An Empirical Analysis of a Relative Performance-Based Incentive Plan: Evidence from a Postal Service

Ella Mae Matsumura*
and
Jae Yong Shin

Department of Accounting and Information Systems
School of Business
University of Wisconsin-Madison
975 University Avenue
Madison, WI 53706-1323

Current Draft: March 15, 2004

*Corresponding author. Email: ematsumura@bus.wisc.edu

We are grateful to Korea Post for allowing us access to the data. We are especially grateful to Joong-Bum Choi at Korea Information Strategy Development Institute (KISDI) and Ki-Hwan Hong at Korea Post for providing us with institutional insights and helpful suggestions. We also thank Steve Hansen, Bill Lanen, participants at 2004 American Accounting Association Management Accounting Section Research Conference, and workshop participants at the University of Wisconsin-Madison and the University of Wisconsin-Milwaukee for their helpful comments.

An Empirical Analysis of a Relative Performance-Based Incentive Plan: Evidence from a Postal Service

Abstract: Using annual performance evaluation data of Korea Post's 214 stores during 1997-1999, we find that the introduction of a relative performance evaluation (RPE)-based incentive plan is positively associated with financial performance and that under the new incentive plan, the degree of common uncertainty is positively associated with store profitability. These results are consistent with Frederickson's (1992) experimental results that agents' effort levels are higher with an RPE contract than with a profit-sharing contract, and that effort increased with the degree of common uncertainty under an RPE contract. We also find evidence that the incentive effect of the RPE-based incentive plan is mitigated in stores at which the level of dysfunctional behavior is high. Finally, we find that the net benefits of introducing the RPE contract may be highly conditional on the degree of common uncertainty, which appears to be inversely related to the likelihood of observing dysfunctional behavior that could be caused by the RPE-based contract.

Key Words: Relative performance evaluation, incentive plans, financial performance, dysfunctional behavior, common uncertainty

Data Availability: Data in this study are available from the authors upon request.

1. INTRODUCTION

Relative performance evaluation (RPE hereafter) entails evaluating individual or organizational unit performance relative to the performance of others. Justification for RPE can be found in economics and behavioral theories. From an economics perspective, RPE provides benefits by deriving information about agents' efforts when the agents face common uncertainty, such as common market conditions. RPE then allows a principal to establish contracts that impose less risk on the agents while motivating the desired effort, and thus develop Pareto superior contracts. Behavioral theories highlight intrinsic (nonmonetary) influences on effort. Although RPE offers potential benefits, it may motivate undesirable behavior, such as sabotaging coworkers, and colluding with coworkers and shirking (Gibbons and Murphy 1990).

Our study uses data from the entire population of 214 post offices (hereafter, stores) managed by Korea Post, which introduced an RPE-based incentive plan in 1998. We examine 1) whether including relative performance-based measures in an incentive plan improves financial performance at the store level (RPE's incentive effect), 2) whether dysfunctional behavioral responses[1] caused by introducing RPE mitigate an RPE-based incentive plan's performance impact, and 3) the role of the degree of common uncertainty in driving store profitability and predicting at which stores the level of dysfunctional behavior are likely to be high.

Due to data availability constraints, prior empirical literature on RPE has focused on the compensation of executives. Moreover, the extant literature has been limited to investigating whether top executives are compensated as if their performance is evaluated

[1] An example of a dysfunctional behavioral response in this context is reduced productivity due to low employee morale when employees felt they could not possibly achieve the RPE bonus.

relative to the performance of competitors (Antle and Smith 1986, Gibbons and Murphy 1990, Janakiraman et al. 1992, Anderson et al. 2001). Overall, little empirical evidence is available on the performance consequences of RPE when there is an important source of common risk. An exception is Frederickson (1992), who provides experimental evidence that agents' effort levels are higher with an RPE contract than with a profit-sharing contract where an agent's compensation depends solely on the absolute performance of that agent. Moreover, there has been no study that empirically documents whether dysfunctional behavior can mitigate an RPE-based contract's incentive effects (Prendergast 1999).

Our study contributes to the RPE literature by providing empirical evidence on RPE's performance impact when agents face varying degrees of common uncertainty. Furthermore, by examining whether RPE's incentive effect is moderated in the sub-sample of stores where the behavioral responses are expected to be highly dysfunctional, this study contributes to a growing body of literature that investigates the effect of compensation policies on performance and dysfunctional responses to compensation schemes. Our study also extends the literature on the role of the degree of common uncertainty (Frederickson, 1992) by demonstrating the importance of the degree of common uncertainty in determining the effectiveness of RPE contracts. In addition, this study contributes to the budgeting literature by providing the first empirical evidence on the impacts of benchmarked performance standards and their motivational effects (Hansen et al. 2003). Finally, by using the data from a state-owned company that is in a quasi-governmental sector, our study contributes to the literature that examines the role of performance measurement systems in improving the efficiency and effectiveness of government operations (Ittner and Larcker 1998).

Our results indicate that the introduction of the research site's RPE incentive plan is

positively associated with store-level financial performance and that under the new incentive plan, a store's degree of common uncertainty is positively associated with profitability. Furthermore, we find evidence that the RPE incentive plan's impact on financial performance is lower in stores with expected high dysfunctional behavior. Finally, we find that a store's degree of common uncertainty is negatively associated with the probability that the store has a high level of dysfunctional behavior and with a proxy for the level of dysfunctional behavior.

This study has the following implications. First, it suggests that given a high level of common uncertainty, an RPE-based incentive plan can induce agents to exert more effort, leading to higher financial performance than with an incentive contract with metrics based on absolute performance. This is consistent with Frederickson's (1992) experimental results that agents' effort levels are higher with an RPE contract than with a profit-sharing contract. Second, prior literature has been silent on the determinants of dysfunctional behavior caused by an RPE incentive plan. Our study suggests that the degree of common uncertainty within reference group can play a critical role in determining agents' level of dysfunctional behavior. A high degree of common uncertainty within reference group appears to decrease the costs caused by an RPE incentive plan.

The remainder of the paper is organized as follows. We review prior literature in Section 2 and describe our research site and data collection in Section 3. Section 4 develops our hypotheses. Section 5 presents estimation models and Section 6 presents our results. Concluding remarks are offered in Section 7.

2. PRIOR LITERATURE

Agency theory suggests that there are benefits associated with evaluating agents on their relative performances when the agents' performances are affected by a common "shock" term (Holmstrom 1979, 1982). The extant archival empirical literature on RPE has been limited to investigating whether top executives are compensated as if their performance is evaluated relative to competitors' performance and provides mixed results (Antle and Smith 1986, Gibbons and Murphy 1990, Janakiraman et al. 1992, Sloan 1993, Aggarwal and Samwick 1998, Anderson et al. 2001). For example, using 39 firms in the chemical, aerospace, and electronics industries during 1947 to 1977, Antle and Smith (1986) find weak evidence that the executives' compensation falls as other firms perform better, holding own performance fixed. Using a comprehensive survey, Gibbons and Murphy (1990) find that executives are penalized when a competitor group fares better, as predicted by theory. However, Janakiraman et al. (1992) find little empirical evidence that the market and industry components of firm performance are completely removed in determining CEO compensation. They conclude that the specific form of the agency model used to develop the RPE prediction is not descriptive for CEOs.

In a laboratory experimental setting, Frederickson (1992) provides evidence that agents' effort levels are higher with an RPE contract than with a profit-sharing contract where compensation is solely tied to agents' absolute performance. He also finds that subjects' effort levels increased significantly as the degree of common uncertainty increased with the RPE contract but not with the non-RPE contract. Using CEO compensation data, Kren (2002) finds that the degree of emphasis on RPE is positively related to the level of common uncertainty.

Combined, little empirical evidence is available on the performance consequences of

RPE-based incentive plans when there is an important source of common risk, especially at non-executive levels.[2] Moreover, no study has empirically documented how dysfunctional behavior influences an RPE-based contract's incentive effects (Prendergast 1999).

In their recent survey on the developments of budgeting practices, Hansen et al. (2003, 106) point out that even though an RPE-based performance standard provides "an increased degree of legitimacy" due to the inherent credibility of performance achieved by similar units, there has been no accounting research that investigates the motivational impacts of benchmarked standards.

Our study fills these gaps in the extant RPE literature by using archival data to provide evidence on the performance impact of RPE given common risk and by examining whether dysfunctional behavioral responses can mitigate the incentive effect. Our study also explicitly explores the degree of common uncertainty as a key factor influencing the net benefits of introducing an RPE-based contract. Furthermore, our study contributes to the growing literature that emphasizes the important role performance measurement systems can play in improving the efficiency of government operations.[3]

---

[2] Several studies that did not focus on RPE have found substantial incentive effects at non-executive levels. Banker et al. (1996) report that the implementation of a performance-based compensation plan increased sales and that the effect persisted and grew over time. Lazear (2000) examines how piece rates affected the performance of workers who installed automobile windshields. He finds that worker output increased after the company switched its compensation scheme from hourly wages to piece rates. Banker et al. (2000) find that firm performance improved after customer satisfaction was included as a performance measure for managers. Sprinkle (2000) provides experimental evidence regarding how incentive-based compensation contracts compare to flat-wage compensation contracts in motivating individual performance. He shows that incentives not only increase the duration of effort but they also appear to increase the intensity of effort. This suggests that incentives improve individuals' analysis and use of feedback information and encourage them to develop strategies to maximize performance.

[3] For example, Ittner and Larcker (1998) suggest that the most fundamental question in governmental performance measurement practice is whether private sector notions of performance measurement and accountability are applicable in the public sector.

3. RESEARCH SITE AND DATA DESCRIPTION

Description of Research Site

The research site for this study is Korea Post (hereafter referred to as POST). POST is the state-owned company that provides comprehensive postal services throughout Korea. POST is chartered as a monopolistic service provider to handle and deliver mail. However, it competes against several large domestic couriers and international couriers, such as Federal Express and DHL, in its expedited delivery market, which is more profitable than regular mail service. Intense competition in the expedited delivery market significantly undermines POST's profitability. Furthermore, rapid diffusion of electronic communication such as e-mail and mobile phone service poses a serious threat to POST's core business. A universal service obligation imposed on POST precludes POST from using its discretion to close unprofitable stores, such as those located in remote islands. These factors, coupled with stagnant mail volume and decreased profit margins caused by strict regulation of postal service rates, collectively contributed to troubled financial performance.

POST's stores are homogeneous in many important operational aspects, such as organizational structure, mail handling and delivery process, incentive system, and basic infrastructure. However, stores vary significantly in terms of exogenous factors such as geographical location, clientele, and competition. Considering that mail volumes are largely determined by exogenous factors, it is highly likely that a store's profitability is conditional on such 'uncontrollable' exogenous factors.[4]

[4] We regressed mail volume handled per employee of each store on major exogenous factors–population, service area, average delivery distance, number of financial institutions (a proxy for income level), and location of store (urban/rural) using 1997 data (pre-RPE). The adjusted R^2 is 73.2%, suggesting that a substantial proportion of variation in mail volume can be attributed to exogenous factors over which a store manager cannot exercise control.

Previous Incentive Plan

Even before it introduced the RPE incentive plan in late 1997, POST's incentive plan tied employees' incentive pay to the performance of the store where the employees worked. Each store's performance score was a weighted average of multiple measures. One of the key performance measures for evaluating each store was its profitability growth rate relative to its previous-year profit. This measure was selected because store profitability is influenced by "uncontrollable" exogenous factors and POST is obligated to provide universal service throughout the country. Therefore, absolute performance cannot accurately reflect managerial effort, that is, "controllable" performance. Similarly, productivity growth rate was another key performance measure for each store.[5]

Each store was ranked based on its weighted average evaluation score. No incentive pay was provided for stores whose relative rank fell within the bottom 50% of all stores. Bonuses ranged between 50%-150% of the monthly salary, and the bonus size was determined by the store's rank.[6] If a store qualified for a bonus, each employee in the store received the same percent of their monthly salary. Employees received the incentive bonus, if any, in April of the following year.

By including the profitability growth rate relative to the previous year as a key performance measure, this incentive plan partially resolved the problem of comparing absolute performance of stores whose business environments are quite diverse. However, it ignored the relative efficiency among stores and accordingly did not effectively motivate

---

[5] A store's profitability for incentive plan purposes was measured as revenue divided by operating cost. Although several measures were used, the weights placed on the profitability growth measure and the productivity growth measures were equally the largest in the formula for determining store employees' incentive pay.

[6] This means that the maximum bonus constituted 12.5% of annual compensation. Management concluded that the range of incentives provided had great potential to motivate workers, especially because performance-contingent bonuses were unusual among Korean state-owned companies.

stores to strive toward achieving greater efficiency. The following comments by a senior manager exemplify the motivational problems:

> *Before we introduced the new incentive plan, individual stores opportunistically manipulated the growth rate in profitability… Stores chose to exert effort on enhancing their profitability every two years. By doing so, in one year, they were able to maximize the growth rate in profitability relative to the previous year and get incentive pay. The next year they just didn't do anything..... Since all of us know that profit growth cannot be sustained eternally, basing incentive pay on profitability growth has an intrinsic problem…*[7]

New RPE-based Incentive Plan

To resolve this problem, POST introduced its new incentive plan for employees at each store in late 1997 with the objective of rewarding stores for their performance relative to an appropriate reference group.[8] Using cluster analysis, the new incentive plan classified all stores into nine reference groups, each of which shares a similar business environment.[9] Both relative profitability (store profit divided by the average profit in the reference group) and relative productivity (mail volume handled per store employee) are key performance

---

[7] To illustrate, a store's operating profit might be 100, 150, 100, and 150 in years 1990, 1991, 1992, and 1993, respectively. This pattern of profits is consistent with the following explanation. In 1991, the agents exerted high effort to maximize profit growth in order to get a bonus. In 1992, effort-averse agents did not exert any effort above the minimally required effort because they knew there was not much room to improve the performance that year. In 1993, they exerted high effort again to maximize the profit growth. This opportunistic behavior is attributable to the incentive plan that measures current year's performance relative only to the previous year's. The incentive problem is similar to that which arises when a reward system ratchets performance standards, inducing employees to reduce effort and expected output in one period to avoid higher standard in the next period (Indjejikian and Nanda 1999).

[8] The reference groups remained fixed until November 1999, when the groups were revised.

[9] Group 1 contains stores in Seoul, which has approximately 11 million people. Groups 2 and 3 contain stores in suburbs of Seoul and the next three largest cities (2 – 4 million people in each city). Groups 4 through 6 contain stores in smaller, mid-sized cities, and Groups 7 through 9 contain stores in rural areas. The number of stores included in each reference group varies from 16 to 33. The arithmetic mean of profitability measures of each reference group varies from 0.4412 to 1.5754 (year 1997). Variables used in the cluster analysis include key factors that are expected to influence store profitability and productivity. Examples include population, square kilometer of the area covered by the store, and number of employees. Those variables are also used as control variables in our regression model in Section 5.

measures on which the largest weights are placed in determining overall performance tied to incentive pay.[10] However, bonus eligibility remains the same as the old plan. That is, incentive pay is provided only to the stores whose weighted averages are in the top 50% of all stores.

Table 1 summarizes the differences between the next incentive plan and the old one. The new plan's performance measures rely on a store's performance relative to its reference group.[11] The old plan relied on growth in a store's performance over the previous year. Both plans incorporated a tournament-type feature in which bonuses were paid only if stores ranked in the top 50% on the weighted average evaluation score. However, the new incentive plan specifically incorporates relative performance evaluation in developing the performance scores. For simplicity, we refer to the old incentive plan as non-RPE and the new system as RPE.

Figure 1 diagrams POST's schedule of introducing the new incentive plan. In December 1997, the framework was circulated to top management. Stores received details of the plan in January 1998. Therefore, in testing the impact of the new incentive plan, January 1998 is considered as the start of the incentive plan. Performance evaluation of stores in calendar year 1997 under the new plan was performed in March 1998 and the

[10] We are unable to disclose specific details of the new incentive plan due to our confidentiality agreement.

[11] In June 2000, the company allowed stores to request to change reference groups, especially if the stores reported significant dysfunctional behavior due to perceived unattainable performance standards. The most conspicuous dysfunctional behavior was decreased productivity resulting from employee morale depressed by a sense of "abandonment," reported to be widespread in stores whose employees perceive their reference group as too "strong" for them. An internal survey of the new incentive plan's impact revealed that employees became the most frustrated when they did not receive a bonus even though they improved performance relative to the previous year. Other examples of store-level dysfunctional behavior include taking profitable customers away from other stores to increase revenue, or intercepting pre-sorted business bulk mail that could be directly received and processed at a mail-processing center. Interception of bulk mail created very inefficient mail processing and handling from top management's perspective, but increased the store's mail volume received, which partially determined the store's revenue.

resulting bonuses were paid in April 1998.[12]

Data Collection

Annual performance evaluation data were obtained for 1997 through 1999 for POST's 214 stores.[13] Financial and operational data for individual stores were collected from POST's performance measurement system. Control variables such as demographics and key statistics for individual stores that will be discussed in Section 5 were hand-collected from the company's other database. In addition, company documents and interviews with store managers provided qualitative data on their expectations and comments about the incentive plan.

4. HYPOTHESES

H1 and H2: RPE Incentives and Degree of Common Uncertainty

The first hypothesis can be motivated by agency theory, with the usual assumptions of risk averse and effort averse agents who maximize their expected utility in response to a principal's offered contract. Moreover, increases in an agent's effort are positively associated with increases in performance measures. Most pertinent to our study, Holmstrom (1982) analyzes agency situations with multiple agents and addresses the question of whether an optimal contract with an agent should depend only on that agent's performance

[12] Since actual performance evaluation of stores for year *t* is performed in March of year *t*+1, one can argue that during year *t*, stores have no way of knowing the benchmarked performance standards against which their performance will be compared. However, because the reference group assigned to each store remained unchanged during 1997-1999 and stores received the results of performance evaluation for 1997 no later than the end of March 1998, we assume that the benchmarked performance standard used in the 1997 evaluation served as a reasonable estimate of a performance standard for 1998 evaluation. Similarly, we assume that the benchmarked performance standard used in the 1998 evaluation served as a reasonable estimate of a performance standard for the 1999 evaluation.

[13] We were allowed to access the performance evaluation data for 2000. However, mergers among stores and changes in performance measures that occurred in 2000 preclude us from using the data for our analysis. Furthermore, in 2000, POST expanded the coverage of bonus eligibility to 80% of the stores from the previous 50% of the stores.

or on that agent's performance and other agents' performance. Holmstrom shows that under mild assumptions, if the agents face common uncertainty, then an optimal RPE-based contract with each agent is Pareto superior to an optimal contract restricted to that agent's performance measures. Intuitively, if the agents face common uncertainty, the principal gains information about one agent's effort from the performance measures of the other agents.

Frederickson (1992) argues that not only agency theory, but also behavioral theories predict that agents' effort levels are higher with an RPE contract than with a profit-sharing contract where an agent's compensation is solely tied to the absolute performance of that agent. In his experiment, agents' effort levels were higher with the RPE contract than with the non-RPE contract. Hypothesis 1 below is similar to Frederickson's H2 and H6 regarding agents' effort under RPE and non-RPE contracts. However, because we can only observe outcomes, not effort, our hypothesis test will involve financial performance.[14]

H1: Ceteris paribus, introduction of the RPE incentive plan is positively associated with financial performance.

As Frederickson (1992) states, increases in the degree of common uncertainty increase the informativeness of the relative performance measure, and theoretically allow the principal to design a Pareto superior contract. This thinking is consistent with POST's formation of reference groups of stores. Based on economic arguments, we therefore predict that increases in the degree of common uncertainty are associated with increases in

[14] We compare POST's new RPE scheme with the previous scheme, which differs from Frederickson's pure profit-sharing contract. However, Holmstrom's arguments rely on the incremental information available in the new contract versus the previous one, not on the particular form of the non-RPE contract.

financial performance.[15]

Social comparison theory argues that comparisons among individuals of similar abilities are most likely (Festinger 1954). Social psychology literature supports the theory by documenting that subjects' performance improved when paired with coworkers at similar ability level (e.g., Seta 1982, among others). Frederickson (1992) argues from social comparison theories that the greater the degree of agents' common uncertainty, the greater the saliency of relative performance information. The greater saliency in turn generates more cognitive activity and competition among agents should increase with the degree of common uncertainty. We use these arguments and the economic argument above to motivate hypothesis 2. As in hypothesis 1, our hypothesis below differs from Frederickson's H5 by ultimately testing the effect on financial performance rather than effort.

H2: With the RPE contract, ceteris paribus, increasing the degree of common uncertainty increases financial performance.

H3 and H4: Dysfunctional Behavior, RPE, and Degree of Common Uncertainty

Although economic theory shows that RPE can provide benefits when there is common uncertainty, several factors rest outside the theory. For example, Holmstrom (1982) did not allow for collusion among agents. Furthermore, RPE-based contracts may generate incentives to sabotage the measured performance of coworkers, to collude with coworkers and shirk, and to apply for jobs with inept coworkers (Gibbons and Murphy 1990).[16]

[15] Frederickson (1992) calculated optimal effort levels with fixed contracts under a variety of conditions and found that in his examples, the agents' optimal effort did not change with the degree of common uncertainty.
[16] However, these dysfunctional responses are not very likely to occur in POST. Since each reference group

In our situation, some stores reported major problems with dysfunctional behavior, such as decreased morale or skepticism, resulting from the perception that membership in a particular reference group created unfair and unattainable standards for achieving a bonus. That is, there was a low degree of common uncertainty with the reference group. We test the association between the degree of common uncertainty and expected high levels of dysfunctional behavior, as stated in H4. We also use the converse of Frederickson's arguments above to develop H3 below. That is, likely dysfunctional behavior is expected to be associated with lower financial performance. This prediction is also consistent with the findings of the budgeting literature that performance decreases at extremely difficult budget level dues to discouragement and less commitment to achieve the target (Merchant and Mazoni 1989, Fisher et al. 2003).

H3: The RPE incentive plan's impact on performance is mitigated (lower) in stores at which the level of dysfunctional behavior is expected to be high, relative to other stores.

H4: The degree of common uncertainty is negatively associated with the probability a store requested a change of reference group. That is, the degree of common uncertainty is negatively associated with expected high levels of dysfunctional behavior.

5. EMPIRICAL MODELS

Test of H1 and H2: RPE Incentives; Degree of Common Uncertainty

H1 is based on the argument that holding other factors fixed, agents' effort levels are higher with the RPE contract than with the non-RPE contract. Therefore, the introduction of the RPE incentive plan is expected to be positively associated with stores' financial performance.

---

comprises geographically dispersed stores, it is not likely that a store manager could sabotage the measured performance of other stores in the same reference group or collude with them.

To test H1, we estimate the following regression equation by using pooled time-series data for 214 stores.[17]

$$PROFRATE_{i,t} = \alpha_0 + \alpha_1 FYR\_PLAN_{i,t} + \alpha_2 SYR\_PLAN_{i,t} + \alpha_3 SERVICE\_AREA_{i,t} + \alpha_4 POP_{i,t} + \alpha_5 N\_BANK_{i,t} + \alpha_6 METRO_{i,t} + \alpha_7 N\_EMP_{i,t} + \alpha_8 DIST_{i,t} + \alpha_9 N\_DELPOINT_{i,t} + \alpha_{10} N\_COUNTER + \alpha_{11} N\_DELSEC + \varepsilon_{i,t} \quad (1)$$

(t = 1997-1999)

Where:

PROFRATE	=	rate of profitability for store *i* measured as revenue divided by operating cost at year *t*
FYR_PLAN	=	dummy variable representing the first year following RPE incentive plan implementation, taking the value 1 for 1998 and 0 otherwise
SYR_PLAN	=	dummy variable representing the second year following RPE incentive plan implementation, taking the value 1 for 1999 and 0 otherwise
SERVICE_AREA	=	square kilometer of the area that store *i* is responsible for at year *t*
POP	=	population of the area that store *i* is responsible for at year *t*
N_BANK	=	number of financial institutions in the area that store *i* is responsible for at year *t*
METRO	=	dummy variable representing the area that store *i* is responsible for at year *t*, taking the value 1 for urban areas, and 0 otherwise
N_EMP	=	number of employees of store *i* at year *t*
DIST	=	average daily delivery distance per mail carrier of store *i* at year *t*
N_DELPOINT	=	average number of daily delivery points per mail carrier of store *i* at year *t*
N_COUNTER	=	number of available counters of store *i* at year *t*
N_DELSEC	=	number of delivery districts of store *i* at year *t*

We use *PROFRATE*, which is defined as revenue divided by operating cost of a store, as our dependent variable. We use this measure as our main dependent variable for the following reasons. First, this measure is actually used in the new incentive plan as a

[17] We use ordinary least square (OLS) estimation unless otherwise stated.

profitability measure. Second, by using *PROFRATE*, we are able to reduce the likelihood of heteroskedasticity that might occur when using alternative financial measures such as revenue or operating profit.[18] Third, by using this measure we can capture the effect of agents' efforts on both revenue increase and cost reduction.[19]

Discussion with POST's management revealed that stores initially responded to the new incentive plan by reducing operating cost rather than trying to increase revenue because mail volume, a store's major revenue driver, is largely determined by "uncontrollable" exogenous factors.[20] Furthermore, stores reportedly exerted effort to increase other revenue sources such as railway or airline ticketing service or e-post (an on-line shopping mall that sells specialty products peculiar to that region) Therefore, *PROFRATE* will effectively capture the plan's impact on efforts for reducing cost and increasing revenue at store level. However, several studies have examined the choice of incentive schemes and their effects on productivity (Lazear 2000, Ho et al. 2002). To examine the impact of the new incentive plan on productivity, we also use *MVOL*, which is defined as the natural log of mail volume handled per employee in a store as an alternative dependent variable.

The above model includes dummy variables *FYR_PLAN* and *SYR_PLAN* to account for any other changes in profitability that may occur after implementation of the new incentive plan. The coefficient on *FYR_PLAN* captures the average difference in profitability measures between the first year after the new incentive plan (year 1998) and

---

[18] Heteroskedasticity is a common problem encountered when the estimation is based on cross-sectional data. One way to reduce this problem is to use a deflated dependent variable. For example, Banker et al. (2000) use normalized revenue and cost deflated by the number of available rooms as a proxy for the size of a hotel. However, in our setting, we are unable to find an appropriate deflator.

[19] Postal rates did not change during 1997-1999.

[20] Examples of cost reduction efforts include eliminating unneeded employees, substituting part-timers for regular workers, encouraging early retirement of senior employees, as well as trimming unnecessary costs.

base year 1997. Similarly, the coefficient on *SYR_PLAN* captures the average difference in profitability measures between the second year after the new incentive plan (year 1999) and base year 1997. Significant positive coefficients on *FYR_PLAN* and *SYR_PLAN* after controlling for other exogenous factors are interpreted as their average impact on financial performance, thus providing support for H1.

We include a set of variables to control for exogenous factors that are expected to influence *PROFRATE*. Including proxy variables that control for differences in the business environments of stores in the estimation model is important to isolate the incentive effect of the new plan (Dopuch and Gupta 1997). We chose these control variables because these variables were actually used in the cluster analysis to form POST's nine reference groups,[21] and consequently, POST believes those control variables are highly correlated with a store's profitability.[22] Of course, one should use caution in claiming that these control variables include all exogenous factors that affect a store's profitability. In addition, some control variables may suffer from measurement error. For example, *N_BANK* is included to control for the effect of average income on profitability. More appropriate proxies for average per capita income, such as income tax paid or average house price of house, were not available. Our concern about not including a more direct proxy for per capita income is alleviated by the fact that the whole nation's average per capita income in 1998 relative to 1997 decreased due to the macroeconomic impact of the nation's financial crisis in late 1997. If income is a significant determinant of a store's profitability, that should bias against our results. Furthermore, the rapid substitution of electronic communication such as e-mail and

[21] In estimating benchmark performance standards for public school expenditure, Dopuch and Gupta (1997) also used proxy variables to reflect differences in socio-economic factors that are used by the State of Missouri in setting a funding formula.

[22] Postmaster's tenure as a proxy for managerial expertise could be included as an additional explanatory variable. However, since POST strictly enforced mandatory job rotation policy about every three years, postmaster's tenure is not expected to be highly correlated with store profitability.

mobile phone service for traditional surface mail in late 1990s should also bias against finding our results.

H2 predicts that under the RPE contract, increasing the degree of common uncertainty increases financial performance. To test H2, we first estimate the following regression equation for each reference group.[23]

$$PROFRATE_{i,t} = \beta_0 + \beta_1 SERVICE\_AREA_{i,t} + \beta_2 POP_{i,t} + \beta_3 N\_BANK_{i,t} + \beta_4 N\_EMP_{i,t} + \beta_5 DIST_{i,t} + \beta_6 N\_DELPOINT_{i,t} + \beta_7 N\_COUNTER_{i,t} + \beta_8 N\_DELSEC_{i,t} + \varepsilon_{i,t} \quad (2)$$

We operationalize the degree of common uncertainty by the adjusted R^2 of a store obtained by running the above reference group-specific regression.[24] The adjusted R^2 obtained by running the above reference group-specific regression captures the extent to which the variation in profitability is explained by the variation in independent variables within each reference group. Considering that the independent variables are those used in the cluster analysis to form the nine reference groups, our adjusted R^2 measure should effectively capture the degree of common uncertainty. Since the magnitude of R^2 could be noisy, we use a ranking on adjusted R^2 that ranges from 1 to 9 with higher values indicating greater R^2.

Then, using store data during 1998-1999 (Post-RPE), we estimate the following regression equation:

$$PROFRATE_{i,t} = \gamma_0 + \gamma_1 COMMON_{i,t} + \gamma_2 SERVICE\_AREA_{i,t} + \gamma_3 POP_{i,t} + \gamma_4 N\_BANK_{i,t} + \gamma_5 METRO_{i,t} + \gamma_6 N\_EMP_{i,t} + \gamma_7 DIST_{i,t} + \gamma_8 N\_DELPOINT_{i,t} + \gamma_9 N\_COUNTER + \gamma_{10} N\_DELSEC + \varepsilon_{i,t} \quad (3)$$

---

[23] Indicator variable *METRO* could not be used in estimating reference group-specific equation (2) because all the stores within each reference group are located in either an urban or rural area.

[24] Considering that a store's profitability is conditional on such "uncontrollable" exogenous factors, the extent to which there is common uncertainty could be measured by the degree of correlation between the store and the relevant peer group with which its performance will be compared. This situation is analogous to the RPE literature on the compensation of executives documenting that the relevant peer group should be companies in the same industry rather than the entire stock market in the sense that there should be more correlation in common shocks within the same industry (Prendergast 1999).

(t = 1998-1999)

Where:

COMMON = R^2 ranking of a store obtained by running reference group-specific regression (2); ranges from 1 through 9 with higher values indicating greater R^2

We expect the coefficient on *COMMON* to be significantly positive, which suggests that with the RPE contract, a store's degree of common uncertainty within its reference group is positively associated with its financial performance.

Test of H3: Dysfunctional Behavior and RPE

H3 predicts that the RPE incentive plan's impact on performance is mitigated in stores at which the level of dysfunctional behavior is expected to be high, relative to other stores. To test H3, we estimate the following regression equations by using pooled time-series data for 214 stores.

$$\begin{aligned} PROFRATE_{i,t} = {} & \delta_0 + \delta_1 FYR\_PLAN_{i,t} + \delta_2 SYR\_PLAN_{i,t} + \delta_3 FYR\_PLAN_{i,t} * DYSFUNCTION_{i,t} \\ & + \delta_4 SYR\_PLAN_{i,t} * DYSFUNCTION_{i,t} + \delta_5 SERVICE\_AREA_{i,t} + \delta_6 POP_{i,t} + \delta_7 N\_BANK_{i,t} + \delta_8 METRO_{i,t} \\ & + \delta_9 N\_EMP_{i,t} + \delta_{10} DIST_{i,t} + \delta_{11} N\_DELPOINT_{i,t} + \delta_{12} N\_COUNTER + \delta_{13} N\_DELSEC + \varepsilon_{i,t} \quad (4) \end{aligned}$$

(t = 1997-1999)

The following alternatives are used for *DYSFUNCTION:*

CHANGE = dummy variable representing stores that requested to change their reference group in 2000, taking the value 1 if the store requested a change, and 0 otherwise

NOCOMP = dummy variable representing stores which did not receive incentive bonus for three consecutive years, taking the value 1 if the stores did not receive a bonus for three consecutive years, and 0 otherwise

PROFRANK = standardized profitability rank of a store within its reference group at year t, with higher values indicating lower relative profitability within a reference group

The key issue is how to identify stores at which the level of dysfunctional behavior will be high. In our analysis, we use three empirical proxies for capturing the level of dysfunctional behavior caused by new incentive plan. First, we use two alternative indicator variables, *CHANGE* and *NOCOMP*, to identify stores at which the level of dysfunctional behavior is expected to be high. *CHANGE* refers to 13 stores that filed a formal request to headquarters to change reference group in mid-2000.[25] We infer that they are likely to have suffered from serious levels of dysfunctional behaviors such as "abandonment," or skepticism of the new plan since the plan's introduction.[26] *NOCOMP* refers to 47 stores that did not receive an incentive bonus for three consecutive years after the new incentive plan's introduction. We infer that they are also highly likely to have suffered from serious levels of dysfunctional behaviors. We expect coefficients on the interaction term between the incentive plan dummy and the dysfunctional behavior dummy, *FYR_PLAN*CHANGE* (*NOCOMP*) and *SYR_PLAN*CHANGE* (*NOCOMP*), to be negative.

For a robustness check, we employ *PROFRANK*, another variable for measuring dysfunctional behavior. *PROFRANK* refers to the standardized profitability rank of a store within its reference group, with higher values indicating lower relative profitability within a reference group. Since the profitability of a store relative to its reference group receives the largest weight in determining bonuses, we expect that the level of dysfunctional behavior

[25] As described in Figure 1, POST adjusted the reference group for each store by performing a second cluster analysis in November 1999. Therefore, the stores whose reference group was changed in November 1999 could be considered those at which dysfunctional behavior is expected to be high. However, since new variables were added in the second cluster analysis, the change of reference group could be attributable to the addition of these new variables in the analysis. Therefore, we chose to use 13 stores that filed a formal request to postal headquarters to change reference group in 2000 as "dysfunctional" stores. These stores were allowed to file a formal complaint to postal headquarters only after a regional postmaster carefully investigated the store's reported problems.

[26] Interviews with senior managers in stores revealed that the key dysfunctional behavior due to the new plan was "abandonment." That is, employees in stores whose reference group is too "strong" for them are not well motivated to exert effort to beat the benchmark performance standard that is set as the arithmetic mean of performance measures of stores that belong to the specific reference group.

will be high at the stores whose ranks are in the bottom of the reference group under the RPE-based plan. Accordingly, we expect coefficients on the interaction term between the incentive plan dummy and the profitability rank, *FYR_PLAN*PROFRANK* and *SYR_PLAN*PROFRANK*, to be negative.

These results would suggest that the new incentive plan's impact is mitigated in stores that suffered from dysfunctional behaviors, indicating that the RPE-based plan's incentive effect is conditional on the level of dysfunctional behaviors.

Test of H4: Dysfunctional Behavior and Degree of Common Uncertainty

H4 predicts that the degree of common uncertainty is negatively associated with the level of dysfunctional behavior.

To test H4, we estimate the following logit regression examining the association between degree of common uncertainty and probability a store requested a change of reference group:

$$Prob(DYSFUNCTION) = F(\alpha_0 + \alpha_1 COMMON_{i,t} + \alpha_2 SERVICE\_AREA_{i,t} + \alpha_3 POP_{i,t} + \alpha_4 N\_BANK_{i,t} + \alpha_5 METRO_{i,t} + \alpha_6 N\_EMP_{i,t} + \alpha_7 DIST_{i,t} + \alpha_8 N\_DELPOINT_{i,t} + \alpha_9 N\_COUNTER + \alpha_{10} N\_DELSEC) \quad (5)$$

Where:

COMMON = R^2 ranking of a store obtained by running reference group-specific regression (2); ranges from 1 through 9 with higher values indicating greater R^2

The following two alternatives are used for *DYSFUNCTION*:

CHANGE = dummy variable representing stores that requested to change their reference group in 2000, taking the value 1 if the store requested a change, and 0 otherwise

NOCOMP = dummy variable representing stores which did not receive incentive bonus for three consecutive years, taking the value 1 if the stores did

not receive a bonus for three consecutive years, and 0 otherwise

To gain further insight into the association between the degree of common uncertainty and the level of dysfunctional behavior, we estimate the following OLS regression of *PROFRANK* on *COMMON* and control variables.

$$PROFRANK_{i,t} = \delta_0 + \delta_1 COMMON_{i,t} + \delta_2 SERVICE\_AREA_{i,t} + \delta_3 POP_{i,t} + \delta_4 N\_BANK_{i,t} + \delta_5 METRO_{i,t} + \delta_6 N\_EMP_{i,t} + \delta_7 DIST_{i,t} + \delta_8 N\_DELPOINT_{i,t} + \delta_9 N\_COUNTER + \delta_{10} N\_DELSEC + \varepsilon_{i,t} \quad (6)$$

PROFRANK = standardized profitability rank of a store within its reference group at year *t*, with higher values indicating lower relative profitability within a reference group

We expect the coefficient on *COMMON* to be significantly negative, which suggests that agents are more (less) likely to suffer from dysfunctional behaviors under the RPE incentive contract if the degree of common uncertainty is low (high) within their reference group.

6. RESULTS

Descriptive Statistics

Descriptive statistics of the variables used in the analyses appear in Table 2.[27] For comparison, we provide descriptive statistics for both the pre-RPE period and the post-RPE period. Overall, the average financial performance of stores improved after the RPE-based plan was introduced. On average, operating profit increased from 60 million (Korean won) during the pre-RPE period to 469 million during the post-RPE period due to increased revenues and reduced operating costs. The large standard deviation in operating profit implies that stores vary significantly in terms of profitability. The rate of profitability (revenue divided by operating cost) increased from 0.84 during the pre-RPE period to 0.95

[27] Due to store-level missing data in some of our control variables, the final number of observations used in estimating our regression models is 619.

during the post-RPE period. An average profitability measure that is smaller than 1 suggests that on average stores are suffering losses. The results of the t-test and Wilcoxon rank sum test show that most of the control variables are not statistically different between the pre-RPE period and post-RPE period. This provides preliminary evidence that the change in profitability, if any, could be attributable to the implementation of the RPE-based plan.

Table 3 provides details on Pearson correlations among variables. Square kilometers of the area (SERVICE_AREA), average number of daily delivery points per mail carrier (*N_DELPOINT*), and average daily distance per mail carrier (*DIST*) are negatively and significantly correlated with operating profit and profitability. Other independent variables such as number of financial institutions (*N_BANK*) and population (*POP*) are all positively and significantly correlated with a store's profit.[28]

Impact of Incentive Plan on Financial Performance

Table 4 presents the estimation results of equation (1) examining the new incentive plan's impact on profitability. The coefficients on *FYR_PLAN* and *SYR_PLAN* after controlling for other exogenous factors are positive and significant ($p < 0.0001$ for both), indicating that the incentive plan change has a positive and significant effect on financial performance. Thus, our results support H1.[29] This result is consistent with the experimental finding in Frederickson (1992) that agents' effort levels are higher under the RPE contract than under the non-RPE contract.

---

[28] A significant level of correlations amongst variables raises a concern about multicollinearity problems in our estimation model. Accordingly, we check variance inflation factors (VIF) of independent variables and the conditions index by Belsley, Kuh, and Welsch (1980).

[29] White's test indicates that the homoskedasticity assumption is violated. Accordingly, we report coefficient estimates from OLS regressions and asymptotic t-statistics based on White (1980) standard errors. Examination of variance inflation factors (VIF) and the conditions index reveals that multicollinearity is not a serious issue in our estimation model.

As seen in Table 4, results are very similar when mail volume handled per employee (MVOL) is used as an alternative dependent variable. This suggests that improved productivity caused by the new incentive plan plays a role in driving profitability.[30] As discussed in section 5, interviews with management and store manager revealed that the initial profitability improvement in response to the new incentive plan is mainly attributable to reducing labor cost and overhead cost. Given that mail volume is largely determined by exogenous factors, a positive and significant incentive effect on productivity appears to be driven by stores' effort to eliminate unnecessary employee hours.

The results in table 4 also indicates that the average store profitability rate in the first year after introducing the RPE-based plan is 0.628 (= 0.540 + 0.088) compared with the pre-RPE (1997) average profitability rate of 0.540 after controlling for important factors influencing profitability. In 1999, the second year after the RPE-based plan, average profitability rate is 0.699 (=0.540+0.159). To provide insights into whether the new plan's impact on performance persisted and grew over time (i.e. continuing improvement), we perform a test of equality of coefficients on *FYR_PLAN* and *SYR_PLAN*. Based on the result of the F-test, we reject the null hypothesis that the two coefficients are equal at the 1% level, suggesting that store employees allocated more effort to learn how to perform better and were motivated to further improve profitability in the second year following implementation of the new plan.[31]

---

[30] Since the use of mail volume handled per employee (MVOL) as an alternative dependent variable produces very similar results, the subsequent analyses use PROFRATE only as the dependent variable.

[31] When *MVOL* is used as the dependent variable, the coefficient on *SYR_PLAN* is still statistically larger than that on *FYR_PLAN*, but the difference between the two becomes smaller than when we use *PROFRATE* as our dependent variable. This suggests that the new plan's impact on continuing improvement of profitability in the second year will be greater than its impact on productivity because, in comparison to the productivity measure over which stores could not exert much control, the profitability rate provides us with a more comprehensive performance measure in the sense that it reflects stores' efforts to trim overhead costs and to increase revenue sources other than those generated by mail volume.

Is the Degree of Common Uncertainty Associated with Performance?

Table 5 provides the results from estimating equation (3), which examines whether with the RPE contract, the degree of common uncertainty is positively associated with store profitability. The coefficient on *COMMON* is positive and significant ($p < 0.001$), suggesting that holding other factors fixed, the greater degree of common uncertainty leads to the higher performance.[32] This is consistent with the experimental finding in Frederickson (1992) that with the RPE contract, ceteris paribus, increasing the degree of common uncertainty increases agents' effort levels. Our result suggests that it is critical to explicitly take into account the major differences in environments of business units when forming a reference group.

Does Dysfunctional Behavior Mitigate the Performance Impact of the Incentive Plan?

Table 6 provides the results from estimating equation (4), which examine whether the RPE contract's performance impact is mitigated due to dysfunctional behaviors. Panel A of Table 6 shows that as in Table 4, the coefficients on *FYR_PLAN* and *SYR_PLAN* are positive and significant. As predicted, the negative and significant coefficients on the *FYR_PLAN*CHANGE* and *SYR_PLAN*CHANGE* interactions support our prediction that the RPE plan's incentive effect will be mitigated in "dysfunctional" stores ($p < 0.01$ and $p < 0.05$, respectively). As seen in panel B of Table 6, when the variable NOCOMP is used to identify "dysfunctional" stores, the *FYR_PLAN*NOCOMP* and *SYR_PLAN*NOCOMP* interactions are still significantly negative ($p < 0.0001$ for both). The results using both *CHANGE* and *NOCOMP* suggest that the first-year cost (dysfunctional effect) of the RPE

[32] The use of R^2 value instead of R^2 ranking provides similar results.

contract exceeded the benefits, but the benefits exceeded the cost in the second year.

Panel C of Table 6 shows the estimation result when we use the variable *PROFRANK*. Again, the coefficients on the *FYR_PLAN*PROFRANK* and *SYR_PLAN*PROFRANK* interactions are negatively significant ($p < 0.0001$ for both).

Taken together, these results indicate that the RPE contract's incentive effect is mitigated in "dysfunctional" stores and that the results are robust to an alternative way of identifying "dysfunctional" stores.[33]

Is the Degree of Common Uncertainty Associated with Dysfunctional Behavior?

Panel A of Table 7 provides the results from estimating the logit regression that examines whether the degree of common uncertainty is negatively associated with the probability of predicting stores at which the level of dysfunctional behavior is expected to be high.[34] Consistent with H4, when *CHANGE* is used as a dependent variable, the coefficient on *COMMON* is negative and significant ($p = 0.007$). This suggests that agents are more (less) likely to suffer from dysfunctional behavior under the RPE incentive contract if the degree of common uncertainty is low (high) within their reference group.[35] The result also implies that the dysfunctional behavior reported in 13 stores after the RPE contract could in part be attributable to low common uncertainty in their reference group.

However, the coefficient on *COMMON* is not statistically different from zero when

[33] For further robustness checks, we estimated equation (4) using the following two alternative ways of measuring the level of dysfunctional behavior. First, we compared the rank of a store on its profitability divided by the mean of its reference group (profitability measure under the new plan) with the rank of the store on profitability growth. We expect a store to suffer from dysfunctional behavior if that store's profitability relative to its peers is low even though it achieves high profitability growth. Therefore, using the difference between those two ranks, we identify the top 10% of stores that exhibited high profitability growth, but low relative profitability and create an indicator variable. Second, we used 37 stores that in November 1999 had their reference group changed to a "lower" reference group relative to their first reference group in December 1997. The results are very similar to those reported in Table 6.

[34] Estimation of a probit regression does not change our inference.

[35] The use of R^2 value instead of R^2 ranking provides similar results.

NOCOMP, the variable indicating stores that did not receive an incentive bonus for three consecutive years, is employed as a dependent variable. Even though these inconsistent results from estimating a logit regression make it difficult to draw unambiguous inferences on the association between the degree of common uncertainty and dysfunctional behavior, the result with the variable *NOCOMP* should be interpreted with caution because we believe that *CHANGE* does a better job of identifying "dysfunctional" stores than *NOCOMP*, for the following two reasons. First, the stores indicated by *CHANGE* were allowed to file a formal complaint to postal headquarters only after a regional postmaster carefully investigated the store's reported problems. Second, the stores indicated by *NOCOMP* are likely to include not only stores that became "dysfunctional" after the new incentive plan, but also stores that have consistently evaluated as inferior across incentive schemes.

For an additional analysis, we estimate equation (6), in which *PROFRANK* is regressed on *COMMON*. Panel B of Table 7 shows that the coefficient on *COMMON* is negative and significant ($p < 0.0001$). Overall, the results, combined with the result in Table 4, imply that the net benefits of introducing the RPE contract might be highly conditional on the degree of common uncertainty. This is because the degree of common uncertainty appears to be inversely related to the likelihood of observing dysfunctional behavior or the level of dysfunctional behavior that could be caused by the RPE contract.

7. CONCLUDING REMARKS

Using annual performance evaluation data of 214 stores of Korea Post during 1997 through 1999, we find that the introduction of an RPE-based incentive plan is positively associated with financial performance and that under the new incentive plan, the degree of

common uncertainty is positively associated with store profitability. Furthermore, we find evidence that the RPE-based plan's incentive effect is mitigated in stores at which the level of dysfunctional behavior is expected to be high. Finally, we find that the net benefits of introducing the RPE contract may be highly conditional on the degree of common uncertainty since the degree of common uncertainty appears to be inversely related to the likelihood of observing dysfunctional behavior and the level of dysfunctional behavior that could be caused by the RPE-based contract.

Our results have several implications for implementations of RPE-based incentive plans. First, even though an RPE-based incentive plan appears to induce agents to exert more effort by introducing competition, a significant level of dysfunctional behavior such as decreased morale or skepticism can occur. Second, the observed dysfunctional behavior caused by introducing an RPE contract may mitigate the incentive effect. Third, our study suggests that the level of dysfunctional behavior increases as the degree of common uncertainty decreases. This suggests that more careful attention should be directed to setting the benchmark standards against which the performance of business units is compared, and it is critical to explicitly take into account the major differences in environments of business units when forming a reference group. This confirms conventional wisdom that benchmark units should be from similar environments (Dopuch and Gupta 1997; Brickley et al. 2003).

REFERENCES

Aggarwal, R., and A. Samwick. 1999. Executive compensation, strategic competition, and relative performance evaluation: theory and evidence. *Journal of Finance* 54: 1999-2043.

Anderson, M., R. Banker, and S. Ravindran. 2001. Another look at relative performance evaluation hypothesis. Working paper, University of Texas-Dallas.

Antle, R., and A. Smith. 1986. An empirical investigation of relative performance evaluation of corporate executives. *Journal of Accounting Research* 24: 1-39.

Banker, R., S. Lee, and G. Potter. 1996. A field study of the impact of a performance-based incentive plan. *Journal of Accounting and Economics* 21: 195-226.

_____, G. Potter, and D. Srinivasan. 2000. An empirical investigation of an incentive plan that includes nonfinancial measures. *The Accounting Review* 75: 65-92.

Beck, H., and J. Seta. 1980. The effects of frequency of feedback on simple coaction task. *Journal of Personality and Social Psychology* 38: 75-80.

Brickley, J., C. Smith, and J. Zimmerman. 2003. *Managerial Economics and Organizational Architecture*. 3rd ed. Chicago: R.D. Irwin, Inc.

Dopuch, N., and M. Gupta. 1997. Estimation of benchmark performance standards: an application of public school expenditures. *Journal of Accounting and Economics* 23: 141-161.

Festinger, L. 1954. A theory of social comparison processes. *Human Relations* 7: 117-140.

Fisher, J., S. Peffer, and G. Sprinkle. 2003. Budget-based contracts, budget levels, and group performance. *Journal of Management Accounting Research* 15: 51-74.

Frederickson, J. 1992. Relative performance information: the effects of common uncertainty and contract type on agent effort. *The Accounting Review* 67: 647-669.

Gibbons, R., and K. Murphy. 1990. Relative performance evaluation for chief executive officers. *Industrial and Labor Relation Review* 43: 30-51.

_____. 1998. Incentive in organization. *Journal of Economic Perspectives* 12: 115-132.

Hansen, S., D. Otley, and W. Van der Stede. 2003. Practice development in budgeting: an overview and research perspective. *Journal of Management Accounting Research* 15: 95-116.

Harkins, S. and J. Jackson. 1985. The role of evaluation in eliminating social loafing. *Personality and Social Psychology Bulletin* 11: 457-465.

Ho, J., A. Wu, and L. Lee. 2002. An empirical study of how changes in compensation plans can hurt employees and company performance – car dealership case. Working paper, University of California-Irvine.

Holmstrom, B. 1979. Moral hazard and observability. *Bell Journal of Economics* 10: 74-91.

_____. 1982. Moral hazard in teams. *Bell Journal of Economics* 13: 324-340.

Indjejikian, R. 1999. Performance evaluation and compensation research: an agency perspective. *Accounting Horizon* 13: 147-157.

_____, and D. Nanda. 1999. Dynamic incentives and responsibility accounting. *Journal of Accounting and Economics* 27: 177-201.

Ittner, C., and D. Larcker. 1998. Innovations in performance measurement: trends and research implications. *Journal of Management Accounting Research* 10: 205-238.

_____, _____. 2001. Assessing empirical research in managerial accounting: a value-based management perspective. *Journal of Accounting and Economics* 32: 349-410.

_____, _____, and M. Rajan. 1997. The choice of performance measures in annual bonus contracts. *The Accounting Review* 72: 231-255.

Janakiraman, S., R. Lambert, and D. Larcker. 1992. An empirical investigation of the relative performance evaluation hypothesis. *Journal of Accounting Research* 30: 53-69.

Jensen, M., and K. Murphy. 1990. Performance pay and top management incentives. *Journal of Political Economy* 98: 225-264.

Kren, L. 2002. Common uncertainty effects on the use of relative performance evaluation for corporate chief executives. *Advances in Accounting* 19: 119-138.

Lazear, E. 2000. Performance pay and productivity. *American Economic Review* 90: 1346-1361.

Merchant, K., and J-F. Manzoni. 1989. The achievability of budget targets in profit centers: a field study. *The Accounting Review* 64: 539-558.

Murphy, K. 1999. Executive compensation. In: Orley, A., and C. David (eds.), *Handbook of Labor Economics*, Vol. 3. North-Holland, Amsterdam.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37: 7-63.

Seta, J. 1982. The impact of comparison processes on coactors' task performance. *Journal of Personality and Social Psychology* 42: 281-291.

Sloan, R. 1993. Accounting earnings and top executive compensation. *Journal of Accounting and Economics* 16: 55-100.

Sprinkle, G. 2000. The effect of incentive contracts on learning and performance. *The Accounting Review* 75: 299-326.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817-838.

FIGURE 1
Summary Schedule of New Incentive Plan Introduction

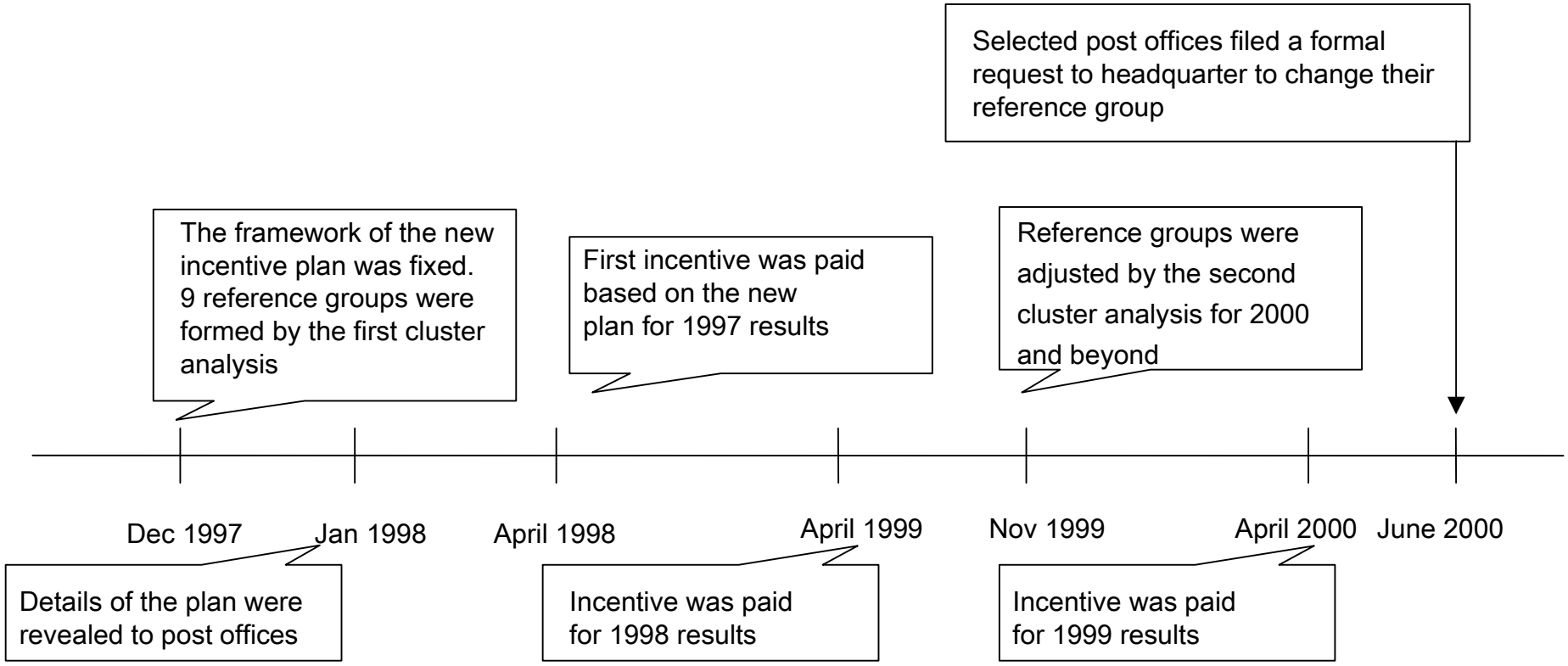

TABLE 1
Comparison of New Incentive Plan with Old Plan

	Old Plan	New Plan
Key Performance Measures	• Store profitability growth relative to previous year • Store productivity growth relative to previous year	• $\frac{\text{Store Profitability}}{\text{Reference Group Profitability}}$ • $\frac{\text{Store Productivity}}{\text{Reference Group Productivity}}$
RPE Component in Performance Measure?	• Others' performance does not affect a store's performance score	• Others' performance affects a store's performance score since relative- performance-based measures are used in calculating performance score
Rank Store Performance?	• Yes	• Yes
Bonus Eligibility	• Top 50% of population	• Top 50% of population

Profitability = store revenue divided by operating cost
Productivity = mail volume handled per employee
Reference group profitability = arithmetic mean of profitability of stores that belong to specific reference group
Reference group productivity = arithmetic mean of productivity of stores that belong to specific reference group

TABLE 2
Descriptive Statistics for Variables Used in the Analyses

	Pre-RPE (1997)				Post-RPE (1998,1999)				*t-test*		*Wilcoxon Rank Sum*	
Variable	N	Mean	Std.	Median	N	Mean	Std.	Median	t-value	p-value	z-value	p-value
REVENUE	214	3,833.57	4,488.26	2,259.28	428	4078.79	4424.33	2504.37	0.66	0.51	-1.26	0.21
COST	214	3,773.09	2,216.69	3,130.90	428	3609.78	2166.93	2896.00	-0.89	0.37	1.30	0.19
PROFIT	214	60.48	2,680.25	-579.83	428	469.00	2661.62	-304.81	1.83	0.07	-3.78	0.00
PROFRATE	214	0.84	0.44	0.75	428	0.95	0.45	0.85	2.85	0.00	-3.26	0.00
N_COUNTER	213	19.05	11.04	16.00	426	22.26	46.25	16.00	1.35	0.18	-0.42	0.67
POP	214	220.71	190.72	144.78	428	222.53	193.75	140.93	0.11	0.91	0.03	0.97
SERVICE_AREA	213	467.54	380.04	463.00	427	468.80	375.21	463.00	0.04	0.97	-0.09	0.93
N_DELPOINT	211	467.97	1,255.26	223.00	418	339.35	595.31	190.00	-1.41	0.16	0.81	0.42
N_BANK	212	70.37	55.00	55.00	420	70.85	60.25	52.50	0.10	0.92	0.60	0.55
DIST	212	37.47	45.81	31.00	424	35.07	44.83	24.50	-0.65	0.52	0.75	0.46
N_EMP	213	127.68	65.82	113.00	426	119.24	61.76	105.00	-1.59	0.11	1.78	0.08
N_DELSEC	213	57.14	28.87	49.00	426	55.48	29.63	48.00	-0.67	0.50	1.08	0.28

REVENUE= Revenue of a store (Million Korean Won)
COST= Operating cost of a store (Million Korean Won)
PROFIT= Operating profit of a store (Million Korean Won)
PROFRATE= Rate of profitability for a store measured as revenue divided by operating cost
N_COUNTER= Number of available counters of a store
POP= Population of the area that a store is responsible for (thousands)
SERVICE_AREA= Square kilometer of the area that a store is responsible for
N_DELPOINT= Average number of daily delivery points per mail carrier of a store
N_BANK= Number of financial institutions of the area that a store is responsible for
DIST= Average daily delivery distance per mail carrier of a store
N_EMP= Number of employees of a store
N_DELSEC= Number of delivery sections of a store

TABLE 3

Pearson Correlations between Variables

	REVENUE	COST	PROFIT	PROFRATE	N_COUNTER	POP	SERVICE_AREA	N_DELPOINT	N_BANK	DIST	N_EMP	N_DELSEC
REVENUE	**1.00**											
COST	0.90	**1.00**										
PROFIT	0.93	0.67	**1.00**									
PROFRATE	0.84	0.67	0.85	**1.00**								
N_COUNTER	0.31	0.32	0.26	0.23	**1.00**							
POP	0.62	0.65	0.50	0.74	0.23	**1.00**						
SERVICE_AREA	-0.40	-0.26	-0.45	-0.57	-0.09**	-0.49	**1.00**					
N_DELPOINT	-0.08*	-0.05 [NS]	-0.09**	-0.08**	-0.04	-0.07*	0.09**	**1.00**				
N_BANK	0.81	0.78	0.71	0.72	0.23	0.65	-0.31	-0.04 [NS]	**1.00**			
DIST	-0.32	-0.26	-0.32	-0.45	-0.05	-0.38	0.43	0.36	-0.27	**1.00**		
N_EMP	0.89	0.97	0.68	0.68	0.33	0.68	-0.24	-0.05 [NS]	0.80	-0.25	**1.00**	
N_DELSEC	0.80	0.87	0.62	0.64	0.32	0.74	-0.21	-0.05 [NS]	0.76	-0.22	0.91	**1.00**

See Table 2 for definitions of variables.

*,** indicates that the correlation coefficient is significantly different from zero at the 10%, 5% level (two-tailed test), respectively.

[NS] indicates that the correlation coefficient is not significantly different from zero at the 10% level (two-tailed test).

All other correlation coefficients are significantly different from zero at the 1% level (two-tailed test).

TABLE 4

OLS Regression Examining the Impact of Incentive Plan Change on Performance

(t-statistics in parentheses)

$$PERFORMANCE_{i,t} = \alpha_0 + \alpha_1 FYR\_PLAN_{i,t} + \alpha_2 SYR\_PLAN_{i,t} + \alpha_3 SERVICE\_AREA_{i,t} + \alpha_4 POP_{i,t} + \alpha_5 N\_BANK_{i,t} + \alpha_6 METRO_{i,t} + \alpha_7 N\_EMP_{i,t} + \alpha_8 DIST_{i,t} + \alpha_9 N\_DELPOINT_{i,t} + \alpha_{10} N\_COUNTER + \alpha_{11} N\_DELSEC + \varepsilon_{i,t}$$

Variable	Estimated Coefficients	
	PROFRATE	*MVOL*
INTERCEPT	0.540***	12.225***
	(17.89)	(308.12)
FYR_PLAN	0.088***	0.214***
	(3.99)	(7.20)
SYR_PLAN	0.159***	0.291***
	(6.88)	(10.38)
SERVICE_AREA	-0.0003***	-0.0004***
	(-8.31)	(-8.56)
POP	0.0005***	0.0009***
	(3.89)	(5.14)
N_BANK	0.002***	0.003***
	(6.84)	(9.25)
METRO	0.131***	0.378***
	(5.25)	(11.37)
N_EMP	0.001**	-
	(2.30)	
DIST	-0.001***	-0.001***
	(-4.01)	(-3.85)
N_DELPOINT	0.00001	0.00002
	(1.39)	(1.57)
N_COUNTER	0.0002	0.0006
	(1.11)	(1.38)
N_DELSEC	-0.001	-0.002**
	(-0.68)	(-2.01)
N	619	619
Adjusted R^2	0.76	0.78
F value of the regression	176.55	217.01

***,**,* Statistically significant at the 1%,5%,10% levels (two-tail),respectively

* Variable *PERFORMANCE* represents the rate of profitability (*PROFRATE)* or mail volume handled *(MVOL)*

FYR_PLAN= dummy variable representing the first year following RPE incentive plan implementation, taking the value 1 for 1998 and 0 otherwise

SYR_PLAN= dummy variable representing the second year following RPE incentive plan implementation, taking the value 1 for 1999 and 0 otherwise

METRO= dummy variable representing the area that a store is responsible for at year t, taking the value 1 for urban area, and 0 for otherwise

MVOL= log of mail volume handled per employee in a store

See Table 2 for other variable definitions.

*Reported t-statistics are based on the White (1980) heteroskedasticity-corrected standard errors.

TABLE 5

OLS Regressions Examining the Association between Degree of Common Uncertainty and Performance

(t-statistics in parentheses)

$$PROFRATE_{i,t} = \gamma_0 + \gamma_1 COMMON_{i,t} + \gamma_2 SERVICE\_AREA_{i,t} + \gamma_3 POP_{i,t} + \gamma_4 N\_BANK_{i,t} + \gamma_5 METRO_{i,t} + \gamma_6 N\_EMP_{i,t} + \gamma_7 DIST_{i,t} + \gamma_8 N\_DELPOINT_{i,t} + \gamma_9 N\_COUNTER + \gamma_{10} N\_DELSEC + \varepsilon_{i,t}$$

Variable	Estimated Coefficients
INTERCEPT	0.479***
	(8.84)
COMMON	0.022***
	(4.47)
SERVICE_AREA	-0.0003***
	(-6.67)
POP	0.0006***
	(5.30)
N_BANK	0.002***
	(6.00)
METRO	0.178***
	(5.18)
N_EMP	0.002**
	(4.17)
DIST	-0.001***
	(-3.37)
N_DELPOINT	0.00002
	(0.71)
N_COUNTER	0.0001
	(0.33)
N_DELSEC	-0.002**
	(-2.27)
N	410
Adjusted R^2	0.75
F value of the regression	127.25

***,**,* Statistically significant at the 1%,5%,10% levels (two-tail), respectively

*Reported t-statistics are based on the White (1980) heteroskedasticity-corrected standard errors.

COMMON= Adjusted R^2 ranking of a store obtained by running reference group-specific regression (ranking ranges from 1 through 9 with the higher value indicating greater R^2)

TABLE 6

OLS Regressions Examining the Impact of Incentive Plan Conditional on Dysfunctional Behavior

(t-statistics in parentheses)

Panel A: Use of 'CHANGE' variable to identify stores at which the level of dysfunctional behavior will be high (n=619)

$$PROFRATE_{i,t} = \delta_0 + \delta_1 FYR\_PLAN_{i,t} + \delta_2 SYR\_PLAN_{i,t} + \delta_3 FYR\_PLAN_{i,t} * CHANGE_{i,t}$$
$$+ \delta_4 SYR\_PLAN_{i,t} * CHANGE_{i,t} + \delta_5 SERVICE\_AREA_{i,t} + \delta_6 POP_{i,t} + \delta_7 N\_BANK_{i,t} + \delta_8 METRO_{i,t}$$
$$+ \delta_9 N\_EMP_{i,t} + \delta_{10} DIST_{i,t} + \delta_{11} N\_DELPOINT_{i,t} + \delta_{12} N\_COUNTER + \delta_{13} N\_DELSEC + \varepsilon_{i,t}$$

Variable	Estimated Coefficients
INTERCEPT	0.543***
	(18.18)
FYR_PLAN	0.096***
	(4.28)
SYR_PLAN	0.165***
	(7.00)
FYR_PLAN*CHANGE	-0.127***
	(-2.72)
SYR_PLAN*CHANGE	-0.104**
	(-1.98)
SERVICE_AREA	-0.0003***
	(-8.46)
POP	0.0005***
	(3.79)
N_BANK	0.002***
	(6.77)
METRO	0.131***
	(5.33)
N_EMP	0.001**
	(2.40)
DIST	-0.001***
	(-4.04)
N_DELPOINT	0.00001
	(1.53)
N_COUNTER	0.0001
	(1.05)
N_DELSEC	-0.0009
	(-0.71)
Adjusted R^2	0.76
F value of the regression	150.87

***,**,* Statistically significant at the 1%,5%,10% levels (two-tail), respectively

*Reported t-statistics are based on the White (1980) standard errors.

Panel B: Use of 'NOCOMP' variable to identify stores at which the level of dysfunctional behavior will be high (n=619)

$$PROFRATE_{i,t} = \delta_0 + \delta_1 FYR\_PLAN_{i,t} + \delta_2 SYR\_PLAN_{i,t} + \delta_3 FYR\_PLAN_{i,t} * NOCOMP_{i,t}$$
$$+ \delta_4 SYR\_PLAN_{i,t} * NOCOMP_{i,t} + \delta_5 SERVICE\_AREA_{i,t} + \delta_6 POP_{i,t} + \delta_7 N\_BANK_{i,t} + \delta_8 METRO_{i,t}$$
$$+ \delta_9 N\_EMP_{i,t} + \delta_{10} DIST_{i,t} + \delta_{11} N\_DELPOINT_{i,t} + \delta_{12} N\_COUNTER + \delta_{13} N\_DELSEC + \varepsilon_{i,t}$$

Variable	Estimated Coefficients
INTERCEPT	0.554***
	(19.08)
FYR_PLAN	0.126***
	(5.32)
SYR_PLAN	0.199***
	(8.04)
FYR_PLAN*NOCOMP	-0.171***
	(-5.99)
SYR_PLAN*NOCOMP	-0.181***
	(-5.97)
SERVICE_AREA	-0.0003***
	(-8.94)
POP	0.0004***
	(3.38)
N_BANK	0.002***
	(6.49)
METRO	0.127***
	(5.34)
N_EMP	0.001**
	(2.44)
DIST	-0.001***
	(-3.95)
N_DELPOINT	0.00001
	(1.34)
N_COUNTER	0.0002
	(0.92)
N_DELSEC	-0.0006
	(-0.47)
Adjusted R^2	0.77
F value of the regression	163.24

***,**,* Statistically significant at the 1%,5%,10% levels (two-tail), respectively
*Reported t-statistics are based on the White (1980) standard errors.

Panel C: Use of 'PROFRANK' variable to measure the level of dysfunctional behavior of store (n=619)

$$PROFRATE_{i,t} = \delta_0 + \delta_1 FYR\_PLAN_{i,t} + \delta_2 SYR\_PLAN_{i,t} + \delta_3 FYR\_PLAN_{i,t} * PROFRANK_{i,t} + \delta_4 SYR\_PLAN_{i,t} * PROFRANK_{i,t} + \delta_5 SERVICE\_AREA_{i,t} + \delta_6 POP_{i,t} + \delta_7 N\_BANK_{i,t} + \delta_8 METRO_{i,t} + \delta_9 N\_EMP_{i,t} + \delta_{10} DIST_{i,t} + \delta_{11} N\_DELPOINT_{i,t} + \delta_{12} N\_COUNTER + \delta_{13} N\_DELSEC + \varepsilon_{i,t}$$

Variable	Estimated Coefficients
INTERCEPT	0.551***
	(20.45)
FYR_PLAN	0.091***
	(4.62)
SYR_PLAN	0.141***
	(6.90)
FYR_PLAN*PROFRANK	-0.133***
	(-10.31)
SYR_PLAN*PROFRANK	-0.147***
	(-9.46)
SERVICE_AREA	-0.0003***
	(-10.07)
POP	0.0004***
	(3.47)
N_BANK	0.002***
	(5.45)
METRO	0.132***
	(5.98)
N_EMP	0.001**
	(2.56)
DIST	-0.001***
	(-3.43)
N_DELPOINT	0.00001
	(1.29)
N_COUNTER	0.0002*
	(1.78)
N_DELSEC	-0.0002
	(-0.17)
Adjusted R^2	0.81
F value of the regression	205.76

***,**,* Statistically significant at the 1%,5%,10% levels (two-tail), respectively

*Reported t-statistics are based on the White (1980) standard errors.

FYR_PLAN= dummy variable representing the first year following RPE incentive plan implementation, taking the value 1 for 1998 and 0 otherwise

SYR_PLAN= dummy variable representing the second year following RPE incentive plan implementation, taking the value 1 for 1999 and 0 otherwise

CHANGE= dummy variable representing stores that requested the change of their reference group in 2000, taking the value 1 and 0 otherwise

NOCOMP= dummy variable equal to 1 for representing stores which did not receive incentive bonus for three consecutive years (1998,1999,2000) and 0 otherwise

PROFRANK= standardized profitability rank of a store within its reference group at year t, with the higher value indicating lower relative profitability within its reference group

See Table 2 for other variable definitions.

TABLE 7
Regressions Examining the Association between Degree of Common Uncertainty and Dysfunctional Behavior (n=619)

Panel A: Logit Regressions Using CHANGE and NOCOMP as Dependent Variables

$$Prob(DYSFUNCTION) = F(\alpha_0 + \alpha_1 COMMON_{i,t} + \alpha_2 SERVICE\_AREA_{i,t} + \alpha_3 POP_{i,t} + \alpha_4 N\_BANK_{i,t} + \alpha_5 METRO_{i,t} + \alpha_6 N\_EMP_{i,t} + \alpha_7 DIST_{i,t} + \alpha_8 N\_DELPOINT_{i,t} + \alpha_9 N\_COUNTER + \alpha_{10} N\_DELSEC)$$

	Dependent Variables			
	CHANGE		*NOCOMP*	
Independent Variables	*Estimated Coefficients*	*Wald Chi-Square*	*Estimated Coefficients*	*Wald Chi-Square*
INTERCEPT	-0.267	0.147	-0.470	1.200
COMMON	-0.204	7.249***	0.006	0.018
SERVICE_AREA	-0.002	6.277**	-0.001	5.523**
POP	-0.009	10.905***	-0.005	14.427***
N_BANK	-0.011	2.066	-0.010	5.810**
METRO	0.508	0.847	0.118	0.152
N_EMP	0.027	6.153**	-0.001	0.067
DIST	-0.001	0.079	0.003	0.993
N_DELPOINT	0.0001	0.328	-0.0005	3.070*
N_COUNTER	-0.139	7.257***	-0.0001	0.0002
N_DELSEC	0.013	0.491	0.024	5.030**
Pseudo R^2	0.12		0.09	
Percent Correctly Predicted	75.1%		70.1%	

* Variable DYSFUNCTION represents indicator variable CHANGE or NOCOMP
***,**,* Statistically significant at the 1%,5%,10% levels (two-tail), respectively

Panel B: OLS Regression Using PROFRANK as a Dependent Variable

$$PROFRANK_{i,t} = \delta_0 + \delta_1 COMMON_{i,t} + \delta_2 SERVICE\_AREA_{i,t} + \delta_3 POP_{i,t} + \delta_4 N\_BANK_{i,t} + \delta_5 METRO_{i,t} + \delta_6 N\_EMP_{i,t} + \delta_7 DIST_{i,t} + \delta_8 N\_DELPOINT_{i,t} + \delta_9 N\_COUNTER + \delta_{10} N\_DELSEC + \varepsilon_{i,t}$$

Variable	Estimated Coefficients
INTERCEPT	0.646***
	(3.95)
COMMON	-0.079***
	(-4.89)
SERVICE_AREA	-0.0003**
	(-2.23)
POP	-0.002***
	(-3.83)
N_BANK	-0.006***
	(-6.16)
METRO	-0.056
	(-0.53)
N_EMP	-0.003*
	(-1.67)
DIST	0.002*
	(1.93)
N_DELPOINT	-0.0001
	(-1.44)
N_COUNTER	0.001**
	(2.26)
N_DELSEC	0.018***
	(4.02)
Adjusted R^2	0.14
F value of the regression	11.31

***,**,* Statistically significant at the 1%,5%,10% levels (two-tail), respectively

COMMON= Adjusted R^2 ranking of a store obtained by running the following reference group-specific regression (ranking ranges from 1 through 9 with the higher value indicating greater R^2)

$$PROFRATE_{i,t} = \beta_0 + \beta_1 SERVICE\_AREA_{i,t} + \beta_2 POP_{i,t} + \beta_3 NO\_BANK_{i,t} + \beta_4 NO\_EMP_{i,t} + \beta_5 DIST_{i,t} + \beta_6 NO\_DELPOINT_{i,t} + \beta_7 NO\_COUNTER_{i,t} + \beta_8 NO\_DELSEC_{i,t} + \varepsilon_{i,t}$$

CHANGE= dummy variable representing stores that requested the change of their reference group in 2000, taking the value 1 and 0 otherwise

NOCOMP= dummy variable equal to 1 for representing stores which did not receive incentive bonus for three consecutive years (1998,1999,2000) and 0 otherwise

PROFRANK= standardized profitability rank of a store within its reference group at year t, with the higher value indicating lower relative profitability within its reference group