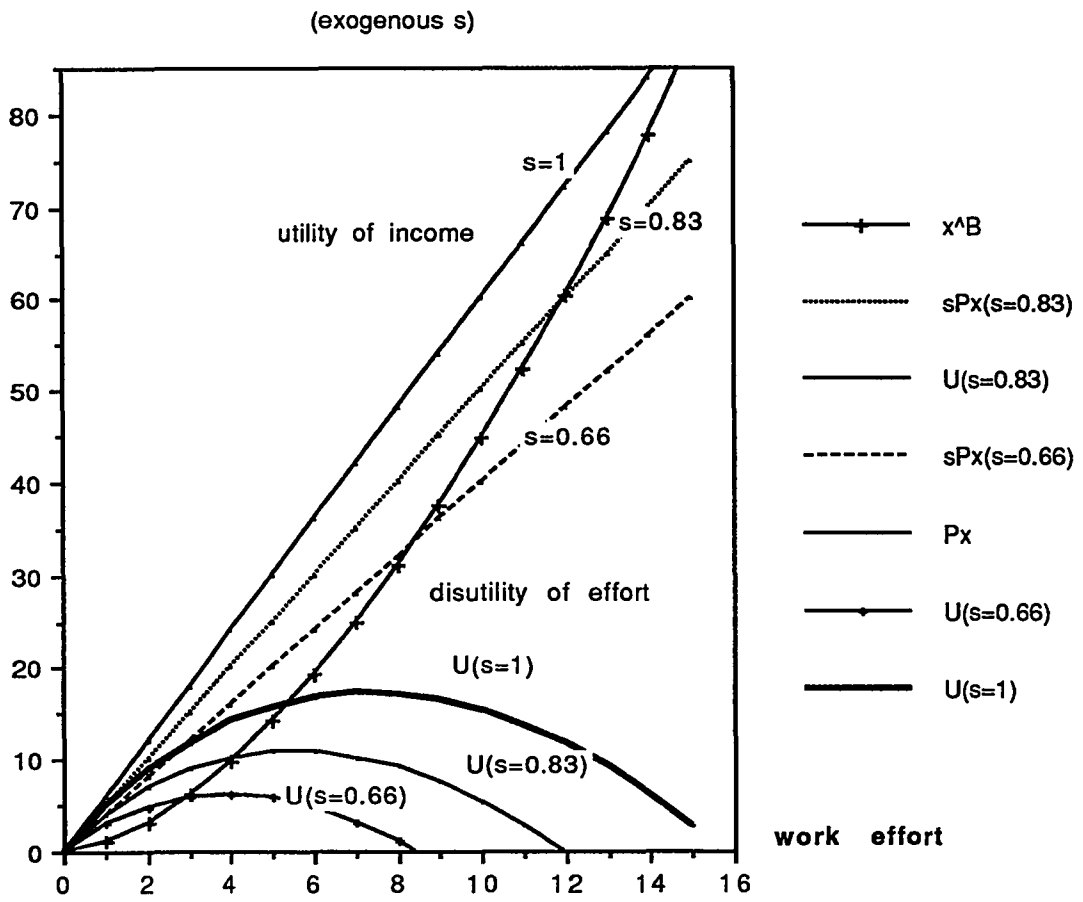
*: The t-ratios followed by * are significant at the 5 percent level.

Figure 1. Optimal effort of the FW System



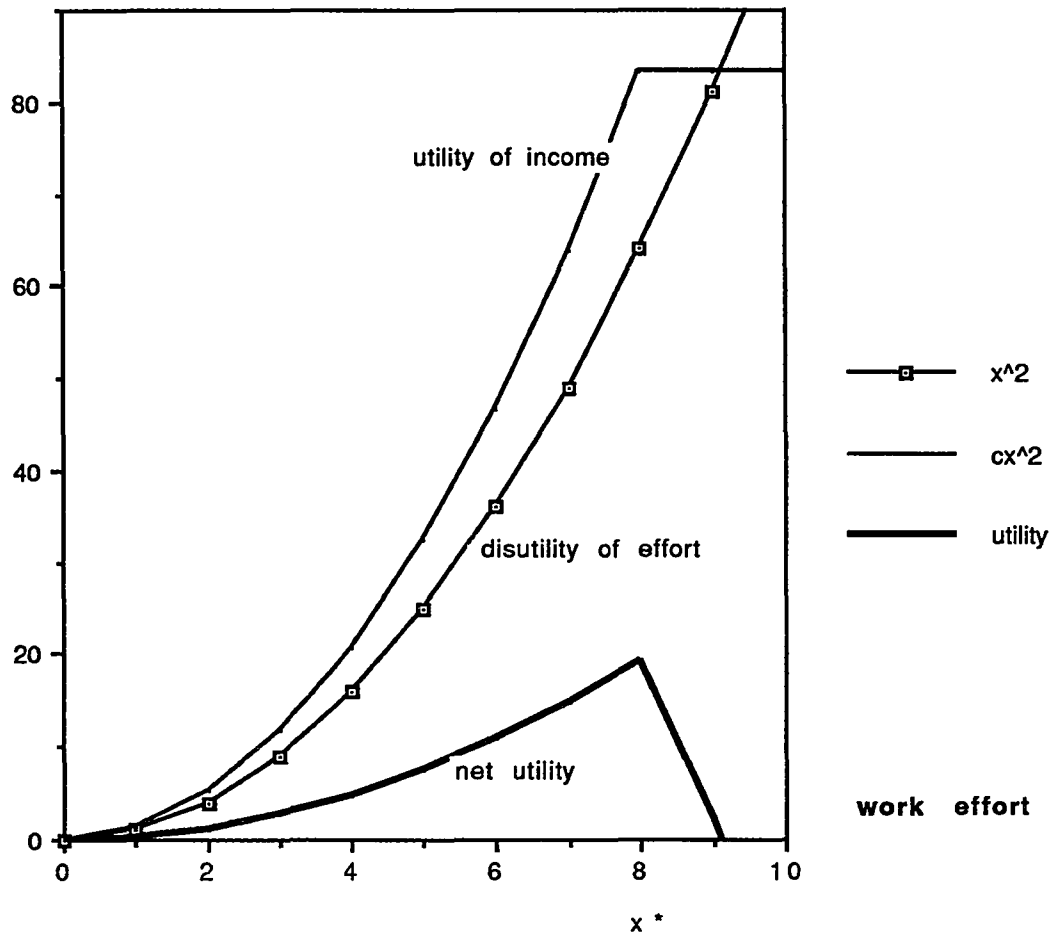
In the fixed wage system, utility of income is independent of work effort; disutility of work effort is increasing with work effort; so net utility of the agent is decreasing with work effort. The optimal work effort of the agent is the minimal work effort under fixed monitoring which is assumed to be 2 in figure 1. Optimal work effort can only be increased through increased monitoring in this system.

Figure 2 Optimal Efforts of the PS System



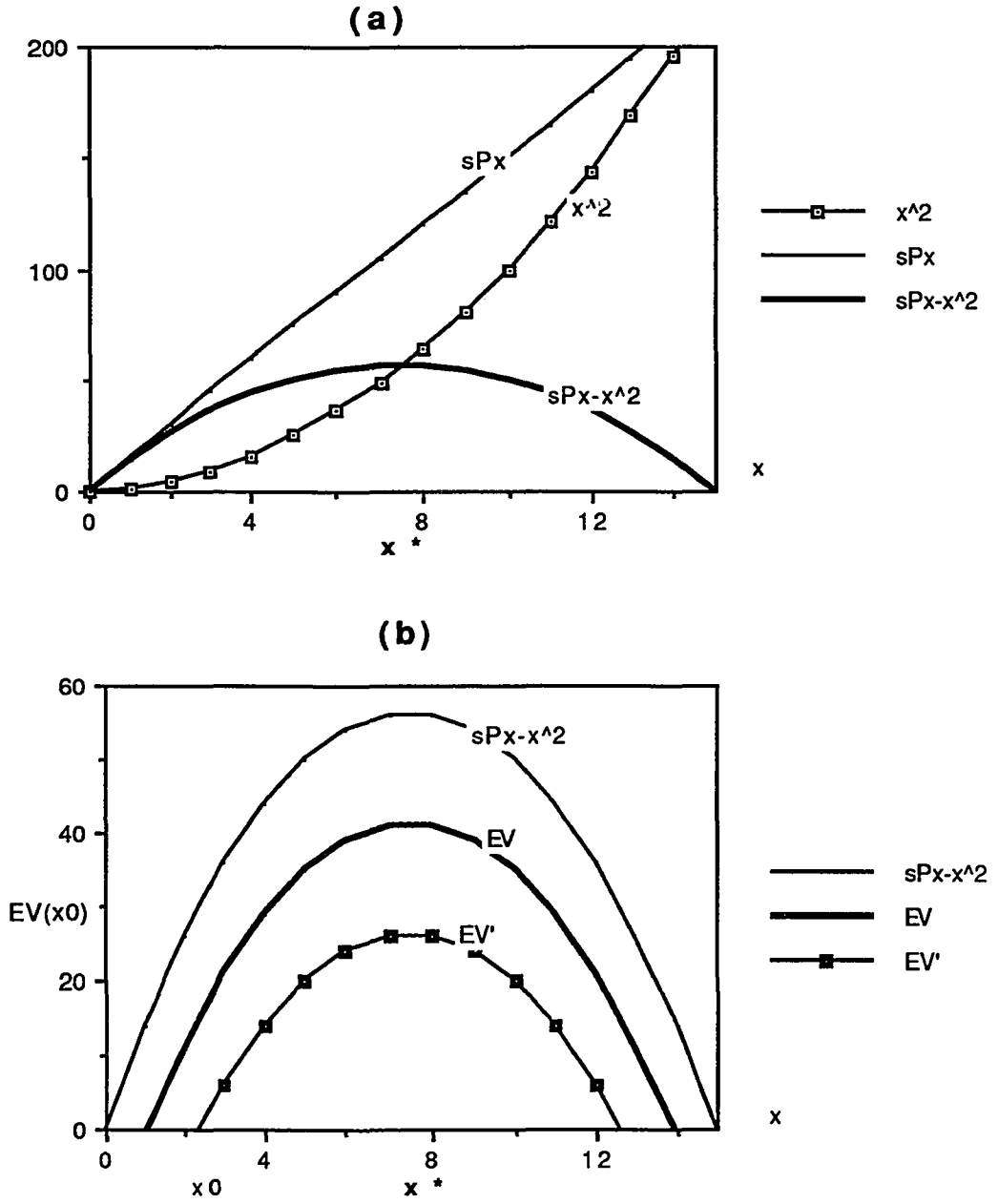
In the profit sharing system, the net utility of the agent is composed by three items: 1) sPx_i which is linearly related to work effort, for given P value, the slope of sPx_i is positively related to s. 2) (w_0-sPx_{i0}) which is independent of work effort. It can only shift the net utility vertically. It cannot affect the optimal work effort. So this item is ignored in figure 2 by assuming that $w_0-sPx_{i0}=0$. 3) the disutility of work effort which is the same as in figure 1. The optimal work effort is positively related to s in this system. When s equals to one, the optimal work effort is the socially optimal work effort.

Figure 3. Optimal Effort of the DEC System



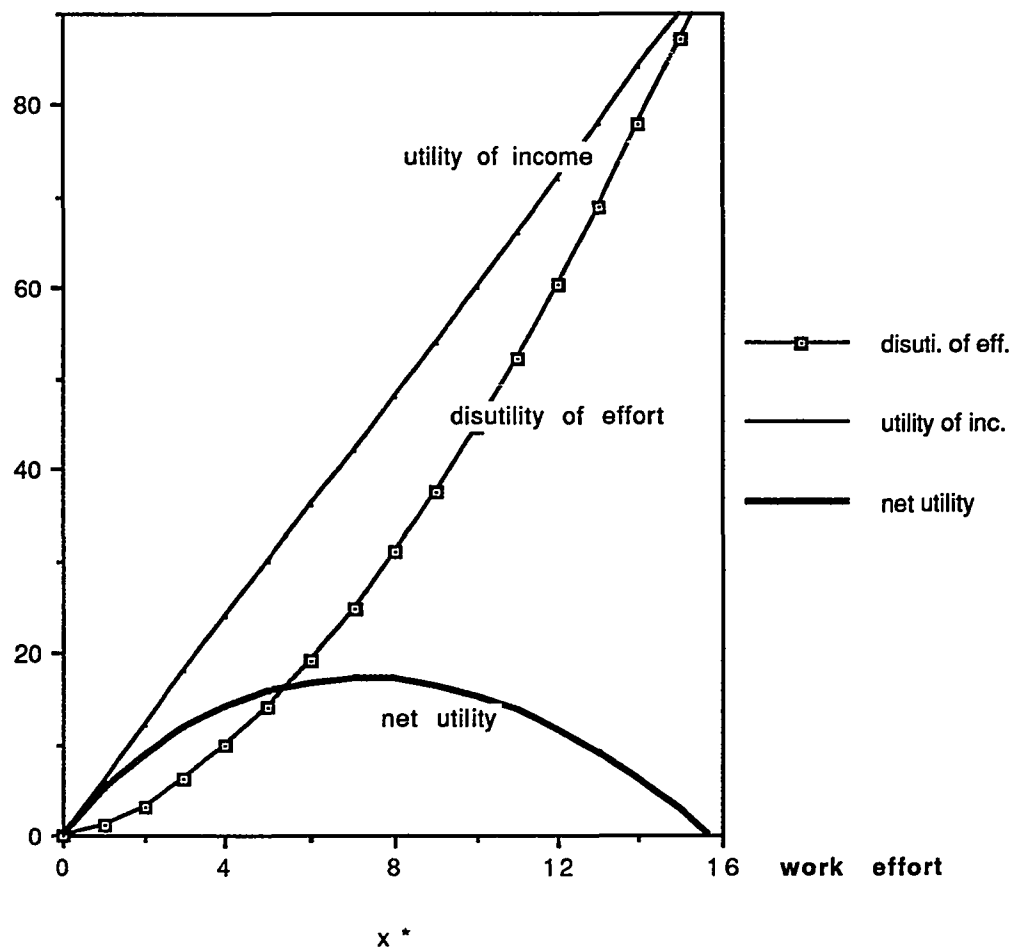
In the disutility of effort compensation system, both the utility of income and the disutility of work effort increase with work effort until the effort level preferred by the principal which is assumed to be 8 in figure 3. After this effort level , utility of income will be fixed, so the optimal work effort of the agent is just the effort level preferred by the principal which may or may not be the socially optimal work effort.

Figure 4. Optimal Effort of the EPS System



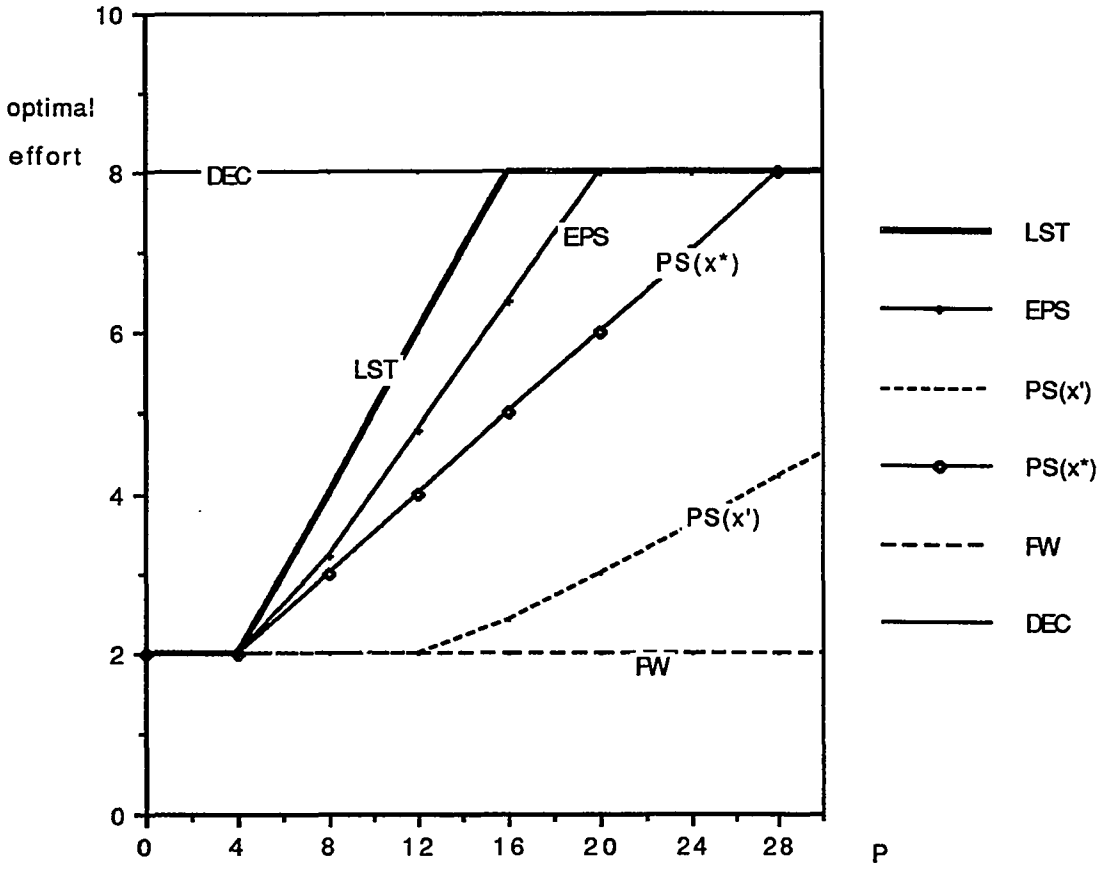
In the EPS system, when $x_{0e}>x_0$ (the subscript i is dropped) the optimal s and optimal effort are greater than the optimal s and optimal effort of the PS system. If x_{0e} is too high, the item of the utility which is independent of work effort (w_0-sPx_{0e}) is very low, which may shift EV curve to EV', and the agent will supply x_0.

Figure 5. Optimal Effort of the LST System



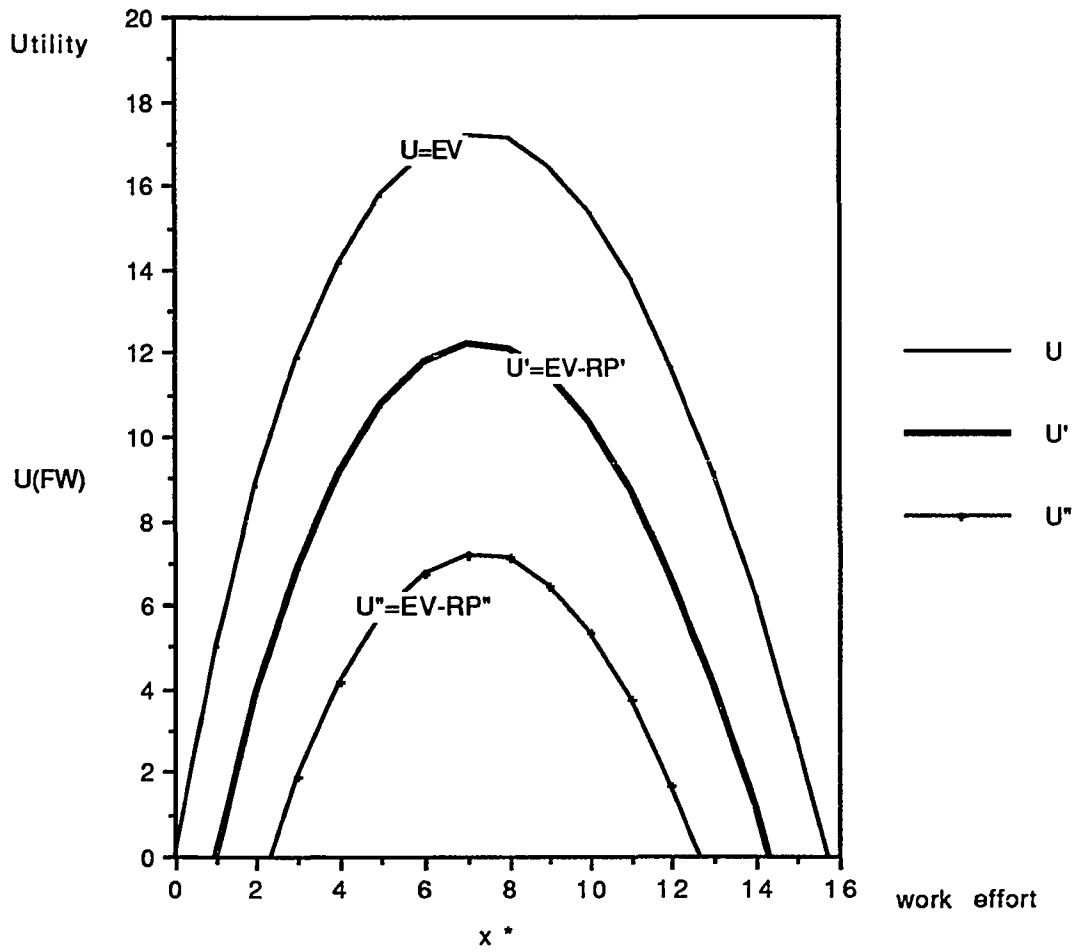
In the lump-sum profit turn over system, the utility of income is linearly related to work effort, the slope of the utility of income is the same as that of the PS or EPS system when s=1. Since the optimal s in the PS or EPS systems is smaller than one in general, the optimal work effort in the LST system is greater than that of the PS or EPS system. The item of the utility which is independent of work effort is ignored in this figure.

Figure 6. Optimal Efforts of Various Payment Systems



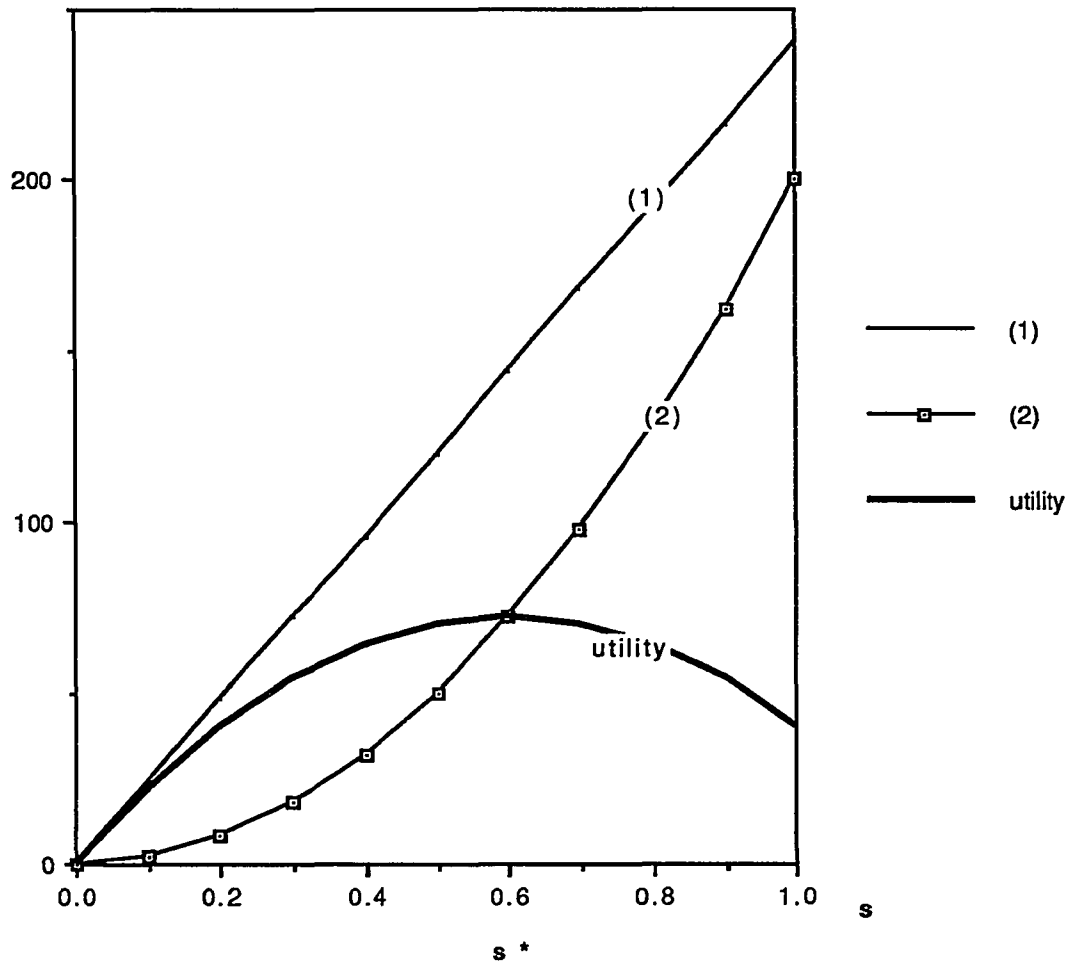
Optimal efforts of all payment systems are compared in this figure. The minimum and maximum work efforts are assumed to be 2 and 8 respectively. the optimal work efforts of all incentive payment systems except the DEC system are increasing with the P value. The optimal work effort of the LST system is the socially optimal work effort; the optimal work effort of the EPS system is closer to the socially optimal work effort than that of the PS system with endogenous s. The optimal work effort of the DEC system is assumed to be 8, but the principal can change it according to his will.

Figure 7. The Effect of Risk Aversion on Optimal Effort



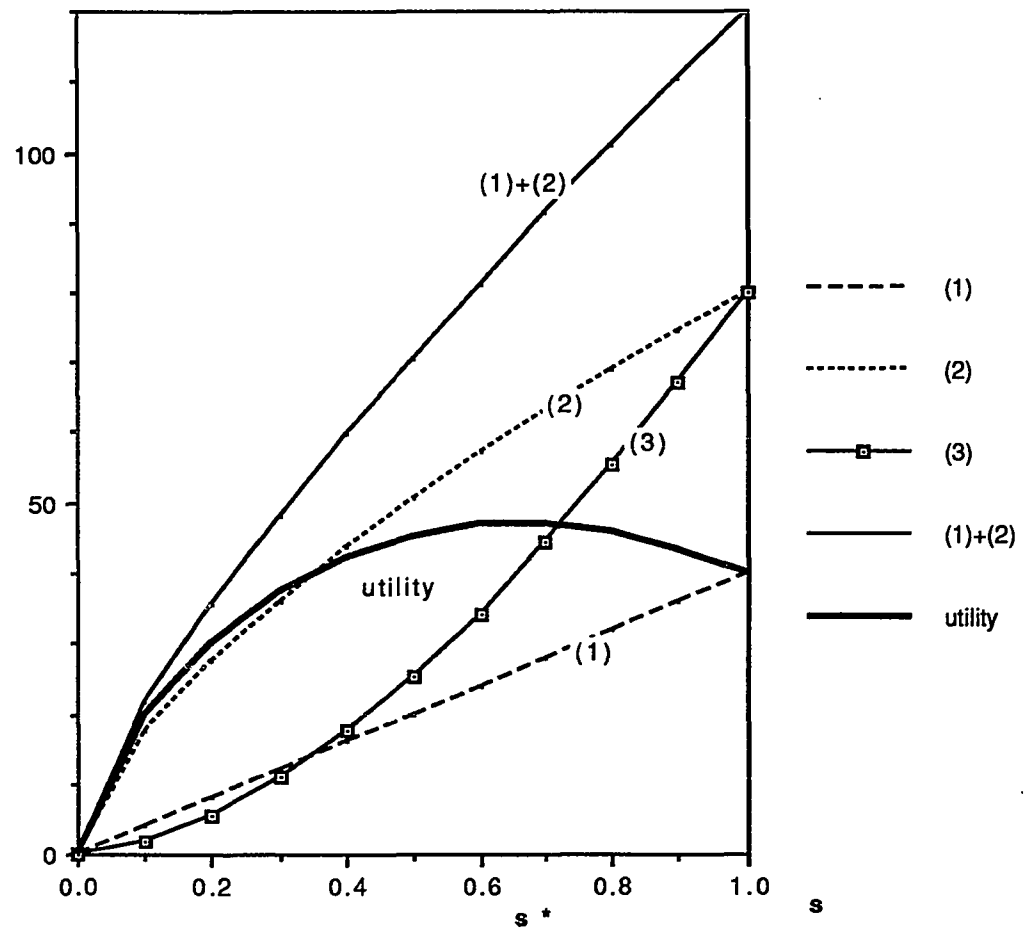
If the agent is risk averse and there is uncertainty, the utility of the agent will equal to the expected utility, EV, minus a risk premium, RP. The risk premium is assumed to be independent of work effort, so it can only shift the utility curve vertically. The utility under the FW system is assumed to be 9 in this figure, if the risk premium is RP', the agent will prefer the optimal work effort, x*, to the minimum work effort; if the risk premium is RP", the agent will prefer the minimum work effort to x*.

Figure 8. Utility of the Principal and Optimal s



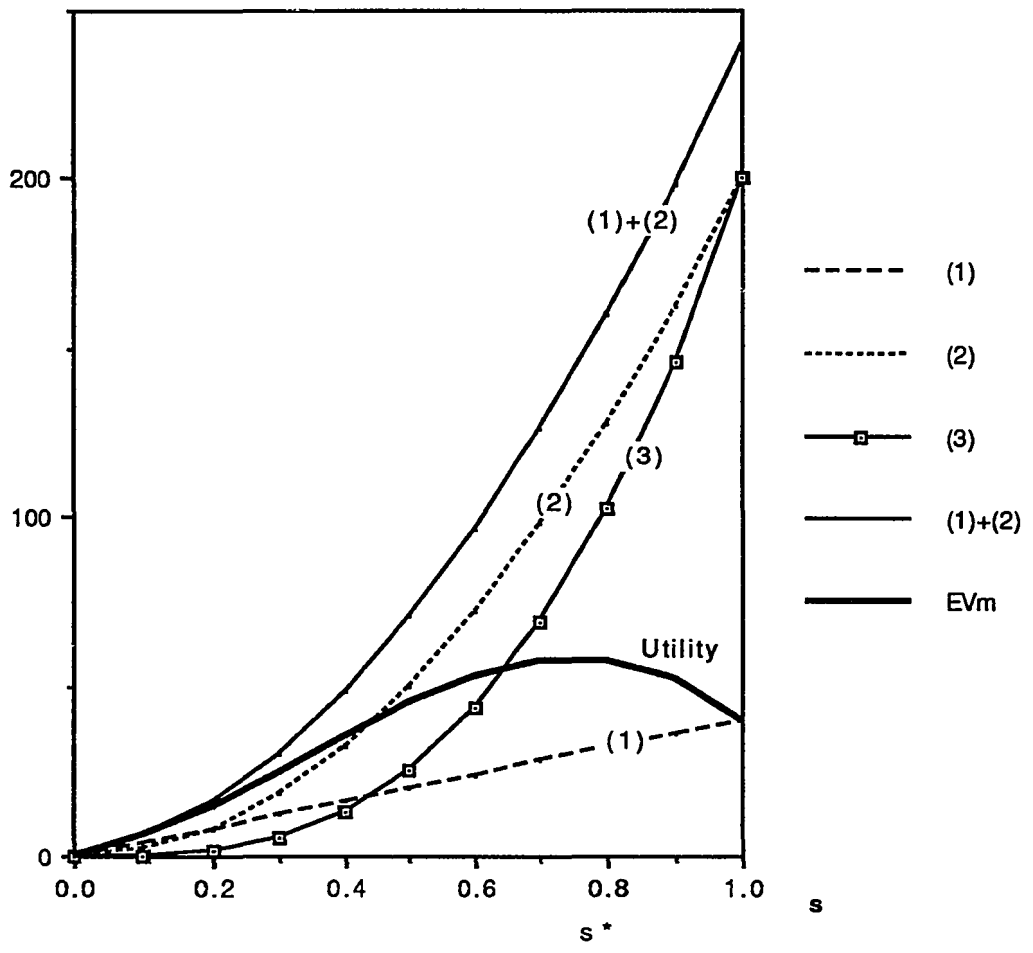
Assume the disutility of work effort is x_i^{β}. When β=2, the utility of the principal can be divided into two items. The first item is linearly related to s. The second item is positively related to s^2. The utility of the principal is maximized at s*. In figure 8, P and x_{i0} are assumed to be 20 and 2 respectively. The optimal s is 0.6.

Figure 9. Utility of the Principal and Optimal s



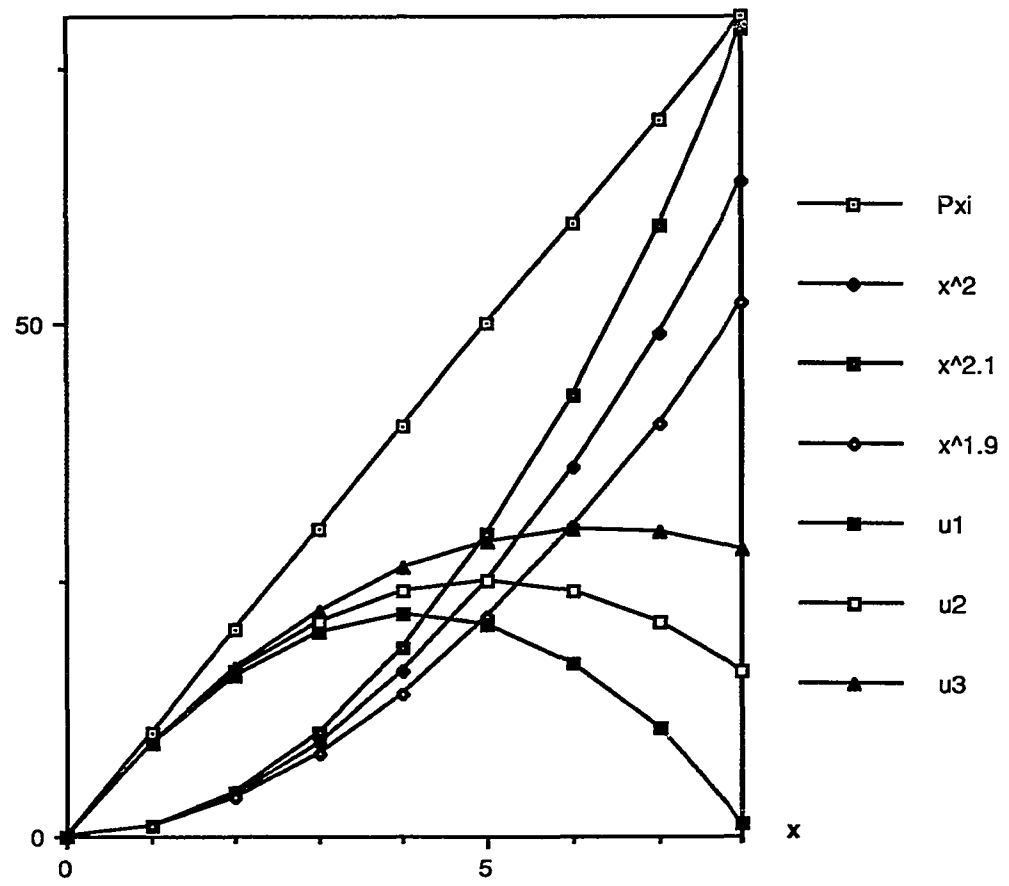
If β is greater than 2 (β=2.5 in this figure), the utility of the principal can be divided into three items. Item (1) is positive and linear with respect to s. Item (2) is also positive but concave with respect to s. Item (3) is negative and convex with respect to s. Utility of the principal is maximized at s*.

Figure 10. Utility of the Principal and Optimal s



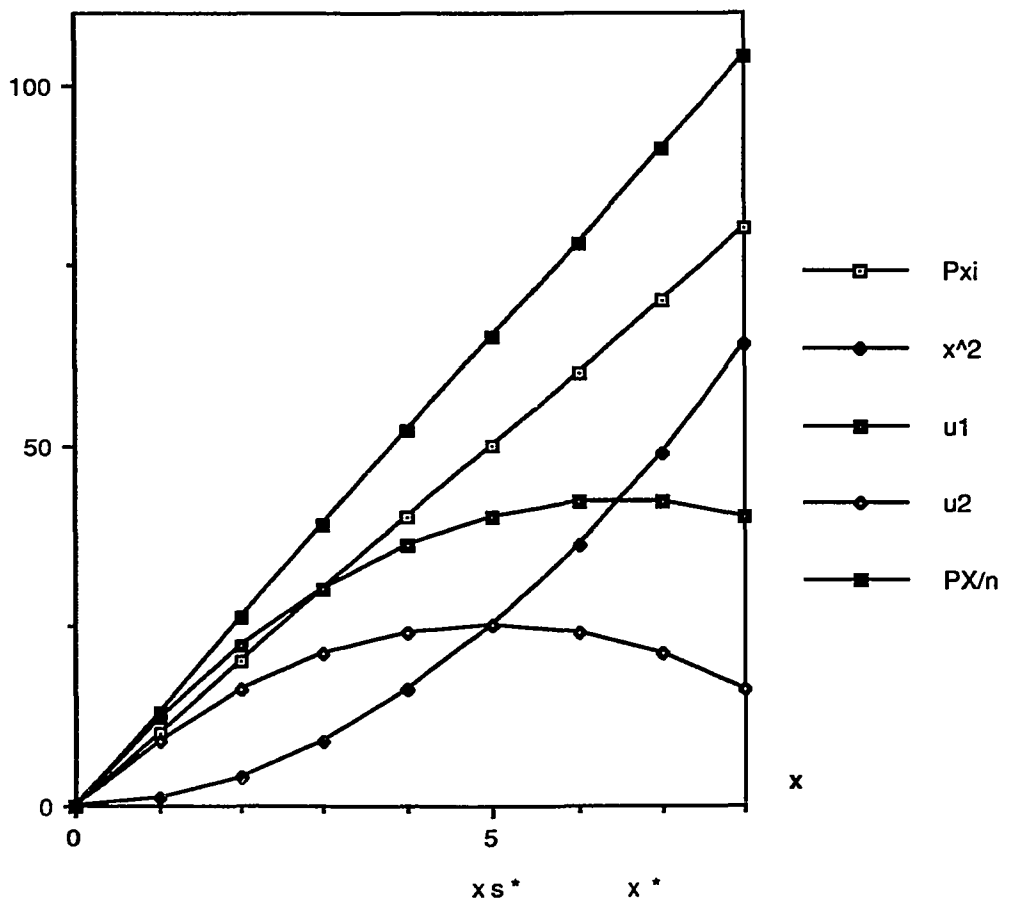
If β is smaller than 2 but greater than 1 (β=1.5 in this figure), the utility of the principal can also be divided into three items. Item (1) is positive and linear with respect to s. Item (2) is positive and convex with respect to s. Item (3) is negative and more convex with respect to s than item (2). Utility of the principal is maximized at s*.

Figure 11. Optimal Efforts of Distinct Agents



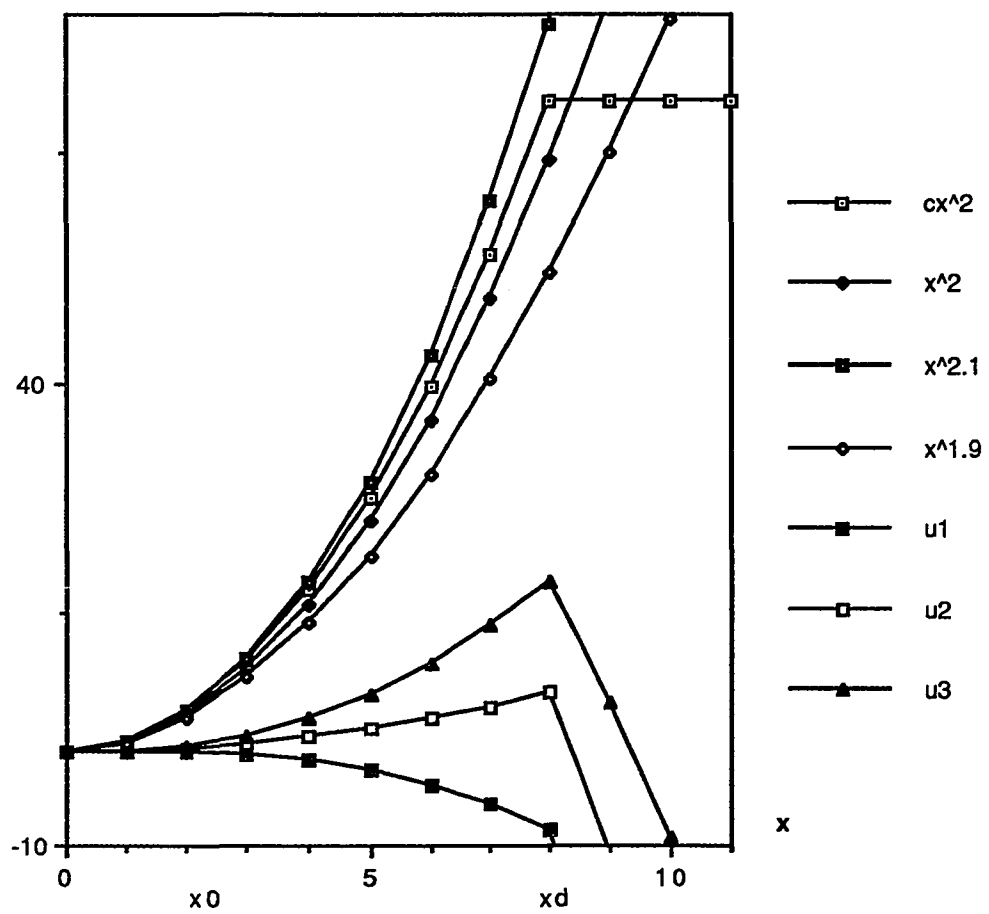
When n=1, the optimal effort is the point where the slop of x_i^{β} equals to the slop of Px_i. Relaxing the assumption that all agents have the same β value, the agent with higher β value will have lower optimal effort, and his socially optimal effort is also lower, the agents would automatically supply the socially optimal work effort in the LST system, even though the principal do not know the β values of each agent.

Figure 12. Optimal and Socially Optimal Efforts (n>1)



If agent i has a higher than average β value, his socially optimal effort which is the point where the slop of x_i^2 (assume β_i=2) equals to the slop of Px_i is lower than the average socially optimal effort. When n>1, the return of work effort is PX/n instead of Px_i. If agent i believes that PX/n is a linear function of x_i, then the optimal effort is determined when the slop of PX/n equals the slop of x_i^2, since PX/n>Px_i, the optimal effort of agent i is greater than his socially optimal effort.

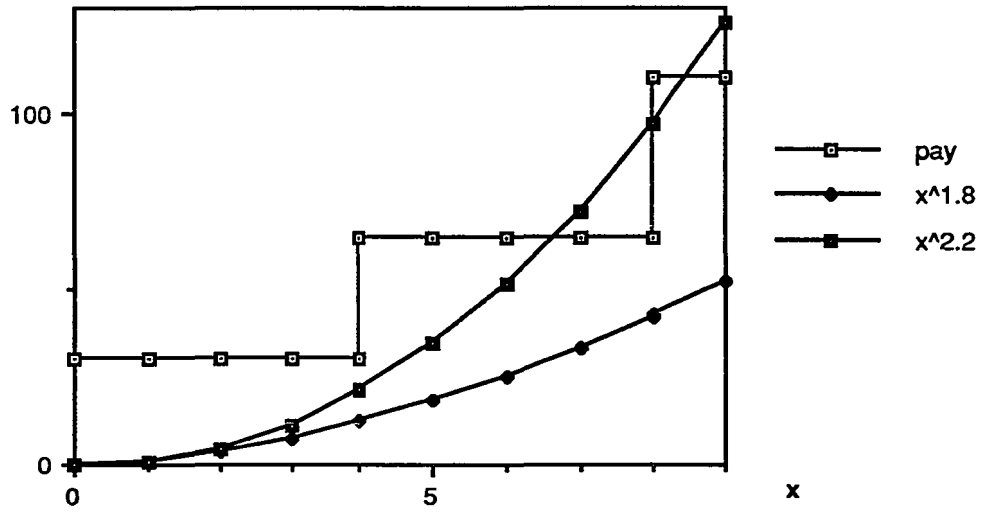
Figure 13. Optimal Efforts of DEC System (Distinct Agents)



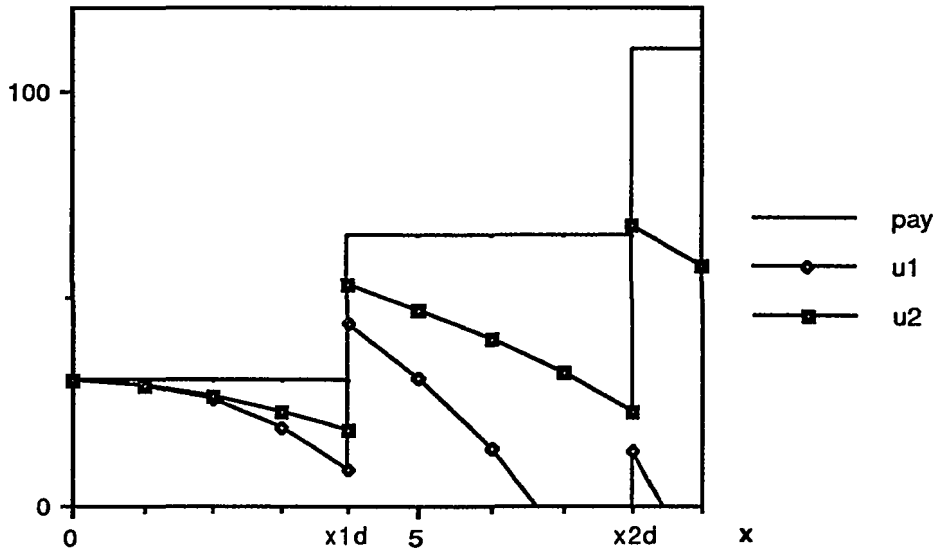
If the principal's estimated β is β_e=2, when actual $\beta_i<\beta_e$ (say β_i=1.9), the optimal effort is x_{id} which is not the socially optimal effort. The socially optimal effort increases as β_i decreases; when $\beta_i>\beta_e$ (say β_i=2.1), the optimal effort is x_{i0}, the minimum work effort.

Figure 14. Optimal Efforts (Discrete Bonuses)

(a)



(b)

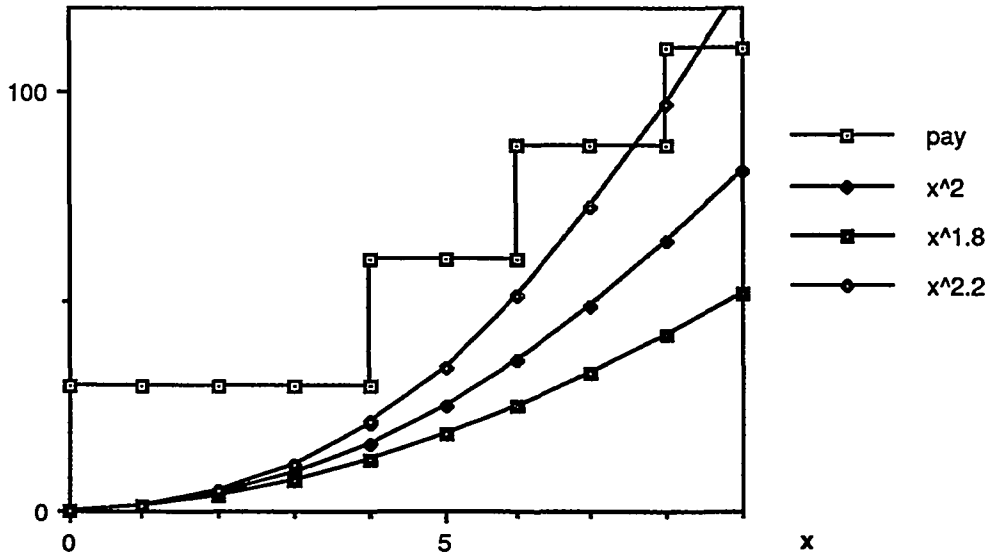


If $\beta_1 > \beta_2$, for a properly designed payment schedule, the optimal effort of agent 1 is x_{1d}; the optimal effort of agent 2 is x_{2d}, (U_1 and U_2 are maximized at x_{id} and x_{2d} respectively) and $x_{id} < x_{2d}$.

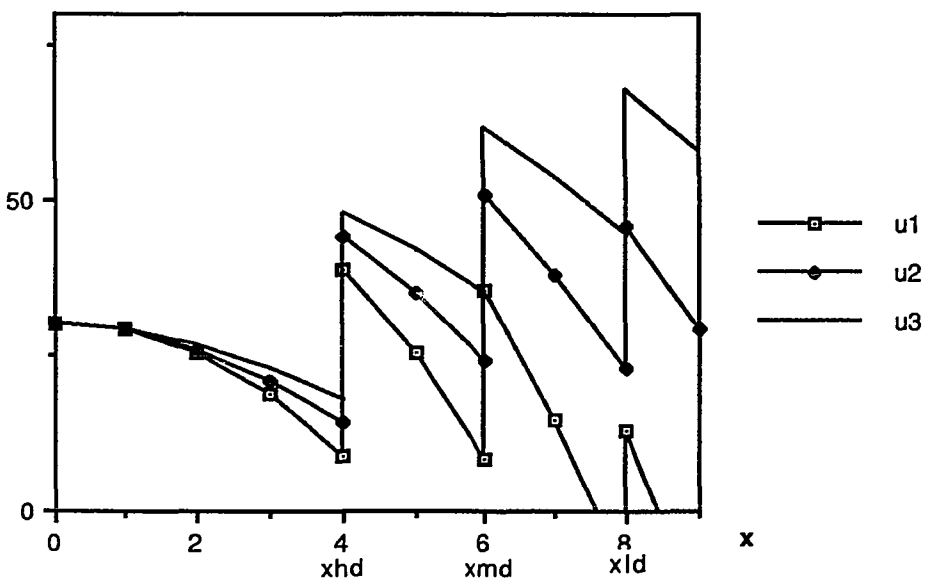
Figure 15. Optimal Efforts (Discrete Bonuses)

(Three Jumps)

(a)



(b)



APPENDIX 2

A MORE GENERAL PS SYSTEM

Up to now, all propositions are developed based on the assumption that the disutility of effort is a quadratic function. Here I relax this assumption and assume that the disutility of effort takes a more general exponential form, x_i^{β}, ($\beta>1$). The reason to assume $\beta>1$ is obvious, in most cases, as effort increases, the extra disutility of effort will become larger and larger. The more general model of the PS system can be written as follows:

$$EV_i=w_0+sP(x_i-x_{i0})-x_i^{\beta} \qquad \text{(a-1)}$$

$$EV_m=n[Px_i-w_0-sP(x_i-x_{i0})] \qquad \text{(a-2)}$$

the optimal effort level under exogenous s can be developed as follows

$$\max_{x_i} EV_i=w_0+sP(x_i-x_{i0})-x_i^{\beta}$$

$$S.T.\ EV(x_i')>EV_i(x_{i0}) \quad \beta>1 \qquad \text{(a-3)}$$

$$F.O.C. \Rightarrow x_i'=\left(\frac{sP}{\beta}\right)^{\frac{1}{\beta-1}} \qquad x_i'=\frac{sP}{2} \quad if\ \beta=2$$

$$\frac{\partial x_i'}{\partial s}=\frac{1}{\beta-1}\left(\frac{P}{\beta}\right)^{\frac{1}{\beta-1}} s^{\frac{2-\beta}{\beta-1}}>0 \qquad \text{(a-4)}$$

$$\frac{\partial x_i'}{\partial P}=\frac{1}{\beta-1}\left(\frac{s}{\beta}\right)^{\frac{1}{\beta-1}} P^{\frac{2-\beta}{\beta-1}}>0 \qquad \text{(a-5)}$$

$$\frac{\partial x_i'}{\partial \beta}=\frac{-sP}{(\beta-1)\beta^2}\left(\frac{sP}{\beta}\right)^{\frac{2-\beta}{\beta-1}}<0 \qquad \text{(a-6)}$$

From (a-4), (a-5), and (a-6) we can see that proposition 1 and 2 still hold in the more general case, and as β increases, x_i' will decrease.

The optimal s can be developed as follows

$$\max_s EV_m=n[(1-s)Px_i'-w_0+sPx_{i0}] \qquad \text{(a-7)}$$
$$F.O.C. \quad \Rightarrow \quad s^{\theta-1}+C=\beta s^{\theta}$$

$$s*=\frac{1}{\beta}+\frac{Cs*^{1-\theta}}{\beta} \qquad \text{(a-8)}$$
$$where \quad \theta=\frac{1}{\beta-1}>0 \quad C=\frac{x_{i0}}{B\theta}>0 \quad B=\left(\frac{P}{\beta}\right)^{\theta}>0$$

if $x_i'<x_{i0}$, then $x_i=x_{i0}$ and $s*=0$; if $x_i'>x_{imax}$, then $x_i=x_{imax}$ and s* can be developed as follows

$$x_{imax}^{\beta}-(x_{imax}-e)^{\beta}=sPe \qquad \text{(a-9)}$$
$$s*=\frac{\beta x_{imax}^{\beta-1}}{P} \qquad as\ e\to 0$$

unfortunately, there is no easy way to solve s* when $x_{i0}<x_i'<x_{imax}$, so we have to use iteration. It is interesting to realize that for whatever value of β, s* is likely greater than 1/2 in general. For example, if P=20, $x_{i0}=2$, and $\beta=1.5$, then s*=0.666; if $\beta=3$, s*=0.78; if $\beta=2$, s*=0.6.

To see why there exists an optimal s in general, we can re-write the expected utility of the principal when $x_i=x_i'$ as follows

$$EV_m = Px_{i0}s + PBs^{\Theta} - PBs^{\Theta+1} \qquad \text{(a-10)}$$

$$\qquad (1) \qquad (2) \qquad (3)$$

where B and Θ are defined as in (a-8) and n is set to 1 for simplicity. In figure 9 and 10, we use three curves to represent the three terms in (a-10). If $\beta>2$ then $\Theta<1$, we can use curve (1) to represent the first term in (a-10) which is linear with respect to s, use curve (2) to represent the second term which is concave since $\Theta<1$, use curve (3) to represent the third term which is convex, and use curve (1)+(2) to represent the sum of term 1 and 2. Curve (1)+(2) is also concave since curve (1) is linear and curve (2) is concave. It's easy to see that EV_m which is the difference between (1)+(2) and (3) takes its maximum value at s* (figure 9). If $\beta<2$, then $\Theta>1$, curve (2) will be convex and so will curve (1)+(2), but curve (3) will be more convex than curve (2) and (1)+(2), so EV_m will also have a maximum value at s* (figure 10).

If $\beta=2$, we return to our special model in chapter 3, and EV_m can be divided into two terms, one is linear with respect to s another is convex with respect to s and $s^*=1/2 + x_{i0}/P$ (figure 8).

Maximizing $Px_i-x_i^{\beta}$ with respect to x_i, the socially optimal effort is

$$x_{is}^* = \left(\frac{P}{\beta}\right)^{\frac{1}{\beta-1}}$$

APPENDIX 3

EXPECTED UTILITY OF THE AGENT UNDER THE PS SYSTEM

In general, the well-being of the agent under the PS system is composed of two parts; work effort has a negative nonrandom effect on agents' well-being, and money income has a positive random effect. The well-being level generated from money income depends, however, on the assumption with respect to the nature of the uncertainty. To simplify the analysis, assume team size n=1, and the subscript i is dropped.

1) Assume value added, Y, takes the following form: $Y=Px+\epsilon$, where P is nonrandom, and ϵ is normally distributed with a zero mean and finite variance σ^2. This is the same assumption made in chapter 3 and in Cauley and Sandler (1991). Under this assumption, the well-being function of the agent is

$$WB=w_0+sP(x-x_0)-x^2+s\epsilon \qquad \text{(a-11)}$$

where WB is the well-being level, other variables are the same as defined in chapter 3.

The mean and variance of WB are

$$MWB=E(WB)=w_0+sP(x-x_0)-x^2 \qquad \text{(a-12)}$$

$$Var(WB)=s^2Var(\epsilon)=s^2\sigma^2 \qquad \text{(a-13)}$$

Assume that the utility function of the agent is twice differentiable, and strictly increasing. Take Taylor expansion of the utility function about MWB we have

$$U(WB)=U(MWB)+U'(MWB)(WB-MWB)+U''(MWB)(WB-MWB)^2/2+.. \qquad \text{(a-14)}$$

The expected utility of the agent is

$$EU(WB)=U(MWB)+U''(MWB)Var(WB)/2 + \ldots \qquad \text{(a-15)}$$

If agents are risk neutral, then $U''=0$, $U'''=0$, ... and the expected utility reduces to

$$EU(WB)=U(MWB) \qquad \text{(a-16)}$$

Because utility functions are invariant with respect to monotonic transformation, we write the expected utility as

$$EV(WB)=MWB=w_0+sP(x-x_0)-x^2 \qquad \text{(a-17)}$$

Because (a-17) is the same as (3-3), the optimal work effort, x', is sP/2, the same as in (3-6).

If agents are risk averse, they maximize the certainty equivalent utility (Laffont, 1989), U(MWB-RP), where RP is the risk premium which satisfies the following equation

$$EV(WB)=U(MWB-RP) \qquad \text{(a-18)}$$

Take Taylor expansion of the certainty equivalent utility about MWB and ignore the higher order terms, we have

$$U(MWB-RP)\approx U(MWB)-RPU'(MWB) \qquad \text{(a-19)}$$

On the other hand, from (a-15) it follows that

$$EV(WB)\approx U(MWB)+U''(MWB)Var(WB)/2 \qquad \text{(a-20)}$$

Substitute (a-19), (a-20), and (a-13) into (a-18), RP can be solved as

$$RP=-\frac{1}{2}\frac{U''s^2\sigma^2}{U'} \qquad \text{(a-21)}$$

If we assume agents possess an exponential utility function with respect to their well-being level

$$U(WB)=C-\exp(-\alpha WB) \quad C=\text{constant} \qquad \text{(a-22)}$$

then $\alpha=-U''/U'$, and $RP=\alpha s^2\sigma^2/2$.

The certainty equivalent utility can be written as

$$U(WBM-RP)=w_0+sP(x-x_0)-x^2-\alpha s^2\sigma^2/2 \qquad \text{(a-23)}$$

The optimal work effort is derived by maximizing (a-23) with respect to x and subject to $U(x)>U(x_0)$.

Because RP is independent of x, the optimal effort, x', is independent of RP in the case of an interior solution, and $x'=sP/2$ is the same as in (3-6). However, if RP is large, such that $U(x')<U(x_0)$, there will be a corner solution, and agents would prefer x_0 to x'.

2) If we assume that $Y=Rx=(P+\epsilon)x$, where R is the random return of work effort, the mean of R is P and the variance of R is σ^2, or ϵ is normally distributed with mean zero and finite variance σ^2. Under this assumption, the well-being function of the agent is

$$WB=w_0+sP(x-x_0)-x^2+sx\epsilon \qquad \text{(a-24)}$$

The mean and variance of WB are

$$MWB=E(WB)=w_0+sP(x-x_0)-x^2 \qquad \text{(a-25)}$$

$$Var(WB)=s^2x^2Var(\epsilon)=s^2x^2\sigma^2 \qquad \text{(a-26)}$$

If the agent is risk neutral, $U''=0$, the expected utility is the same as (a-17) or (3-3). If the agent is risk averse, assume that the utility function of the agent is the same as (a-22), the certainty equivalent utility is

$$U(MWB-RP)=w_0+sP(x-x_0)-x^2-\alpha s^2\sigma^2x^2/2 \qquad \text{(a-27)}$$

where the last term is the risk premium, RP, which is a function of x. The optimal work effort can be derived by

maximizing (a-27) with respect to x, subject to $U(x) \geq U(x_0)$.

In case of interior solution, the F.O.C. implies

$$x' = \frac{sP}{2+\alpha s^2 \sigma^2} < \frac{sP}{2} \qquad \text{(a-28)}$$

x' decreases as α, s, and σ^2 increases. If RP is large, such that $U(x') < U(x_0)$, we end up with a corner solution, and the agent prefers x_0 to x'.

In general, RP has two effects on the optimal work effort; it reduces the optimal work effort in case of interior solution, and it may cause the agent to select a corner solution. The negative effect of risk aversion and uncertainty on optimal effort in case of interior solution is less important, because it only reduce work effort below sP/2, it does not reduce work effort all the way to the minimum level x_0. So our concern is focused on the possibility that risk aversion and uncertainty may drive the optimal choice of the agent to a corner solution, so we assume that $Y=Px+\epsilon$ instead of $Y=(P+\epsilon)x$.

APPENDIX 4

A MODEL WITH DISTINCT AGENTS

In chapter 3, I assumed that agents are identical, and the disutility of work effort is the same for all agents (β=2). In appendix 2, I relaxed the assumption that β=2. Here I relax the assumption that all agents are identical, and assume that the disutility of work effort is different among agents, that is, β_i not equals to β_j in general.

1) The PS and the LST system with n=1. If the team size n=1, relaxing the assumption that $\beta_i=\beta_j$ for all i,j=1, 2, ...n, nothing changes in the PS and the LST systems. We only need substitute β_i for β in all relevant equations. From (a-6) we can see that agents with higher β values would supply lower work efforts, their socially optimal work efforts are also lower. In the LST system, the optimal work efforts of all agents are still the socially optimal work efforts. Even if the principal do not know the β_i values, the agents would automatically supply the socially optimal work effort in the LST system (figure 11).

2) The PS and the LST systems with n>1. To simplify our analysis, I assume there is no free-riding behavior, or the free-riding problem is solved by peer-pressure from co-workers. In this case, because agents' money payment is based on the average work effort of the team, the slop of the utility of income curve become steeper relative to the n=1

case for agents with high β values, and their optimal work efforts are greater than their socially optimal work effort in the LST system. This result can be seen in figure 12. If agent i has a higher than average β value, his optimal effort (also his socially optimal effort in LST system) is lower than average work effort, or $x_i^*<X^*/n$ (a-6). If n=1, x_i^* is the socially optimal effort, x_{is}^*; if n>1, $Px_i^*<PX^*/n$ and $x_i^*>x_{is}^*$. Similarly, the optimal work efforts of agents with lower than average β values are smaller than their socially optimal efforts.

3) The ideal DEC system. In the ideal DEC system of chapter 3, the principal has to estimate the β value. If his estimated β is β_e, the expected utility of agent i is

$$\begin{aligned} EV_i &= w_0 + c(x_i^{\beta_e} - x_{i0}^{\beta_e}) - x_i^{\beta_i} \quad for\ x_i < x_{id} \\ EV_i &= w_0 + c(x_{id}^{\beta_e} - x_{i0}^{\beta_e}) - x_i^{\beta_i} \quad for\ x_i \geq x_{id} \end{aligned} \qquad \text{(a-29)}$$

where x_{id} is the desired value of x_i by the principal.

For the agent with $\beta_i \leq \beta_e$, his optimal effort is x_{id} (see figure 13). Since the socially optimal efforts are different for agents with different β values, the optimal efforts are not socially optimal for β_i not equals to β_e. For the agent with $\beta_i > \beta_e$, since c is greater and close to 1, the optimal effort is the minimum effort, x_{i0}.

4) The DEC system with discrete bonus payments. In the real bonus system, money payment may not be continuous with respect to work effort. If bonus payments are discrete, the

optimal efforts are different for agents with different β_i. This result can be seen in figure 14. In figure 14, I assume there are only two agents, and $\beta_1 > \beta_2$. The principal can design the payment schedule such that the optimal efforts are his desired values. The only thing he need to do is let the payment schedule "jumps" at the desired effort levels. If β_1 and β_2 are known by the principal, he can design the payment schedule such that the optimal efforts are the socially optimal efforts. However, if there are more than two agents, and all have distinct β values (so their socially optimal efforts are different), assume that β_1 is the highest β and β_2 is the lowest β among these agents, then the optimal efforts would be either x_{1d} or x_{2d}, depending whether β_i is closer to β_1 or β_2. Optimal efforts are not socially optimal in general. If the payment schedule includes more "jumps" (see figure 15), there would be more optimal efforts, and the optimal efforts would be closer to the socially optimal values in general.

REFERENCES

Alan B. Krueger and Lawrence H. Summers, 1988. "Efficiency Wages and the Inter-Industry Wage Structure," Econometrica, 56(2), March 1988, PP.259-293.

Alan, Gelb, 1990. "TVP Workers' Incomes, Incentives, and Attitudes" China's Rural Industry, Chapter 13. (New York: Oxford University Press).

Alchian,A.A. & Demsetz, H. 1972."Production, Information Costs, and Economic Organization," American Economic Review, Dec. 777-95. Arrow, K.J. 1971. "Essays in the Theory of Risk-bearing," Markham, Chicago, 1971.

Arrow, 1985. "The Economics of Agency," In Pratt and Zeckhauser, Eds., "Principals and Agents," Boston.

Baron, D. and Myerson, R. 1982. "Regulating a Monopolist with Unknown Costs," Econometrica (July 1982): 911-30.

Becker,G.S. & G.J.Stigler, 1974. "Law enforcement, malfeasance, and compensation of enforcers." Journal of Legal Studies 3, 1-18.

Brown, C. 1990. "Firms' Choice of Method of Pay," Industrial and Labor Relations Review,43, 1990, pp.165-182.

Buchanan, James M. 1975. "Public Finance and Public Choice." National Tax Journal 28, no.4 (December):383-94.

Burns,J.P. 1987."China's nomenclature system." Problems in Communism 36(5), 36-51.

Byrd, William A. 1987. "The Impact of the Two-tier Plan/Market System in Chinese Industry," Journal of Comparative Economics, September 1987, 11, 295-308.

Byrd, William and Gene Tidrick, 1987. "Factor Allocation and Enterprise Incentives," in Tidrick and Chen Jiyuan (eds.), China's Industrial Reform (New York: Oxford University Press).

Byrd, W. 1990. "Plan and Market in the Chinese Economy: a simple general equilibrium model." Journal of Comparative Economics 13 (2), 177-204.

Carmichael, H. 1989. "Self-Enforcing Contracts, Shirking, and Life Cycle Incentives,"Journal of Economic Perspectives, 3(4), 65-83.

Cauley Jon and Sandler Todd, 1991. "Agency thoery and the Chinese enterprise under reform" Working paper. Department of Economics, University of Hawaii-Hilo.

Chen, C. Y. 1984. "Fundamentals of Thoughts and Political Education Work in Enterprises. Beijing, Enterprise Management Press (in Chinese).

Chen, Kuan et al. 1988. "Productivity Change in Chinese Industry: 1953-1985," Journal of Comparative Economics, Dec. 12, 570-591.

Chen, Yanling, 1990. "A concealed income sustaining early-maturing consumption in China." Economic Research, No.2, 1990.

Chen, Yizi & Wang, Xiaoqian, 1987. "Reform: Results and Lessons from the 1985 CESRRI Survey," Journal of Comparative Economics, Sep. 1987, 11, 462-478.

Cheung, 1968. "Private property rights and Sharecropping." Journal of Political Economy 76, 1107-22.

Cheung, 1969. "Transaction Costs, Risk Aversion, and the Choice of Contractual Arrangements," Journal of Law & Economics (April 1969).

Cheung, Steven N., 1990. "Privatization vs. Special Interests: The Experience of China's Economic Reform," in James A. Dorn and Wang Xi Eds., Economic Reform in China: Problems and Prospects (Chicago: University of Chicago Press).

Chow, Gregory C., 1985. "The Chinese economy" (Harper and Row, New York).

Chuan, Zhehua, et al. 1982. "Democratic Management Enhances More Vitality to the Industrial Enterprises". Shehui Kexue (Social Science, Shanghai), No.2 (in Chinese).

Conlisk, 1989. "Three Variants on the Allais Example," American Economic Review 79 (June 1989): 392-407.

David M. Kreps, 1985. "A Course in Microeconomic theory" pp503-515. Princeton University press.

Dollar, David. "Economic Reform and Allocative Efficiency in China's State-owned Industry," Forthcoming, Economic Development and Cultural Change.

Dubey, P. 1983. "Inefficiency of Nash equilibria." Mimeo.

Du, Haiyan et al, 1990. "Autonomous rights of state-owned enterprises, market structure and incentive systems--investigation and analysis of 403 state enterprises." Economic Research, Jan. 1990.

Encyclopedia of China's Industrial Management 1984. Beijing, Enterprise Management Press, Vols. 1 and 2 (in Chinese).

Eswaran and Kotwal, 1985. " A Theory of Contractual Structure in Agriculture," American Economic Review (June 1985): 352-367.

Fama,E.F. 1980. "Agency problems and the theoty of the firm." Journal of Political Economy 8, 288-307.

Fei, J. and Reynolds, B. 1988. "A tentative plan for the rational sequencing of overall reform in China's economic system." Journal of Comparative Economics 11(3), 490-502.

Feng, Lanrui & Zhao, Lukuan 1984. "Urban Employment and Wages", in China's Socialist Modernization. Yu Guangyuan (eds.), Foreign Languages Press, Beijing.

Firchau, V. 1987. "Information systems for principal-agent problems." In Bamberg, G. and Spremann, K. Eds. "Agency Theory, Information,and Incentives," pp.81-92. Berlin: Springer-verlag.

Fishburn, 1989. "Retrospective on the Utility Theory of Von Neumann and Morgenstern," Journal of Risk and Uncertainty 2 (June 1989): 127-58.

Fureng, D. 1987. "Increasing the vitality of enterprises." In G. Tidrick and J. Chen, Eds. "China's Industrial Reform," pp.44-59. New York: Oxford University Press.

Gardner, R., E. Ostrom & J.M. Walker. 1990, "The Nature of Common-Pool Resource Problems," Rationality and Society, Vol. 2, No.3 (Jule,1990) pp. 335-358.

Gao, Shangquan, 1987. "Nine Years of Chinese Economic System Reforms, [in Chinese], Beijing: People's Publishing House.

Gjesdal, F. 1982. "Information and incentives: the agency information problem." Review of Econimic Studies 64, 373-64.

Granick, D., 1987. "The industrial environment in China and the CMEA countries," in G. Tidrick and J. Chen, eds., op cit.

Green, J.R. and N.L. Stokey, 1983. "A comparison of tournaments and contracts." Journal of Political Economy 91, 394-64.

Grossman, S.J. and O.D. Hart, 1983. "An analysis of the principal-agent problem." Econometrica 51, 7-45.

Groves, T. 1973. "Incentives in teams," Econometrica 41, 617-31.

Guo, Fei, 1989. "New research on the objective model of salary reformation in the state-owned firm" Economic Research, Nov. 1989.

Harberger, Arnold C. 1984. "Basic Needs Versus Distributional Weights in Social Cost-Benefit Analysis," Economic Development and Cultural Change, Feb. 1984, 32, 455-474.

Harper, Paul. 1971. "Workers' Participation in management in Communist China". Studies in Comparative Communism, July-October, pp. 111-40.

Harris, M. and A. Raviv, 1979. "Optimal incentive contracts with imperfect information." Journal of Economic Theory 20, 231-59.

Harris, M. and R.M. Townsend, 1981. "Resource allocation under asymmetric information." Econometrica 49, 33-64.

Harsanyi, J.C. 1967. "Games of incomplete information played by 'Bayesian' Players" Part I. Management Studies 14, 159-82.

Hart, O. D. 1988. "Incomplete Contracts and the Theory of the firm," Journal of Law, Economics, and Organization, Spring 1988, 4(1) 119-139.

Hashimoto, Masanori, 1990. "The Japanese Labor Market: in a comparative perspective with the United States" W.E. UPJOHN Institute.

Henley, J. S. and Nyaw, M. K. 1986. "The Development in Managerial Decision Making in Chinese Industrial Enterprises", in Stewart Clegg and Gordon Redding eds. "The Enterprise and Management in East Asia. Centre of Asian Studies, University of Hong Kong.

Henley, J. S. and Nyaw, M. K. 1987. "The Development of Work Incentives in Chinese Industrial Enterprises--Material Versus Non-Material Incentives." In Warner, M., Eds., "Management Reforms in China," 127-148. London: Frances Printer.

Holmstrom, B. 1978. "Moral hazard and observability" Discussion paper No.7816, CORE, April 1978; Bell J. Econ., in press.

Holmstrom, B. 1979. "On incentives and control in organizations," Bell Journal of Economics 10, 74-91.

Holmstrom, B. 1982. "Moral hazard in teams." Bell Journal of Economics 13(2), 324-340.

Holmstrom, B. 1983. "Managerial incentive problems - a dynamic perspective." Mimeo.

Holyer, H. 1990. "Wages, Employer Costs, and Employee Performance in the Firm," Industrial and Labor Relations Review, 43, 1990, pp. 147-164.

Hong, Ng Sek, and Lansbury, Russell D. "The Workers' Congress in Chinese Enterprises." In M. Warner (eds.), Management Reforms in China. London: Frances Printer, 1987.

Hua Sheng et al., 1987. World Economic Herald 1/19/87.

Hutchens, R. 1989. " Seniority, Wages and Productivity: A Turbulent Decade,"Journal of Economic Perspectives,3-(4), 1989, pp.49-64.

Jefferson, Gary. 1989. "Potential Sources of Productivity Growth within Chinese Industry," World Development 17(1).

Jefferson, Gary. 1989. "Book Review: Bruce Reynolds Eds, Reform in China: Challenges and Choices," Journal of Comparative Economics, 1989b, 13, 354-356.

Jensen, M. C. & Meckling, W. H. 1976. "Theoty of the firm: Managerial Behavior, Agency Costs and Ownership Structure," Journal of Financial Economics 1976, 3, 305-360.

Jiang, Yiwei & Shen, Hungshen 1985. "Industrial Enterprise Management". Beijing, Economic Management Press (in Chinese).

Jin, Qi 1984. "New Bonus System Lifts Limits." Beijing Review, 27, No. 26, 25 June.

Johnson, D.G. 1950. "Resource allocation under share contracts." Journal of Political Economy 58, 111-23.

Kahn, L. and Sheur, P. 1990. "Contingent Pay and Managerial Performance," Industrial and Labor Pelations Review,43, 107-20.

Katz, L. 1988. "Some Recent Development in Labor Economics and their Implications for Macroeconomics," Journal of Money, Credit, and Banking, 20(3), 1988, pp.507-522.

Klein, Benjamin. 1983. "Contracting Costs and Residual Claims: The Separation of Ownership and Control," Journal of Law and Economics, Oct. 1983, 26(2), 367-374.

Klein, B. et al. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," Journal of Law and Economics, Oct. 1978, 297-326.

Kornai, Janos. 1985. "The Dual-Dependency of State Industry," Economic Research [Chinese Journal] 10, 1985.

Kornai, Janos. 1986. "Contradictions and Dilemmas: Studies on the Socialist Economy and Society, Cambridge, Mass.: The MIT Press.

Korzec, M. 1988. "Efficiency wages and enterprise behavior in China." The Journal of Communist Studies 3(1), pp.3-26.

Krueger, A. & Summers, L. 1988. "Efficiency Wages and the Inter-Industry Wage Structure," Econometrica, 56(2), March, pp.259-293.

Laffont and Tirole, 1986. "Using Cost Observation to Regulate Firms," Journal of Political Economy (June 1986): 614-41.

Laffont, 1989. "The Economics of Uncertainty and Information," Ch. 2.

Lal, D. 1990. "The fable of the three envelopes: the analytics and political economy of the reform of Chinese State owned enterprises." European Economic Review 34(6), 1213-1231.

Lal, R. and R. Staelin, 1981. "A theory of optimal salesforce compensation plans." Mimeo.

Lambert,R.A. 1983. "Long term contracts and moral hazard." Bell Journal of Economics 14, 441-52.

Lazear, E.P. 1979. "Why is there mandatory retirement?" Journal of Political Economy 87, 1261-84.

Lazear, E.P. and S. Rosen. 1981. "Rank-order tournaments as optimum labor contracts." Journal of Political Economy 89, 841-64.

Lazear, 1986. "Salaries and Piece Rates," Journal of Business 59 (1986): 405-431.

Lazear and Rosen, 1981. "Rank-Order Tournaments as Optimum Labor Contracts," Journal of Political Economy 89 (Oct.): 841-864.

Lee, K. 1990. "The Chinese Model of the Socialist Enterprise: an Assessment of its Organization and Performance." Journal of Comparative Economics 14(3), pp.384-400.

Lee,K. and Mark, S. 1990. "Socialist Egalitarianism and the Agency Problem." Seoul Journal of Economics 2(4), pp.383-401.

Li Yining, 1986. Beijing Daily, 5/19/86.

Lin, Justin, Y. 1987. "The Household responsibility System Reform in China: A peasant's Institutional Choice," American Journal of Agricultural Economics, May 1987, 69(2), 410-415.

Liu, Xiaoxuan, 1990. "Analyze the effects of income incentives in enterprises". Economic Research, No.5, 1990.

MacDonald, G. 1984. "New Directions in the Economic Theory of Agency." Canadian Journal of Economics 17(3), 415-440.

Malueg, D. 1981. "Pareto optimality of repeated insurance contracts when there are probliems of moral hazard and adverse selection." Mimeo.

Manion, "The cadre management system post-Mao" Loc. cit., p205.

Marshall, A. 1920. "Principles of Economics, eighth ed." New York: Macmillan.

Mirrlees, J. 1974. "Notes on welfare economics, information and uncertainty." In Balch et al., eds, Essays in Economic Behavior Under Uncertainty (Amsterdam: North-Holland).

Mirrlees, J. 1975. "The theory of moral hazard and unobservable behaviour-Part I." Nuffield College, Oxford.

Mirrlees J. 1976. "The optimal Structure of incentives and authority within an organization." Bell Journal of Economics 7, 105-31.

Mueller, Dennis C. 1976. "Public Choice: a Survey." Journal of Economic Literature 14, no.2(June): 395-433).

Murrell, 1983. "The Economics of Sharing: A Transactions Cost Analysis of Contractual Choice in Farming," Bell Journal of Economics (Spring 1983): 283-93.

Myerson, R.B. 1979. "Incentive compatibility and the bargaining problem." Econometrica 47, 61-74.

Nalebuff, B. and J.E. Stiglitz, 1983. "Prizes and incentives: towards a general theoty of compensation and competition." Bell journal of Economics 14, 21-43.

Niskanen, William A. 1968. "Nonmarket Decision Making: the Peculiar Economics of Bureaucracy." American Economic Review 68, no.2 (May): 293-305.

Niu, Yuesheng et al, 1991. "Some considerations of perfecting the individual income distribution system" Economic Research, Jan. 1991.

Nove, A., 1977. "The Soviet economic system" (Allen and Unwin, london).

O'Keefe,M., W.K. Viscusi, and R.J. Zeckhauser. 1984. "Economic contests:comparative reward schemes." Journal of Labor Economics 2,27-56.

Olsen, M., 1965, "The Logic of Collective Action, Cambridge, MA: Harvard University Press, (1965).

Otsuka, K. & Hayami, Y. 1989. "Theories of Share Tenancy: A Critical Survey," Economic Development and Cultural Change, 37(1).

Pan, Jinyun et al, 1990. China Industrial Economic Research, No.3, 1990. pp49-54 (In Chinese).

Perkins, D. 1988. "Reforming China's economic system." Journal of Economic Literature 26(2), 601-645.

Pindyck. Robert S. & Daniel L. Rubinfeld, 1981. "Econometric Models and Economic Forecasts" McGraw-Hill Book Company.

Pratt, 1964. "Risk Aversion in the Small and in the Large," Econometrica 32 (1964): 122-36

Putterman, Louis, 1987. "The Incentive Problem and the Demise of Team Farming in China," Journal of Development Economics 26.

Radner, R. 1981. "Monitoring cooprative agreement in a repeated principal-agent relationship." Econometrica 49, 1127-48.

Rasmusen, C. 1987. "Moral hazard in risk -averse teams." Rand Journal of Economics 18(3), 428-435.

Reynolds, Bruce, Eds. 1987. "Reform in China: Challenges and Choices, Armonk, NY: M.E.Sharpe.

Ricketts, M. 1984. "Principal and agent, risk sharing and incentives: a simple diagrammatic exposition." Mimeo.

Riskin, Carl 1973. "Maoism and Motivation: Work incentive in China". Bulletin of Concerned Asian Scholars, july: 10-24.

Rogerson, W. P. 1982. "The role of reputation in a repeated agency problem involving information transmission." Institute for Mathematical Studies in the Social Sciences, Technical Report #377.

Rogerson,W.P. 1983. "The first-order approach to principal-agent problems." Mimeo.

Rosen, S. 1982. "Authority, control and the distribution of earnings." Bell Journal of Economics 13,311-23.

Rosen, 1986. "Prizes and Incentives in Elimination Tournaments," American Economic Review 76 (Sept. 1986): 701-715.

Ross, S. A. 1973. "The economic theory of agency: the principal's problem." American Economic Review 63(2), 134-139.

Rothschild, 1973. "Models of Market Organization with Imperfect Information: A Survey," Journal of Political Economy: 1, 283-1, 308.

Roumasset, James A. 1979. "Sharecropping, Production Externalities, and the Theory of Contracts," American Journal of Agricultural Economics (Nov. 1979): 640-7.

Roumasset, James A. and Laine, Charles. 1991, "A Constitutional Approach to Free Riding and Common Property Resource Management" Working paper. Department of Economics, Univercity of Hawaii.

Rubenstein,A. and M.E. Yaari, 1983. "Repeated insurance contracts and moral hazard." Journal of Economic Theory 30,74-97.

Schenck, K. 1988. "New Institutional Dimensions of Economics." New York: Springer-Verlag.

Shavell, S. 1979. "Risk sharing and incentives in the principal agent relationship." Bell Journal of Economics 10, 55-73.

Shirk, Susan L. 1981. "Recent Chinese labour policies and the transformation of industrial organization in China," The China Quarterly, No. 88. (December 1981) pp 584-90.

Shubik, M. 1982. "Game Theory in the Social Sciences (Cambridge, MA: M.I.T. Press.)

Simon, H. 1965. "Administrative Behavior: A Study of Decision Processes in Administrative Organizations, second ed. (New York: Modern Library).

Spence,A.M. and R.J. Zeckhauser 1971. "Insurance, information and individual action." American Economic Review 61, 380-91.

Spremann, K. 1987. "Agency Theory and Risk Sharing." In Bamberg, G. and Spremann, K Eds. "Agency Theory, Information, and Incentives, PP. 3-37. Berlin: Springer-Verlag.

State Economic Planning Commission 1984. "Ideological and Political Work in Industrial Enterprises". Beijing, Economic Science Publication (in Chinese).

Stiglitz, J.E. 1974. "Incentive and risk sharing in sharecropping." Review of Economic Studies 41, 219-55.

Stiglitz, J.E. 1975. "Incentives, risk, and information: notes toward a theory of hierarchy," Bell J. Econ. 6(1975), 552-579.

Stiglitz, J. 1987. "Risk Sharing and Incentives in Sharecropping." Review of Economic Studies 61(2) 219-256.

Tang, Hongqian & Guo Xiaoming, 1991. "On improving the family contract system in China's agriculture" Economic Research, Feb, 1991.

Tang, Z. 1987. "Supply and Marketing." In Tidrick, G. and Chen, J. Eds., "China's Industrial Reforms," PP.210-236. New York: Oxford University Press.

Tian, Jiyun 1986. " On the Present Economic Situation and Restructuring the Economy". (A speech made by Vice-Premier Tian Jiyun on 6 January, at a meeting of cadres from central organization). Beijing Review, 29, Nos. 6 & 7, 10 February.

Tidrick, G. 1987. "Planning and Supply." In Tidrick, G. and Chen, J. Eds., "China's Industrial Reforms," pp.175-209. New York: Oxford University Press.

Tirole, J. 1986. "Hierarchies and Bureaucracies: on the Role of Collusion in Organizations." Journal of Law, Economics, and Organizations 2(2), 181-214.

Townsend, R. M. 1976. "Efficient contracts with costly state verification," GSIA working paper, Carnegie-Mellon University, 1976.

Tong, Yujing et al, 1991. "Some problems in the payment system which relates wage to performance". Research on Economics and Management, No.1, 1991.

Tuck, R. H. 1954. "An essay on the economic theory of rank" (Oxford: Basil Blackwell).

Tung, Rosalin L. 1981. " Patterns of Motivation in Chinese Industrial Enterprises". Academy of Management Review, 6, No.3: 481-9.

Varian, Hal R. 1984. "Microeconomic Analysis" chapter 1. W.W. Norton & Company New York. Landon.

Vepa,Ram K. 1979. "Mao's China" Abhinav Publications New Delhi.

Walder, A. G. 1986. "Communist Neo-Traditionalism." Berkeley: University of California Press.

Walder, A. G. 1987. "Wage Reform and the Web of Factory Interests" The China Quarterly 109(1), 22-41.

Walder, A. G. 1989. "Factory and Manager in an Era of Reform" The China Quarterly, No. 118, June, 1989.

Wang Dingding, 1990. "An Inquiry into the Nature of Producer's Behavior in a Reforming Economy." Dissertation in the University of Hawaii.

Wang, Xuefeng, 1990. "More on reduction of share of the financial income to the national income" Economic Research, No.4, 1990.

Weitzman, Martin L. 1984. "The Share Economy" Harvard University Press. Combridge, Massachusetts and London, England.

Weitzman, Martin L. and Kruse, Douglas L, 1990. "Profit Sharing and Productivity" In Paying for Productivity: A Look at the Evidence, ed. Alan S. Blinder. Washington DC.: Brookings Institution.

Wiemer, Calla, 1990. "Price Reform Stalled: An Inherent Obstacle, a Missed Opportunity," Journal of Asian Economics 1(2).

Wiemer, Calla, 1991. Paper prepared for Modern China symposium on "Institutional Segmentation, Structural Change, and Economic Reform in China."

Williamson, O.E. 1967. "Hierarchical control and optimum firm size." Journal of Political Economy 75, 123-38.

Williamson, O. E. 1988. "Enterprise Reform in China: a Transaction Cost Economics Perspective," Mimeo.

Wilson, R. 1978. "Information, efficiency and the core of an economy." Econometrica 46,807-16.

Winiecki, J. 1990. "Why Economic Reform Fail in the Soviet System--a Property-Rights Based Approach." Economic Inquiry 28(2), 195-221.

Wong, T. 1989. "The Salary Structure, Allowances and Benefits of a Shanghai Electronics Factory." The China Quarterly 117(1), 135-144.

Wu, J. and Reynolds, B. 1988. "Choosing a Strategy for China's Economic Reform." American Economic Review 78(2), 461-466.

Xiang, Qiyuan, 1982. "Economic Development and Income Distribution", in China's Search for Economic Growth, Xu Dixin et al. Beijing, New World Press.

Xiao, Geng 1990. "The Impact of Property Rights Structure on Productivity, Capital Allocation and Labor Income in Chinese State and Collective Enterprises," Manuscript, September 1990a.

Xiao, Geng 1990. "Managerial Autonomy, Fringe Benefits, and Ownership Structure" Forthcoming in China Economic Review, Dec.

Xin, Changxing & Pan, Dapeng, 1989. "Inflation of consume demand and reformation of salary" Economic Research, Nov. 1989.

Xu, Liencang & Lin, Wenchuan, 1980. "A Psychological Study of the Workers' Ideological Dynamics". Kuanming Daily, 22. August.

Xun, Dazhi, 1990. "The contracted system: difficulty and outlet--reformation of enterprise management mechenism during the period of adjustment." Economic Research, No.3, 1990.

Yang, Guansan et al, 1987. "Enterprise Cadres and Reform," in Reynolds, Reform in China, p83.

Yang, peixin, 1990. "On sticking to and profecting the contracted system". Economic Research, No.3, 1990.

Yu, Guangyuan 1985. "Reform During Socialist Construction in terms of World and Chinese History." Social Sciences in China, 3: 23-38.

Yuan, Engzhen & Gu, Guangqing, 1991. "On some problems in the economic reform theory". Economic Research, Feb. 1991.

Zhang, Fan et al, 1991. "The change of the basic structure of national economic distribution and its adjusting orientation". China Industrial Economic Research, No.1, 1991.

Zhang, shaojie & Zhang Amei. 1987. "The Present Management Environment in China's Industrial Enterprises," In Reynolds, Bruce, Eds. "Reform in China: Challenges and Choices, Armonk, NY: M. E. Sharper, 1987.

Zhou, Liqun, 1989. "Stimulation and restrain--on the transformation of enterprise function mechanism" Economic Research, Sept.1989.

Zhuang, S. C. et al. 1991, "American and Chinese Attitudes Towards Profit and Loss" Working paper. Center for Chinese Studies, Contemporary China Seminar, UHM.

Zhu, Ping, 1990. Research on Economics & Management, No.5, 1990, p46. (In Chinese).