The Effect of Prior Performance Information on Leniency in Subjective Performance Evaluation

Tae Sik Ahn
College of Business Administration
Seoul National University

Jeong-Hoon Hyun
College of Business Administration
Seoul National University
hjh275@snu.ac.kr

May, 2014
The Effect of Prior Performance Information on Leniency in Subjective Performance Evaluation

ABSTRACT:

This paper examines whether supervisors’ leniency in subjective performance evaluation is influenced by the prior performance information of subordinates. While prior studies have used proxies for leniency based on aggregate objective performance level or median of subjective scales, we develop the proxy for leniency at the performance measure level based on prior performance level which is found to be more relevant to the current subjective performance score. We analyze archival performance evaluation data of multiple state-owned enterprises (SOEs) in Korea over multiple time periods which enables us to develop the new definition of leniency. Utilizing this data set, we empirically find that lenient rating persists over time, and that larger amount of leniency is applied to low prior performers and harshly evaluated performers in the previous period and smaller amount of leniency is applied to high prior performers and leniently evaluated performers in the previous period. The results have important implications for understanding the incentives of raters which differ with respect to their previous performance information.

Keywords: Performance evaluation, Subjective measure, Prior performance, Asymmetric leniency.

Data Availability: Data are collected from public sources indicated in the text.
I. INTRODUCTION

A long body of research in performance evaluation has investigated costs and benefits of subjective performance evaluation. Subjective performance evaluation provides useful performance information of an organization when objective measures are inadequate, incomplete, and prone to manipulation (Banker and Datar 1989; Baker et al. 1994; Bushman et al. 1996; Gibbs et al. 2004; Holmström and Milgrom 1991; Murphy 1999). Despite its importance and benefits, subjectivity has problems that supervisors make unfair and biased judgments (Baker et al. 1988; Bol 2008; Prendergast 1999). Research in psychology, organizational behavior, and accounting finds that subjectivity suffers from severe leniency and compression effects (Bol 2011; Bol et al. 2013; Moers 2005; Prendergast and Topel 1993). However, due to limited access to data, little research investigates why both effects occur and how prior performance rating affects the decisions of lenient ratings.

Leniency effect is defined as the tendency that evaluators provide performance ratings higher than those warranted by ratee performance (Saal and Landy 1977). However, it is difficult for raters to observe warranted performance of ratees. To estimate the warranted performance, prior empirical research employs indirect benchmarks, such as the median of subjective rating scales or the objective performance score (Bol 2011; Bretz et al. 1992; Merchant et al. 2010; Moers 2005). In subjective performance evaluations, however, all ratees perform at the different level according to their resources and their abilities (Murphy and Balzer 1989), which is inconsistent with the assumption of benchmark of the median of rating scale. Also, the objective performance score may not be an appropriate proxy for the subjective performance dimension because one of the most important roles of subjective performance measure is to modify or supplement the objective performance score (Merchant et al. 2010). Objective and subjective performance measures are supposed to represent different perspectives on performance.
Meanwhile, prior success (failure) directionally affects current expectation of success (failure) (Feather 1966). It is difficult for raters to determine whether the ratees’ current performance is clearly discrepant from previous performance score (Smither et al. 1988). When the behavior of the ratees is consistent with the raters’ expectations, it is automatically stored (Feldman 1981), and revealed in forms of serially correlated evaluation score. Therefore raters and ratees will create their expected score of a specific performance measure based on their prior performance level. The prior performance score of a specific measure is perceived as more relevant benchmark to its current performance (e.g., Thorsteinson et al. 2008; Woods 2012) than aggregated level of overall objective performance measures when both prior performance score of the same specific measure and overall objective performance level are available, which is also supported in our empirical setting. Therefore, to estimate leniency, prior performance level can be a better benchmark not only for the raters but also for ratees. However, most of prior studies rely on single-year (Moers 2005), or two-year samples (Bol 2011; Merchant et al. 2010), making it almost impossible to use the prior performance information as the benchmark. We develop the proxy for leniency at the performance measure level based on prior performance level by collecting and utilizing performance evaluation data of multiple years for multiple SOEs in Korea.

In addition to the leniency measurement issue, if the leniency is expected in most subjective evaluations, the following questions will be interesting: Is leniency a one-time event or can it persist over time? If the prior subjective performance score is high, then is current performance score evaluated leniently? Once prior period subjective evaluation is lenient, then does the next period evaluation become harsh? All of the above questions cannot be explored explicitly without the time-series data. In this paper, we examine whether leniency persists over time and how prior performance information affects the incentives of raters. In doing so, we aim to provide a deeper understanding about the leniency.
When raters bias their rating toward leniency, they may estimate the expected utility of lenient rating against the disutility of the accurate rating (Murphy and Cleveland 1995, p.243). Prior studies show that raters have more incentives to rate inaccurately than accurately when they are neither penalized for biased rating nor rewarded for accurate rating (Spence and Keeping 2011), and that the incentives and characteristics of raters do not change easily. Thus, we expect that the leniency persists over time unless there is a significant change in reward (or penalty) structure for rating accuracy.

Supervisors have their own incentives to inflate ratings for low or harshly evaluated performers in the prior period so that they avoid conflicts with subordinates (Levy and Williams 2004; Mitchell and O’Reilly 1983). Also, the raters are likely to show larger leniency to poor performers than to high performers because motivation effect is greater for poor performers (Bol 2011; Kane 1994; Woods 2012). If we assume the concave utility functions of the ratees as performance score rises, the sensitivity of utility to lenient or harsh rating will be higher for the low prior performers than high prior performers. Therefore we can predict that raters show larger amount of leniency for low prior performers (or harshly evaluated performers) and smaller amount of leniency for high prior performers (or leniently evaluated performers) because raters want to reduce overall confrontation costs by maximizing total utility of ratees without violating the grading distribution in the evaluation guideline. Our findings confirms the asymmetry of lenient ratings depending on the level of prior performance and the degree of prior leniency. This findings lead to additional insights that asymmetric lenient rating for low performers over time can increase the average subjective performance scores and decrease the score gap between superior performer and poor performer, resulting in both leniency and compression effects. This finding shows that the asymmetry of leniency bias is one of the contributing factors to compression bias for
subjective measures. In other words, compression bias is a result of asymmetric leniency bias over time.

Our study contributes to the literature in the following ways. First, we redefines the proxy for leniency based on prior performance level while prior studies have used proxies based on aggregate objective performance level or median of subjective scales. The multiple year and multiple SOE performance evaluation data make it possible to develop our new definition of leniency. We construct an alternative proxy for the degree of leniency in micro-level. The proxy for leniency is based on modified previous performance score and reflects the expected performance level of both raters and ratees in multiple period setting. Although contract theory presents that performance is a function of both ratees’ effort and noise (bias), the leniency proxies in prior studies have not explicitly discriminated between ratees’ effort part and rating bias part. This paper fills this void by excluding the ratee effort portion from our leniency proxy and therefore focusing on systematic bias – leniency effect itself.

Second, this study contributes to the literature by examining the time-series pattern of leniency from the perspective of prior performance information. For example, we analyze the time series pattern of leniency responding to the necessity of time-series data study (Bol 2011; Moers 2005), and find that leniency effect persists over time. These findings increase the pervasiveness of leniency effects over time across firms. Furthermore, we extend previous research by analyzing the effects of prior performance information on the degree of raters’ leniency. Our findings show asymmetric leniency depending on the prior performance level and the prior leniency tendency. This indicates that the supervisors manage their dual roles as evaluators and motivators and adjust the ratings differently conditional on the ratees’ past performance level. In short, our study shows that the consideration of the prior performance information is crucial in subjective evaluation, and suggests that the incentives of raters might
differ with respect to their performance information set such as prior performance level and prior leniency level.

The remainder of the paper is organized as follows. We review the theoretical background and develop our hypotheses in Section II. In Section III, we provide institutional background on performance evaluation systems of SOEs in Korea. In Section IV, we describe research design including sample selection and empirical measures. Our primary empirical results are provided in Section V. Section VI concludes.

II. THEORY AND HYPOTHESIS DEVELOPMENT

Background

If the adoption of objective performance measures in incentive contracts is effective by itself in incorporating all efforts of ratees properly, any other performance signal should not be additionally valuable for contracting purposes (Holmström 1979). However, objective performance measures are often inadequate, incomplete, and prone to manipulation (Murphy 1999). Thus, to compensate for the weakness of the objective performance measure, subjective performance measure is usually combined, because it provides a more aligned incentive and reduce uncertainty to ratees by filtering out uncontrollable factors mixed in objective measures (Baker et al. 1994; Bushman et al. 1996).\(^1\) Despite its benefit, subjective performance evaluation also has a number of problems. One example is the lenient rating. Prior studies in various disciplines such as psychology, organizational behavior, and accounting find that subjective performance evaluation suffers from severe leniency bias, the

---

\(^1\) Bol (2008) shows three different types of subjectivity: (1) using subjective performance measure, (2) allowing for flexible weighting, and (3) allowing for ex post discretion in bonus. The setting of this paper naturally controls two latter ones because the performance evaluation system prohibits supervisors from arbitrarily adjusting weight and bonus.
tendency of evaluators to provide performance ratings higher than those warranted by ratee performance (Bol 2011; Bretz et al. 1992; Moers 2005; Saal and Landy 1977).²

Hypothesis

Persistence of Lenient Rating

Performance evaluation literature has been premised on the assumption that rating biases come from supervisors’ cognitive limitations (Murphy et al. 1982). Supervisors tend to focus more on a limited set of performance dimensions such as common measures and financial outcomes and less on unique measures and financial inputs (Ittner et al. 2003; Lipe and Salterio 2000). Specifically, research on assimilation effects and anchoring effects has shown that prior performance evaluation serves as an anchor and affects current performance rating (Huber 1989; Smither et al. 1988; Woods 2012). These studies provide evidence that cognitive limitations can deteriorate the full benefits of using subjectivity in contracting, on the assumption that the goal of raters is to rate their ratees accurately. However, recent studies argued that accuracy may not be the sole goal of raters and explain that other rater incentives cause leniency bias.

Supervisors inflate ratings of subordinates to serve their self-interests (Banks and Murphy 1985; Fried and Tiegs 1995; Longenecker et al. 1987; Spence and Keeping 2011). The probability of lenient rating is determined by the difference between raters’ expected utility of biased rating and that of accurate evaluation (Murphy and Cleveland 1995). If supervisors are not rewarded for accurate ratings or punished for biases, supervisors may have insufficient motivation to invest time in gathering information (Bol 2011). Also, supervisors prefer to maintain a good relationship with their subordinates, or to minimize confrontation costs (Harris 1994; Murphy and Cleveland 1995), and tend to favor specific employees from...
political considerations (Prendergast and Topel 1993). Especially, when the rewards of supervisors are tied to the performance of subordinates, rating inflation occurs more frequently (Ilgen et al. 1981; Prendergast and Topel 1996). Golman and Bhatia (2012) also presents asymmetric utility of managers as a source of leniency, suggesting that supervisors feel worse about unfavorable errors than about favorable errors.

While we expect the leniency in most subjective evaluations, only a few studies examine the time series tendency of leniency. Conducting three experiments with two-period settings, Kane et al. (1995) find that leniency tendency exist in both periods and their correlations are significant across all three experiments. Dalla Via et al. (2011) empirically examine the difference between the ratings assigned to development purposes and those for administrative purposes by using the multi-period setting. Although both studies utilize the multi-period sample, they do not explicitly analyze the pattern or the extent of leniency over time. In this study, we try to directly analyze the extent of leniency variations during the evaluation period.

The incentives of raters do not change easily unless the incentives of accurate evaluation exceeds those of lenient evaluation. Without serious penalty for inaccurate rating or reward for accurate rating, the leniency effects become habitual (Murphy and Cleveland 1995). Therefore, unless there is a systematic change in evaluation guidelines against leniency overturning the utility structure of raters and ratees, we may expect the leniency persist over time. We then set up the hypothesis as follows.

**H1**: The leniency tendency persists over time.

**Asymmetric Lenient Rating**

Supervisors’ rating incentives theory features more (less) leniency for low (high) performers and harshly (leniently) evaluated performers. In other words, asymmetric lenient rating is an outcome of self-interested rating errors because it avoids conflicts with workers (Levy and Williams 2004). If we assume a typical concave utility function for a risk-averse
subordinate, marginal utility will decline as performance score rises with respect to previous performance level. Then, given the limited authority to adjust ratings and suggested rating guidelines, supervisors are more likely to inflate scores of low performers than those of high performers. Intuitively, the marginal utility of one level change in performance score is greater for low performers than high performers. Accordingly, raters are likely to inflate (deflate) ratings of low (high) performers more because the sensitivity of utility to lenient or harsh rating is greater for low past performers given the constraints of score distribution. Raters can maximize the total utility of ratees reducing the total potential confrontation costs with ratees. Also, supervisors prefer to avoid confrontation with poor performers (Mitchell and O’Reilly 1983), who tend to be less acceptable to accurate rating (Murphy and Cleveland 1995). Thus, supervisors are more likely to evaluate low performers more leniently and extend good relationships with ratees. Second, supervisors may intentionally encourage poor performers by inflating ratings than deserved, in order to influence ratees’ future behavior (Bol 2011; Kane 1994; Woods 2012). Motivating low performers is considered an essential role of performance evaluation (Murphy and Cleveland 1995). Telling a poor worker that he or she is performing poorly will demotivate and discourage him or her (Longenecker et al. 1987). Therefore, targeted inaccuracy is realized as the pattern of asymmetric leniency where supervisors evaluate low performers more leniently than high performers.

If we view leniency as a result of unconscious inferential processes, however, raters’ asymmetrical perception about prior score can drive asymmetric leniency. Both high performance score and low performance score produce anchoring effects when people are influenced by previous performance information that serves as an anchor for judgments (Chapman and Johnson 1999; Tversky and Kahneman 1974). But the anchoring effect is

---

4 According to Tversky and Kahneman (1974), the anchoring effect is the disproportionate influence on judgment which is biased toward the initially presented value. They suggest that people make insufficient adjustments to yield a final estimation based on the anchor. In this paper, we limit the concept of the anchoring
larger for high performance scores than low performance scores because raters perceive low score as an extreme anchor (Thorsteinson et al. 2008). That is, the plausibility of an anchor determines the magnitude of anchoring effects (Blankenship et al. 2008; Wegener et al. 2001). Consequently, high performance score in the previous period is more likely to persist, while low performance score is more likely to be reverted to the higher score. These unintended cognitive limitations may lead to asymmetric leniency effects.

With respect to leniency in the previous period, when ratees fail to receive expected rating, they perceive that the raters are responsible for their receiving a harsh evaluation. If subordinates believe that rewards or performance score have not been distributed equitably, the supervisor might have a strong incentive to restore perceptions of equity (Murphy and Cleveland 1995). A possible solution to restore them is to give lenient rating in the current period to the harshly evaluated subordinates in the prior period, hence raters can reduce confrontation costs of harshly evaluated ratees in the prior period.

In sum, both theories of intended and unintended rating inflation predict that raters are likely to be more lenient toward low past performers or harshly evaluated subordinates than toward high past performers or leniently evaluated ones.

**H2a:** The leniency tendency is greater for low past performers than high past performers.

**H2b:** The leniency tendency is greater for harshly evaluated performers in the previous period than leniently evaluated performers.

### III. INSTITUTIONAL BACKGROUND

**Overview of Performance Evaluation Systems of SOEs in Korea**

The Korean government enacted the Law for Management of SOE in 2003 that requires SOE be evaluated annually by a group of auditors assigned by the government (Ministry of Strategy and Finance) and disclose the result of performance evaluation by June of the
subsequent year. The government develops performance measures with which the rater and the ratee mutually agree, and distribute the guideline by the end of the year for evaluation. Then, at the end of the March, SOEs submit performance reports on the subsequent year. From April to June, raters evaluate the performance of the SOEs.

The performance evaluation system employs an incentive bonus plan in which bonus size is based on the results of performance evaluation. The bonus size is limited and determined by ranking-based peer comparison via distribution analysis. Thus, it is important to compare relative scores among peers. To provide ratees (SOEs) with appropriate incentives and sound evaluation systems, the Korean government has developed and actively amended the performance measurement structure for SOEs. It uses various measures under three categories: overall management, main business, and business management. Panel A in Table 1 presents an example of a performance rating of an SOE in 2005.

The third column of Panel A in Table 1 shows that each evaluation criterion classifies performance measures into objective and subjective measures. Objective measures are assessed using four quantitative, formula-based methods: actual-to-target analysis, target-range assignment analysis, trend analysis, and beta analysis. In contrast to objective measures, subjective measures have a single evaluation method: grading. While objective

---

5 After all the evaluations, the government transforms total performance score of SOEs into standardized Z-score. If Z-score of an SOE is above 2σ, then the SOE is graded ‘S’ and paid the maximum 500% bonus of monthly salary.

6 (1) Actual-to-target analysis represents actual performance divided by target performance. Sales volume, labor expenses, and plant construction progress are examples of measures that use the actual-to-target method. (2) Target-range assignment analysis uses the ratio of actual performance minus minimum target performance to maximum target, minus minimum target performance. Capital productivity, capacity utilization, and customer satisfaction index use this evaluation method. (3) Trend analysis is a regression analysis that computes standard performance coefficients using past actual performance (e.g., the prior 15 years’ figures). Actual performance is evaluated against expected performance via the standard performance coefficient. This method is, in general, used to assess how effectively SOEs manage their inventory, cost of capital, cost of goods sold, and administration expenses. (4) Beta analysis is similar to trend analysis in using past data to obtain a benchmark to assess current performance; however, it uses the beta distribution instead of the regression. Typical measures that use this method are labor productivity, economic value added (EVA), and plant power management. After conducting the quantitative assessment, each SOE is assigned into five or nine grades, in accordance with the predetermined score ranges.

7 Subjective measures complement objective measures for activities that are difficult to quantify, but that are important to achieving firm strategic goals—for example, efficient management in strategic plans, improvement
measures are calculated by comparing actual performance based on predetermined formula for benchmark performance, assigning subjective measures to the various grade levels depends entirely on the rater’s subjective evaluation with a guideline. The guideline for subjective evaluation can be summarized into two rules. First, the guideline suggests that subjective performance should be rated scored on 50 percent of performance level and 50 percent of performance improvement against the past performance level. This rule lets raters to compare current performance to previous performance for the measure. Second, it is strongly recommended that subjective performance score be distributed as following: A+/A0: 10%, B+: 10%, B0: 20%, C: 40%, D+/D0/E+/E0: 20%. This distribution guideline is applied to each SOE class and performance category.

Formation of Raters

All SOEs are categorized into 8 classes based on their business characteristics – SOC, Service, Inspection management, Culture, Industry promotion I, Industry promotion II, Education, and Pension, as shown in Panel B in Table 1. Each class is audited by four groups of rating committee – one objective measure team and three subjective measure teams that grade overall management, main business, and business management. For performance evaluation of SOEs, rating committees consist mainly of academic professors, certified public accountants, and industrial experts. Raters are allocated with matrix structure in order to harmonize between class characteristics and measurement characteristics. For a measurement group of common measures (overall management and business management), raters consist of experts in organization management such as accounting, public administration, business, and economics. On the other hand, for a measurement group of unique measures (main business), raters are mostly professionals with experiences in the ratees’ industries.

---

in control systems, appropriateness in assessment and implementation of investments, development of organizational culture, appropriateness of budgeting, cooperation with community, and employee education.

8 Panel B in Table 1 provides an example of composition of raters in 2008. The number of classification category varies six to nine as the rule changes.
The government also provides various safeguards against rater bias. First, if a rater is assigned to a certain rating committee in the current year, the rater would usually move to a different committee every year to promote independence (i.e. rotation system). Second, after three years of rating service, the rater should take one year’s leave of absence. Moreover, raters should sign a code of ethics in evaluation and receive a training about how to deal with interest conflicts. All these rules guarantee independence and fairness of evaluation process and reduce familiarity risk.

Raters can perform an additional on-the-spot inspection to gather supplementary information. After the inspection, score rating is normally decided unanimously by all committee members. Ratees have opportunities to express their opinions about the processes and the results of evaluation. Although the government provides the rating committee with a checklist to facilitate the subjective evaluation, it is likely to be vulnerable to leniency bias, as discussed in Section IV.

IV. RESEARCH DESIGN

Sample Selection

Our data are collected from the Korean Government’s annual reports on its performance evaluations for SOEs from 2005 to 2011. The reports have been released since 2004, and the first-year data are used as prior performance information in our sample. We also exclude newly adopted measures due to the construction process of proxy for leniency. After imposing the data requirements for the leniency computation, we obtain a final sample of

---

9 The Korean government enacted the Law for Management of SOEs in 1984. External raters have evaluated the performance of 13 SOEs since 1985. However, this data of performance results makes it difficult to compare inter-temporal rating tendency because (1) before and after year 2004, there is a big difference in rating systems that can mislead the effects of leniency over time, (2) in 1998, the minimum base score at 75% was eliminated, and hence there is a large difference in score rates between before-1998 and after-1998 (Ahn et al. 2010).

10 Although we do not directly use the sample of new measure in our analysis, we utilize the distribution of new measures when we construct the alternative leniency proxies.
7,470 measure-SOE-year observations, representing 109 Korean SOEs.

**Leniency Model**

We develop a cross-sectional model for leniency and estimate Equations (1) and (2) using OLS to test if the past performance level asymmetrically affects the raters’ likelihood of lenient rating:

\[
LEN_{ijt} = \alpha_0 + \alpha_1 \text{SUB}_{ij,t-1} \text{ or } LEN_{ij,t-1} + \alpha_2 \text{UNIQUE}_{ijt} + \alpha_3 \text{WEIGHT}_{ijt} + \alpha_4 \text{EXPERIENCE}_{jt} + \alpha_5 \text{WOMAN}_{jt} + \alpha_6 \text{SIZE}_{jt} + \text{Fixed Effects} + \epsilon_t. \tag{1}
\]

\[
LEN_{ijt} = \alpha_0 + \alpha_1 \text{High SUB}_{ij,t-1} \text{ or } \text{High LEN}_{ij,t-1} + \alpha_2 \text{Low SUB}_{ij,t-1} \text{ or } \text{Low LEN}_{ij,t-1} + \alpha_3 \text{UNIQUE}_{ijt} + \alpha_4 \text{WEIGHT}_{ijt} + \alpha_5 \text{EXPERIENCE}_{jt} + \alpha_6 \text{WOMAN}_{jt} + \alpha_7 \text{SIZE}_{jt} + \text{Fixed Effects} + \epsilon_t. \tag{2}
\]

where \(i, j,\) and \(t\) indicate performance measure \(i\), firm \(j\), and year \(t\), respectively.

All regressions in this paper are estimated with Huber-White robust standard errors clustered by performance measure level. The standard errors are robust to both serial correlation and heteroskedasticity (Rogers 1993). We build upon related literature on both leniency and ratees’ effort allocation, in order to construct the measures used in our empirical tests. First, we define our proxy for leniency and describe the factors that prior research suggests could affect leniency effects at firm-level, rater-level, and individual measure-level.

**Leniency Proxy**

Leniency effect is the tendency of evaluators to provide performance ratings higher than those warranted by rateee performance (Saal and Landy 1977). However, as stated before, it is difficult for raters to observe warranted performance of ratees. In this study, we assume that the warranted performance of subjective performance reflects the expected performance score of the raters and ratees at performance measure level. Ratees who receive positive (negative) prior performance information will have the expectation of favorable (unfavorable) ratings than the ratees who receive no prior performance information (Salvemini et al. 1993).
Therefore, the expected performance score is mainly based on the previous performance level (Feather 1966; Feldman 1981; Smither et al. 1988).

Contract theory assumes that performance equals to effort plus noise: \( q = a + e \), where \( q \), \( a \), and \( e \) refers to performance, ratee’s effort level, and noise, respectively. The performance \( q \) can be transformed into performance score rate (\( SUB_{Attain} \)) minus base performance, or ‘benchmark’ (\( BENCHMARK_t \)). Meanwhile, noise term \( e \) can be decomposed into two – systematic noise such as leniency bias (\( LEN_t \)) and random noise \( \varepsilon \) which is assumed to be normally distributed with zero mean and variance \( \sigma^2 \). Moreover, change in objective performance (\( \Delta OBJ_{Attain} \)) can be a proxy for effort level \( a \) (Ahn et al. 2010; Bol 2011). In sum, we calculate the leniency proxy at the individual measure level as in Equation (3).

\[
LEN_t = SUB_{Attain} - BENCHMARK_t - \Delta OBJ_{Attain}.
\] (3)

Here, the variable \( BENCHMARK_t \) is supposed to represent an warranted performance level although it cannot be directly measured. Prior studies use median of subjective scale (i.e. performance score “C” in our setting) or overall objective performance score as a benchmark (e.g. Bol 2011; Bretz et al. 1992; Moers 2005). First, the benchmark of the median scale is based on the assumption that true mean level of performance corresponds to the scale midpoint. This assumption implies that all ratees perform at the same level regardless of their resources and their abilities (Murphy and Balzer 1989). However, this assumption seems illogical because (1) it is entirely possible that ratee A is better performer than ratee B depending on their efforts and abilities, (2) subjective performance evaluation is normally designed to produce a skewed distribution of performance on purpose (O’Boyle and Aguinis 2012).11 Second, the leniency measure in Bol (2011) is defined as the difference between

---

11 Our sample, on average, shows attainment ratio of 0.109 above the median value (grade C). This descriptive statistic implies that our sample shows leniency pattern and negatively skewed distribution (skewness = -0.244). Also, the forced distribution rule admits the skewed distribution. Therefore, our benchmark for leniency based on prior performance already incorporates skewness (Bretz et al. 1992; Kane et al. 1995; Ng et al. 2011). Our leniency proxy (\( LEN \)) captures the degree of leniency above guaranteed skewness which implies our leniency proxy is derived based on the assumption that performance score does not follow normal distribution.
objective performance score and subjective performance score, which depends on the assumptions that the variance in the objective performance score is similar to the true performance variance and that performance standards are similar for the objective and subjective performance. However, this assumption defeats the purpose of the introduction of subjective performance measure whose role is to mitigate incentive distortions. While, objective performance measure is often incomplete, imperfect, and prone to manipulation (Banker and Datar 1989; Bushman et al. 1996; Holmstrom and Milgrom 1991), the introduction of subjective performance measure can improve incentive contracting because it allows value-enhancing efforts which is not captured by objective performance measure (Bol 2008). Therefore, it is not appropriate to use objective performance level as a proxy for warranted performance level of a subjective performance measure. Moreover, we estimate leniency at an individual subjective performance measure level. Overall objective score cannot be used as overall benchmark for every subjective performance measures in calculating leniency of individual subjective performance measures.

In performance evaluations in a time series setting as in our study, prior performance level will play an important role. Raters may refer to the prior performance level of a specific performance measure in forming an opinion and providing a final evaluation point. Appendix A shows empirical results that current subjective performance score is significantly affected by prior subjective performance score, not by objective performance score at both performance measure level and SOE level tests.\textsuperscript{12} Evaluation guideline also suggests that raters should compare the current performance with the last year performance in evaluation. In other words, main reference point or benchmark for evaluation of current performance of a specific subjective performance measure should be prior performance level reflected in a performance score. Time-series data allows us to identify prior performance level as a

\textsuperscript{12} We also compare the relevance effect of prior performance level to the current performance score with that of other benchmarks, and find that the prior performance level is more relevant to the current performance score than other benchmarks such as total prior performance and prior objective performance level in Appendix A.
Although the prior performance score can be an improved benchmark, it also has a shortcoming because this benchmark implies that the average current rating score is just as the same as the prior performance score. If we regard the absolute score of prior performance level as the benchmark without any adjustment, the expectation level of raters and ratees can deviate from prior performance level. For example, very poor prior performers with ‘zero’ score are assumed to have expectation level of prior performance score (i.e. zero score), again. However, they usually expect the “mean reversion” tendency of performance score. In other words, they expect that their zero past performance score is more likely to be reverted to more higher score level than zero. If they are evaluated as zero score in the current period again, they perceive the rating is very severe. We attempt to develop a benchmark performance measure in our study, exploiting the time-series performance score data.

In our setting, “mean reversion” is partly incorporated by requesting the combination of 50 percent midpoint scale and 50 percent prior performance score which is based on the performance guideline of SOE. For example, the benchmark for the poor prior performers with zero score can be the sum of a half of midpoint (0.5) and a half of prior score (0.0). Similarly, the benchmark for the superior prior performer with 1.00 score can be 0.75 (= 1.00 * 50% + 0.50 * 50%). These combination reflect the rating distribution of ratees with top and bottom scores better.

In summary, prior subjective performance score in its raw form (SUB_Attaint-1) or adjusted form (BENCHMARKt), not the aggregate objective score (OBJ_Attain), is the

---

13 From the ratees point of view, they are also concerned with current performance score compared with the last year performance. Especially in our performance evaluation setting, SOEs assign the responsibility to their employees for individual performance measure. Employees assigned a specific responsibility for improving that performance score will analyze the prior performance evaluation report by the raters, remedy the shortcomings mentioned in the report, develop some positive actions to improve the current situations. All these effort are directed to increase the performance score. Those employees in charge of a specific performance measure want to influence the performance score by actions mentioned above but also by generating the polished performance report regarding that performance dimension. In all these efforts by raters and ratees, prior performance score will provide the most important benchmark.
critical benchmark in analyzing the persistence of leniency over time and the effect of prior performance information on leniency. In SOE setting, raters are recommended to follow the well-documented rating guideline that (1) raters should give 50 percent weight on absolute performance level and 50 percent weight on performance improvement from the prior year performance level, and (2) final performance rating is strongly required to follow a forced distribution (A+/A0: 10%, B+: 10%, B0: 20%, C: 40%, D+/D0/E+/E0: 20%). Therefore, we apply these two conditions in the guideline to estimate BENCHMARK. The example of estimation process of BENCHMARK and LEN is shown in Appendix B.

This leniency proxy has intuitive appeal as it can observe biases when the actual rating distribution is deviated from the assumed distribution for the job based on the prior performance level. We calculate the leniency proxy based on the assumptions and findings that subjective performance measure tends to be independent from objective performance measure (Merchant et al. 2010), and that adjusted prior performance score, on average, stands for the assumed distribution of job at the performance measure level. Also, low (high) past performance level leads to higher (lower) level of BENCHMARK than its original performance level. Therefore, the ratees’ expectation based on “mean reversion” can be also incorporated into the measure in forms of BENCHMARK and effort level (∆OBJ.Attain). Moreover, our leniency proxy can be generated from an individual performance measure, making possible the analysis of the incentives of both raters and ratees at the performance measure level.

Factors Influencing the Leniency

Prior performance level (SUB.Attain) is an important factor not only in measuring

---

14 As stated in Section III, there are separate rating committee – one objective measure team and three subjective measure teams that grade overall management, main business, and business management. This setting conceptually prevents from spillover bias which means raters’ subjective evaluations are directionally influenced by an accompanying objective performance score (Bol and Smith 2011). Also, we empirically find that spillover bias is not presented in our setting. Rather, current subjective score is mainly determined by prior subjective score as shown in Appendix A.
leniency but also in influencing leniency. The test variables of interest are the past performance level – $\text{SUB}_{\text{Attain}_{t-1}}$, $\text{Low}_{\text{SUB}_t}$, and $\text{High}_{\text{SUB}_t}$, and the past leniency tendency – $\text{LEN}_{t-1}$, $\text{Low}_{\text{LEN}_t}$ and $\text{High}_{\text{LEN}_t}$. The independent variable, $\text{Low}_{\text{SUB}_t}$ ($\text{High}_{\text{SUB}_t}$) is an indicator variable that equals to one if the score rate belongs to the lowest (highest) quartile of $\text{SUB}_{\text{Attain}}$ in the previous year, and zero otherwise. $\text{Low}_{\text{LEN}_t}$ ($\text{High}_{\text{LEN}_t}$) is an indicator variable that equals to one if the score rate belongs to the lowest (highest) quartile of $\text{LEN}$ in the previous year, and zero otherwise. The model also includes a comprehensive set of control variables. $\text{UNIQUE}$ is an indicator variable that equals one if a performance measure is unique, and zero if it is common. Unique measures are cognitively more difficult to evaluate than common measures (Lipe and Salterio 2000). In addition, some dimensions of performance in public organizations are often hard to measure (Burgess and Ratto 2003). Therefore, raters are more likely to inflate ratings for unique measures that may incur high information gathering costs (Bol 2011). However, there is an opposite prediction regarding the effect of unique measure. Because of high comparability of common measures among ratees, raters are more pressured to inflate those ratings (Fried et al. 1999). Hence, we predict no signal on this variable. We use the weight assigned to a measure ($\text{WEIGHT}$) to proxy for the relative importance among performance measures. Raters are more pressured to inflate ratings of highly weighted measures due to higher confrontation costs. We expect measurement weight to be positively associated with the tendency toward leniency. $\text{EXPERIENCE}$ measures the experience of raters with the average tenure of raters at a rating committee. As tenure of raters increases, favoritism may increase, and hence raters are more likely to inflate rating (Prendergast and Topel 1996). However, rotation rule and sabbatical year system in our sample prevent raters from exercising favoritism. Research in psychology also shows that greater degrees of expertise and experience raters make evaluation of ratees more reliable and less lenient (Brown 1968; Schneier 1977). Therefore, we expect a negative
relation between experience of raters and the lenient ratings. WOMAN measures the gender effect of raters as the number of female raters divided by total number of raters within a class. In a number of studies in performance evaluation, the gender effects of rater have been considered as possible sources of variance in ratings (Landy and Farr 1980). Although some laboratory studies find that female raters are more lenient than male raters (Bartol and Butterfield 1976), others show that gender does not typically affect evaluations (Peters et al 1984; Pulakos and Wexley 1983). Therefore, we do not have a signed prediction on the effect of woman raters on the leniency. SIZE is included because most previous studies support that large firms have better performance scores because they have sufficient amount of resources and capabilities, more knowledge or experiences, and well-established system (Lee 2009; Nooteboom 1993). The higher scores may partially come from leniency effect (Ahn et al. 2011). Large firms with abundant resources prepare refined reports and present them in a polished manner. Also, raters may face larger confrontation costs from larger firms. Thus, large firms are likely to get higher score ratings than small firms do. On the other hand, there is a great deal of pressure on raters to evaluate large firms accurately. SOEs with more employees receive more public attention and media coverage (Du et al. 2012), and hence, raters and the government are more cautious when they evaluate and inspect the performance of large firms. These conflicting findings of prior studies preclude us from making a signed prediction on the effect of firm size on leniency.

V. EMPIRICAL RESULTS

Descriptive Statistics

Figure 1, Panel A shows the histogram of subjective performance score rate (SUB_Attain). The mean (median) of subjective performance score rate is 0.664 (0.650) and the standard deviation is 0.131. Figure 1, Panel B presents that the mean (median) of objective
performance score rate is 0.880 (1.000) and the standard deviation is 0.205. In our sample data, the mean and median rating achievement ratio of the subjective measures are lower than those of objective measures. This is inconsistent with the conventional leniency bias that has been documented in the prior literature that measures leniency by comparing subjective performance score to objective performance score (Bol 2011; Moers 2005). Participative or formula-based target setting process for objective performance measure may allow the ratees to set easier targets and to reduce efforts after the targets are met, especially when raters are not residual claimants as in the setting of this paper (e.g. Ahn and Choi 2010; Merchant and Manzoni 1989).

[INSERT FIGURE 1 ABOUT HERE]

We present summary statistics for the leniency, subjective performance score rate, and characteristics of measurement, rater, and firms in Table 2. The average (median) leniency (LEN) is 0.025 (0.024) and its standard deviation is 0.143. More importantly, the positive value of LEN persists over the sample period, as shown in Figure 2, Panel A and all the values of LEN are statistically different from zero and positive except for year 2005. These descriptive statistics and t-test results present that leniency persists over time in our sample period, which supports H1.  

The asymmetric leniency may also induce compression effects, the tendency of raters to provide performance ratings that fail to distinguish between ratees (Ahn et al. 2010; Bol 2011). If high (low) past performers are continuously rated lower (higher) in subsequent periods, then the distribution of rating will be concentrated around the median, and compression effects increase over time. As shown in Figure 2, Panel B, the standard deviations of both LEN and SUB_Attain monotonically decrease over time, representing the  

15 The leniency cannot infinitely persist over time because of the score constraint. Therefore, the continuous leniency might be controlled in several ways by prohibiting leniency (mandatory forced distribution) and changing performance evaluation system and measures. Ahn and Kim (2013) find one of the possibilities that leniently rated subjective performance measures are likely to be dropped. Other possibilities are beyond scope of this paper, and remain as a subject of future study.
time-series compression effects.\textsuperscript{16}

Subjective performance score rate ($SUB\_Attain$) monotonically increases over time, with the mean of $\Delta SUB\_Attain$ at 2.1 percent. We conjecture that the positive $LEN$ represents the time trend of subjective performance score rate ($SUB\_Attain$). The objective performance score rate ($OBJ\_Attain$) shows relatively stable pattern over time. The tenure of raters is, on average, about two years. The average fraction of woman raters is 7.5 percent, indicating that most raters consist of male raters. Table 3 presents the Pearson correlation matrix. In general, the control variables are not highly correlated with each other.

\[\text{[INSERT FIGURE 2 ABOUT HERE]}\]

\[\text{[INSERT TABLE 2 ABOUT HERE]}\]

\[\text{[INSERT TABLE 3 ABOUT HERE]}\]

**Impact of Prior Performance Information on Leniency**

Columns (1) to (6) of Table 4 present the estimation results of Equations (1) and (2), where the leniency is the dependent variable. We replace prior subjective performance score rate ($SUB\_Attain_{t-1}$) with indicator variables of the highest and the lowest quartile of prior subjective performance score rate ($High\_SUB_{t-1}$ and $Low\_SUB_{t-1}$) in column (2). We find that measures with higher (lower) past performance score rate tend to exhibit a lower (higher) leniency, consistent with H2a. The negative relation between the previous performance level and leniency supports the notion that raters will positively inflate ratings of low-performing ratees. Also, in column (4) and (5) of Table 4, we find that measures with leniently (harshly) rated performance score rate tend to show a lower (higher) leniency, consistent with H2b.

\textsuperscript{16} We find the significant decreasing time trend by estimating the regression. When we regress annual and industrial average standard deviations of $LEN$ ($SUB\_Attain$) on time trend, the coefficient is -0.005 (-0.007) at p-value < 1%.
These results show the reverting tendency of leniency because raters have incentives to restore perceptions of equity.\(^{17}\)

With respect to control variables, we find that leniency is significantly higher in unique measures than in common measures. The superior requires specific knowledge about an individual subordinate’s characteristics to correctly evaluate the performance of the unique measure. The high information gathering costs let raters promote leniency (Bol 2011). We also find that measures of high \textit{WEIGHT} are more likely to be rated leniently. Moreover, firms with a more experienced raters receive a lower leniency in column (1) and (2), stressing the importance of raters’ characteristics in interpreting the leniency effects. Firm size is negatively related to leniency in column (4) and (5). These results indicate that raters are misled by firm size or not sophisticated enough to evaluate the performance of large firms. However, the gender effect on leniency is statistically insignificant.

The results in Columns (1), (2), (4) and (5) in Table 4 do not address the endogenous nature of the leniency. If some of the \textit{unobserved} measurement characteristics of leniency are also determinants of the previous performance, then the previous performance score rate may spuriously affect leniency. To address the effect of endogenously determined leniency, we estimate Equation (2) using a performance measure fixed-effects model to control for time-invariant unobserved heterogeneity. As shown in column (3) and (6) of Table 4, the coefficients on \textit{High\_SUB}\(_{t-1}\), \textit{Low\_SUB}\(_{t-1}\), \textit{High\_LEN}\(_{t-1}\), and \textit{Low\_LEN}\(_{t-1}\) continue to be significant at the 1 percent level, suggesting that our results are largely robust to corrections for the endogenous nature of the leniency. In sum, we interpret these results as confirming our previous findings of a negative relation between the prior performance score rate, the prior leniency tendency and the current tendency toward leniency.

\[\text{[INSERT TABLE 4 ABOUT HERE]}\]

\(^{17}\) We can also use a median of subjective measure scale (performance score “C”) as a benchmark performance. We empirically test and find the similar results to the main findings of this paper.
Additional Analysis

**Alternative Proxies for Leniency**

We check the robustness of our analysis on the asymmetric leniency by using the alternative leniency proxies. The previous results of estimating the relation between leniency and prior performance are based on the two government guidelines for SOE evaluation—(1) to use a forced distribution, and (2) to incorporate 50 percent of performance level and 50 percent of year-on-year performance improvement. To test the robustness of asymmetric leniency effects, we relax the two assumptions one by one.

First, raters can choose not to follow the forced distribution guideline although it is strongly recommended.\(^{18}\) We employ “the distribution of new measure”, the proxy for the actual rating tendency of raters.\(^{19}\) We utilize the distribution of new measures on annual base and apply the distribution to calculate \(BENCHMARK2\). Then we calculate leniency \((LEN2)\) by using the same Equation (3). Second, we assume that change in objective performance score rate represents rater efforts. When we eliminate the change in objective performance score from Equation (3), we get the additional proxy for leniency, which is free from the effort adjustment. In other words, \(LEN3\) is calculated by extracting the change in objective performance score rate part \((\Delta OBJ\_Attain)\) from the original leniency \((LEN)\). Third, even though the “50 percent of performance level and 50 percent of performance improvement” rule is mandatory when we calculate the \(BENCHMARK\), we relax this rule for the purpose of the robustness check. Then the benchmark is 100 percent of prior performance score level. If change in objective performance score rate is also ignored, then change in subjective performance score rate itself \((\Delta SUB\_Attain)\) is the proxy for leniency. Finally, to check the possibility that raters think of not only one-year prior performance but also prior performance

---

18 Merchant et al. (2010) find no evidence of a leniency effect in the subjective performance ratings, when forced distribution is mandated.
19 Our database consists of existing performance measures and newly adopted performance measures. In the previous analysis, we only utilize existing performance measures because the consecutive two-year observations are necessary to calculate our proxy for leniency.
trend as the benchmark, we use the average performance score rate in the previous three years as the benchmark \((SUB\_Attain2)\). We incorporate these four alternative measures of leniency into the Equation (2) as shown in the Table 5. The results show the evidence consistent with our prediction that lower (higher) past performance is positively (negatively) associated with lenient ratings.\(^{20}\) In an untabulated analysis, we also find the reverting pattern of leniency, consistent with H2b.

**Alternative Proxies for Prior Performance**

We additionally check the possibility of another standard of judgment about prior performance. Raters normally have information of both past subjective performance scores and current objective performance scores when they conduct evaluation of subjective performance measures. In other words, the overall objective performance rates of current year are likely to be known by raters before subjective performance evaluation. Therefore, there is a possibility that raters would judge ratees’ performance level by their current objective performance rate and evaluate ratees leniently or harshly according to their incentives. In an untabulated analysis, we find the same asymmetric leniency pattern – lenient (harsh) for low (high) performers – when we classify superior and poor performer by using current objective performance score.\(^{21}\)

**Same Rater Bias**

Typical argument is that if a rater consecutively evaluates the same ratee for two periods, familiarity and favoritism will increase, resulting in lenient rating (Prendergast and Topel 1996). It is impossible, however, to test this possibility in our setting. The rater rotation rule

---

\(^{20}\) The previous results with our original leniency proxy are more conservative than those with alternative proxies in terms of statistical significance. Benchmark with prior score only can induce much higher effect of past performance level on leniency.

\(^{21}\) When we replace the past subjective performance \((SUB\_Attain_{-1})\) with the current objective performance score rate \((OBJ\_Attain)\) as a proxy for the superiority of ratee’s performance, the result remains the same. The coefficient of the \(OBJ\_Attain\) is -1.040 at p-value < 1%. 

---

[INSERT TABLE 5 ABOUT HERE]
is mandated in Korean SOE setting, and thus, the same rater cannot evaluate the same SOE for consecutive two years. Alternatively, we manually trace all the raters’ evaluation committee, not individual level, and make a proxy for the fraction of raters who evaluate the same industry for consecutive two years. Untabulated results suggest that there is statistically significant and negative relationship between the portion of raters who rate the same industry for consecutive two years and lenient rating tendency when we add the proxy in Equation (2). In other words, similar to EXPERIENCE measure, raters who have experience with evaluating a SOE in the same industry are more likely to evaluate ratees more reliably and less leniently because of reduced information gathering costs (Brown 1968; Schneier 1977).

VI. CONCLUSION

The purpose of this paper is to examine how leniency behaves over time in terms of persistence, whether lenient ratings are affected by the past performance information. Using the dataset of multiple SOEs for multiple years, we construct the proxy for leniency at the performance measure level, based on the prior performance information. We find that (1) leniency effect persists over time, (2) leniency is reverted, and (3) high (low) past performers tend to be rated less (more) leniently. Also, these findings are robust to four alternative proxies for leniency. Interestingly the persistence of leniency over time and asymmetric leniency contributes to compression bias. Leniency and compression effects do not stem from different motivations, but from the same incentive to evaluate the low performers more leniently.

The results of this study have several practical implications. We examine the incentive and reaction of raters with regard to lenient rating. Our study shows that the consideration of the prior performance information is very important in subjective evaluation, and suggests that the incentives of raters and ratees differ with respect to their previous performance level.
and their prior leniency. Also, the intrinsic motivation of public sector employees is a major premise condition of effort and performance (e.g., Prendergast 2007; Wright 2001), and is higher than that of private sector employees (Lyons et al. 2006). Given its importance in this sector, our results suggest that raters should comprehensively consider the motivational effects of leniency in subjective evaluation focusing on the prior performance information.

Our findings are subject to important caveats. Most importantly, generalizability of this study is limited because this study is based only on public entities in Korea and the quality of the variables employed in this study may not be optimal. Societal norms and values in Korea may also have some influence on our results. However, our sample firms employ a wide range of performance measures (e.g., subjective versus objective, common versus unique) and adopt a performance-based bonus system that is parallel to performance evaluation system in most successful enterprises. Second, our proxy for leniency is constructed based upon the field-specific assumptions in Korean SOEs. Therefore, we should exercise caution when generalizing the results based on our leniency proxy. Lastly, even though we suggest that leniency persists over time, the leniency tendency cannot infinitely persist because of the score constraint. Future research can address these limitations by incorporating the age of performance measure and the timing of performance measure drop.
### APPENDIX A

Test for Relevance of Benchmarks to Current Subjective Score

<table>
<thead>
<tr>
<th>Independent Variables (b)</th>
<th>Predicted Sign</th>
<th>(1) PM level Coefficient (t-value)</th>
<th>(2) PM level Coefficient (t-value)</th>
<th>(3) SOE level Coefficient (t-value)</th>
<th>(4) SOE level Coefficient (t-value)</th>
<th>(5) SOE level Coefficient (t-value)</th>
<th>(6) SOE level Coefficient (t-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(SUB_{\text{Attain}}_{t-1})</td>
<td>+</td>
<td>0.337*** (14.90)</td>
<td>0.311*** (8.81)</td>
<td>0.634*** (22.23)</td>
<td>0.465*** (8.75)</td>
<td>0.614*** (10.78)</td>
<td>0.558*** (2.83)</td>
</tr>
<tr>
<td>(OBJ_{\text{Attain}}_{t-1})</td>
<td>+</td>
<td>0.052*** (3.61)</td>
<td>-0.010 (-0.49)</td>
<td>0.026 (0.89)</td>
<td>0.016 (0.45)</td>
<td>0.024 (0.80)</td>
<td>0.013 (0.38)</td>
</tr>
<tr>
<td>(TOT_{\text{Attain}}_{t-1})</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td>UNIQUE(_i), WEIGHT(_i), EXPERIENCE(_i), WOMAN(_i), SIZE(_i)</td>
<td>UNIQUE(_i), WEIGHT(_i), EXPERIENCE(_i), WOMAN(_i), SIZE(_i)</td>
<td>EXPERIENCE(_i), WOMAN(_i), SIZE(_i)</td>
<td>EXPERIENCE(_i), WOMAN(_i), SIZE(_i)</td>
<td>EXPERIENCE(_i), WOMAN(_i), SIZE(_i)</td>
<td>EXPERIENCE(_i), WOMAN(_i), SIZE(_i)</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>0.321*** (2.81)</td>
<td>1.140*** (4.45)</td>
<td>0.207*** (6.69)</td>
<td>0.242*** (4.70)</td>
<td>0.200*** (5.93)</td>
<td>0.228*** (4.62)</td>
</tr>
<tr>
<td>Year Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SOE Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SOE Clustering</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure Clustering</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>7,470</td>
<td>3,310</td>
<td>500</td>
<td>295</td>
<td>500</td>
<td>295</td>
<td>295</td>
</tr>
<tr>
<td>Adjusted R(^2)</td>
<td>0.443</td>
<td>0.378</td>
<td>0.689</td>
<td>0.596</td>
<td>0.688</td>
<td>0.593</td>
<td></td>
</tr>
</tbody>
</table>

T-statistics are reported in parentheses under each estimated coefficient. The symbols *, **, and *** correspond to 10 percent, 5 percent, and 1 percent significance levels, respectively, for two-tailed t-tests. \(TOT_{\text{Attain}}_{t-1}\) is defined as the prior total performance score rate of a SOE. See Table 2 for the definitions of other variables.
Panel A: Numerical Example of Estimation Process of LEN

Panel B: Example of BENCHMARK in 2005

<table>
<thead>
<tr>
<th>Prior Grade</th>
<th>50% of Level</th>
<th>50% of Improvement</th>
<th>BENCHMARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+ (1.000)</td>
<td>40% A+ = 0.400</td>
<td>0.569</td>
<td>50% * 0.975</td>
</tr>
<tr>
<td></td>
<td>40% A+ = 0.400</td>
<td></td>
<td>50% * 0.569</td>
</tr>
<tr>
<td></td>
<td>20% A0 = 0.175</td>
<td></td>
<td>0.772</td>
</tr>
<tr>
<td>A0 (0.875)</td>
<td>40% A+ = 0.400</td>
<td>0.569</td>
<td>50% * 0.900</td>
</tr>
<tr>
<td></td>
<td>40% A0 = 0.350</td>
<td></td>
<td>50% * 0.569</td>
</tr>
<tr>
<td></td>
<td>20% B+ = 0.150</td>
<td></td>
<td>0.735</td>
</tr>
<tr>
<td></td>
<td>0.900</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E+ (0.125)</td>
<td>40% D0 = 0.100</td>
<td>0.569</td>
<td>50% * 0.150</td>
</tr>
<tr>
<td></td>
<td>40% E+ = 0.050</td>
<td></td>
<td>50% * 0.569</td>
</tr>
<tr>
<td></td>
<td>20% E0 = 0.000</td>
<td></td>
<td>0.360</td>
</tr>
<tr>
<td></td>
<td>0.150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E0 (0.000)</td>
<td>40% E+ = 0.050</td>
<td>0.569</td>
<td>50% * 0.050</td>
</tr>
<tr>
<td></td>
<td>40% E0 = 0.000</td>
<td></td>
<td>50% * 0.569</td>
</tr>
<tr>
<td></td>
<td>20% E0 = 0.000</td>
<td></td>
<td>0.310</td>
</tr>
<tr>
<td></td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Raters are supposed to follow these guidelines while keeping the prior performance score in mind as the reference point. First, ‘50 percent of improvement’ part would be applied by the forced distribution rule (A+/A0: 10%, B+: 10%, B0: 20%, C:40%, D+/D0/E+/E0: 20%). Because this part should be independently determined and not be affected by the previous performance score in accordance with the rule, “50 percent of improvement” portion for
every measure has the same distribution and weighted average of each rating of about 0.569.
The subjective performance rating has nine grades, from A+ to E0. This rating determines score rates: a grade of A+ translates to the score rate of 100 percent, diminishing by 12.5 percent for each lower grade, and ending at 0 percent for E0. Thus, multiplying the score level by its weight which is based on the forced distribution rule yields 0.569 \[= 1.000 \times (A+) \times 5\% + 0.875 \times (A0) \times 5\% + 0.750 \times (B+) \times 10\% + 0.625 \times (B0) \times 20\% + 0.500 \times (C) \times 40\% + 0.375 \times (D+) \times 20\%\].

Second, the calculation of “50 percent of performance level” part starts from the prior score level and slightly adjusts the score according to “50 percent of improvement” part. In other words, the current performance level can be upgraded or degraded or unchanged from the prior score rating. As a numerical example in Panel A, let’s assume that the previous subjective performance score (\(SUB\_Attain_{t-1}\)) of “Organization management” measure in a SOE is D+ (0.375). If raters evaluate that performance improves with a probability of 40 percent (i.e. the sum of A+/A0: 10%, B+: 10%, B0: 20%), then performance level will be upgraded by one level, (i.e. C: 0.50). No change in performance with probability of 40 percent (C: 40%) earns grade D+ (0.375), and deteriorated performance earns grades below D+ in 50 percent of improvement, and performance level will be degraded by one level D0 (0.25) with probability of 20 percent. Hence, the “50 percent of level” part is the sum of the three cases, 0.40 \[= 0.5 \times (C) \times 40\% + 0.375 \times (D+) \times 40\% + 0.25 \times (D-) \times 20\%\] which is mainly subject to the prior performance level. So, the benchmark level (\(BENCHMARK_i\)) is 0.485, equals to the sum of 1) the weighted average 0.285 from “50 percent of improvement” part, and 2) 0.200 from “50 percent of performance level” part. Panel B presents the example of benchmark level (\(BENCHMARK_i\)) in 2005. Annual adjustment to the benchmark is necessary because of intermittent rule change in scaling. This produces a benchmark table at ‘grade-year’ level that has, on average, nine benchmarks per year.

Also, if we assume that current performance achievement of “Organization management"
measure is B0 (0.625) and changes in objective performance achievement is 0.060, then the value of leniency ($LEN_t$) of the performance measure is 0.080 [ = $SUB_{\text{Attain}}$ (0.625) – $BENCHMARK_t$ (0.485) – $\Delta OBJ_{\text{Attain}}$ (0.060) ].
REFERENCES


Panel A: Distribution of $SUB\_Attain$

Panel B: Distribution of $OBJ\_Attain$

Panel A shows the distribution of the score rate of subjective performance measure ($SUB\_Attain$). The mean (median) value of subjective measure is 0.664 (0.650) and the standard deviation is 0.131. Panel B shows the distribution of score rate of objective performance measure ($OBJ\_Attain$). The mean (median) value of objective measure is 0.880 (1.000) and the standard deviation is 0.205. Both panel have eight bins which have the interval of 0.125. The sample includes 7,470 subjective performance measures and 6,617 objective performance measures from 2005 to 2011.
FIGURE 2
Time Trends of Performance Score Rate and Leniency

Panel A: Time Trends of Means of Variables

Panel B: Time Trends of Standard Deviations of Variables

The sample includes 7,470 subjective performance measures (SUB_Attain), changes in subjective performance measures (ΔSUB_Attain), leniency measures (LEN) and 6,617 objective performance measures (OBJ_Attain) from 2005 to 2011 in both Panel A and Panel B.
### Table 1: Examples of Performance Rating and Composition of Raters

#### Panel A: Example of Performance Ratings for an SOE in 2005\(^a\)

<table>
<thead>
<tr>
<th>Performance Category</th>
<th>Individual Measure</th>
<th>Measure-ment (^b)</th>
<th>Commonality</th>
<th>Weight</th>
<th>Rating</th>
<th>Score</th>
<th>Attain (=Score rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Overall management</td>
<td>Capital productivity</td>
<td>OBJ</td>
<td>Common</td>
<td>5</td>
<td>-</td>
<td>3.998</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>Customer satisfaction</td>
<td>OBJ</td>
<td>Common</td>
<td>2</td>
<td>-</td>
<td>1.512</td>
<td>0.756</td>
</tr>
<tr>
<td></td>
<td>Restructuring or management innovation</td>
<td>SUB</td>
<td>Common</td>
<td>4</td>
<td>C</td>
<td>2.000</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>Managing board of directors</td>
<td>SUB</td>
<td>Common</td>
<td>6.25</td>
<td>C</td>
<td>3.125</td>
<td>0.500</td>
</tr>
<tr>
<td>2. Main business</td>
<td>Maintaining high load factor</td>
<td>OBJ</td>
<td>Unique</td>
<td>3</td>
<td>-</td>
<td>3.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Effort for demand management</td>
<td>SUB</td>
<td>Unique</td>
<td>4</td>
<td>A0</td>
<td>3.500</td>
<td>0.750</td>
</tr>
<tr>
<td></td>
<td>Efficiency of overseas business</td>
<td>SUB</td>
<td>Unique</td>
<td>2</td>
<td>B+</td>
<td>1.500</td>
<td>0.750</td>
</tr>
<tr>
<td>3. Business management</td>
<td>Financial structure</td>
<td>OBJ</td>
<td>Common</td>
<td>5</td>
<td>B+</td>
<td>2.165</td>
<td>0.433</td>
</tr>
<tr>
<td></td>
<td>Labor union management</td>
<td>SUB</td>
<td>Common</td>
<td>2</td>
<td>A0</td>
<td>1.500</td>
<td>0.750</td>
</tr>
<tr>
<td></td>
<td>Budget management</td>
<td>SUB</td>
<td>Common</td>
<td>5</td>
<td>-</td>
<td>4.375</td>
<td>0.875</td>
</tr>
</tbody>
</table>

### Panel B: Example of Composition of Raters in 2008\(^c\)

<table>
<thead>
<tr>
<th>SOE Class</th>
<th>Overall management</th>
<th>Main business</th>
<th>Business management</th>
<th>Objective measure</th>
<th>No. of raters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>3 professors</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1. SOC (14 SOEs)</td>
<td>4 professors</td>
<td>5 professors</td>
<td>6 professors</td>
<td>2 professors 2 CPAs</td>
<td>23</td>
</tr>
<tr>
<td>2. Service (10 SOEs)</td>
<td>4 professors 1 industrial expert</td>
<td>4 professors 1 industrial expert</td>
<td>5 professors</td>
<td>1 professor 3 CPAs</td>
<td>19</td>
</tr>
<tr>
<td>3. Inspection management (6 SOEs)</td>
<td>3 professors</td>
<td>4 professors 1 industrial expert</td>
<td>3 professors</td>
<td>2 CPAs</td>
<td>13</td>
</tr>
<tr>
<td>4. Culture (9 SOEs)</td>
<td>3 professors 1 industrial expert</td>
<td>4 professors</td>
<td>3 professors 1 industrial expert</td>
<td>1 professor 2 CPAs</td>
<td>15</td>
</tr>
<tr>
<td>5. Industry promotion I (10 SOEs)</td>
<td>6 professors</td>
<td>4 professors</td>
<td>3 professors 1 CPA</td>
<td>3 CPAs</td>
<td>17</td>
</tr>
<tr>
<td>6. Industry promotion II (6 SOEs)</td>
<td>3 professors</td>
<td>3 professors</td>
<td>2 professors 1 industrial expert</td>
<td>1 professor 1 CPA</td>
<td>11</td>
</tr>
<tr>
<td>7. Education (6 SOEs)</td>
<td>2 professors 1 industrial expert</td>
<td>3 professors</td>
<td>2 professors 1 industrial expert</td>
<td>1 professor 2 CPAs</td>
<td>12</td>
</tr>
<tr>
<td>8. Pension (14 SOEs)</td>
<td>5 professors 1 industrial expert</td>
<td>5 professors 1 industrial expert</td>
<td>7 professors</td>
<td>4 CPAs</td>
<td>23</td>
</tr>
<tr>
<td>Total (75 SOEs)</td>
<td>33 professors 6 industrial experts</td>
<td>32 professors 5 industrial experts</td>
<td>31 professors 3 industrial experts 1 CPA</td>
<td>6 professors 19 CPAs</td>
<td>136</td>
</tr>
</tbody>
</table>

Panel A shows the example of performance ratings for KEPCO (Korea Electronic Power Corporation) in 2005. OBJ and SUB in the measurement column in Panel A are objective and subjective performance measures, respectively. Panel B presents the example of raters composition in 2008.
TABLE 2
Descriptive Statistics of the Sample

<table>
<thead>
<tr>
<th>Variables a</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEN</td>
<td>7,470</td>
<td>0.025</td>
<td>0.144</td>
<td>-0.608</td>
<td>-0.067</td>
<td>0.024</td>
<td>0.117</td>
<td>0.644</td>
</tr>
<tr>
<td>SUB_Attain</td>
<td>7,470</td>
<td>0.664</td>
<td>0.131</td>
<td>0.125</td>
<td>0.600</td>
<td>0.650</td>
<td>0.750</td>
<td>1.000</td>
</tr>
<tr>
<td>∆SUB_Attain</td>
<td>7,470</td>
<td>0.021</td>
<td>0.124</td>
<td>-0.550</td>
<td>-0.050</td>
<td>0.000</td>
<td>0.125</td>
<td>0.625</td>
</tr>
<tr>
<td>UNIQUE</td>
<td>7,470</td>
<td>0.296</td>
<td>0.456</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>7,470</td>
<td>3.706</td>
<td>1.830</td>
<td>0.300</td>
<td>3.000</td>
<td>3.000</td>
<td>4.100</td>
<td>12.000</td>
</tr>
<tr>
<td>EXPERIENCE</td>
<td>7,470</td>
<td>2.107</td>
<td>0.376</td>
<td>1.364</td>
<td>1.829</td>
<td>2.125</td>
<td>2.250</td>
<td>3.222</td>
</tr>
<tr>
<td>WOMAN</td>
<td>7,470</td>
<td>0.075</td>
<td>0.046</td>
<td>0.000</td>
<td>0.045</td>
<td>0.073</td>
<td>0.104</td>
<td>0.167</td>
</tr>
<tr>
<td>∆OBJ_Attain</td>
<td>7,470</td>
<td>0.003</td>
<td>0.100</td>
<td>-0.441</td>
<td>-0.051</td>
<td>0.004</td>
<td>0.051</td>
<td>0.388</td>
</tr>
<tr>
<td>SIZE</td>
<td>7,470</td>
<td>6.391</td>
<td>1.430</td>
<td>2.303</td>
<td>5.176</td>
<td>6.430</td>
<td>7.244</td>
<td>10.363</td>
</tr>
<tr>
<td>ROOM</td>
<td>7,470</td>
<td>2.350</td>
<td>0.662</td>
<td>0.000</td>
<td>1.951</td>
<td>2.402</td>
<td>2.808</td>
<td>3.920</td>
</tr>
</tbody>
</table>

The sample includes 7,470 unique measure-firm-years and 109 unique SOEs from 2005 to 2011. Data for performance rating and rater characteristics are manually collected from the performance evaluation report of SOEs (www.alio.go.kr).

a Variable Definitions:
- LEN = defined in the “RESEARCH DESIGN” section;
- SUB_Attain = performance score rate of a subjective measure;
- ∆SUB_Attain = change in subjective performance score rate (=SUB_Attain_{ijt} – SUB_Attain_{ij,t-1});
- UNIQUE = 1 if the measure is a unique one, 0 if it is a common one;
- WEIGHT = weight assigned to the measure;
- EXPERIENCE = average tenure of raters within rating committee;
- WOMAN = fraction of female raters, which is the number of female raters within an SOE class divided by the number of total raters within an SOE class;
- ∆OBJ_Attain = change in sum of objective performance score rate (=OBJ_Attain_{ijt} – OBJ_Attain_{ij,t-1});
- SIZE = natural logarithm of the number of employees; and
- ROOM = room for improvement (= log_e (101 – OBJ_Attain_{ij,t-1}*100))

where i, j, and t indicate performance measure i, firm j, and year t, respectively.
### TABLE 3
Pearson Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>LEN</th>
<th>SUB_Attain</th>
<th>UNIQUE</th>
<th>WEIGHT</th>
<th>EXPERIENCE</th>
<th>WOMAN</th>
<th>SIZE</th>
<th>ROOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEN</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUB_Attain</td>
<td>0.592***</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNIQUE</td>
<td>0.030**</td>
<td>0.060***</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEIGHT</td>
<td>0.063***</td>
<td>0.081***</td>
<td>-0.185***</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPERIENCE</td>
<td>0.030**</td>
<td>0.138***</td>
<td>0.002</td>
<td>0.086***</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAN</td>
<td>-0.026**</td>
<td>-0.015</td>
<td>-0.005</td>
<td>0.026**</td>
<td>-0.059***</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIZE</td>
<td>0.068***</td>
<td>0.290***</td>
<td>0.005</td>
<td>-0.025**</td>
<td>0.133***</td>
<td>0.081***</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>ROOM</td>
<td>0.346***</td>
<td>-0.087***</td>
<td>0.009</td>
<td>-0.021*</td>
<td>-0.037***</td>
<td>-0.139***</td>
<td>-0.019</td>
<td>1.000</td>
</tr>
</tbody>
</table>

P-values are reported in parentheses under each estimated correlation value. The symbols *, **, and *** correspond to 10 percent, 5 percent, and 1 percent significance levels, respectively, for two-tailed t-tests.
<table>
<thead>
<tr>
<th>Independent Variables ( b )</th>
<th>Predicted Sign</th>
<th>(1) Coefficient (t-value)</th>
<th>(2) Coefficient (t-value)</th>
<th>(3) Coefficient (t-value)</th>
<th>(4) Coefficient (t-value)</th>
<th>(5) Coefficient (t-value)</th>
<th>(6) Coefficient (t-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( SUB_{Attain, t-1} )</td>
<td>–</td>
<td>-0.147*** (-6.79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( High_{SUB, t-1} )</td>
<td>–</td>
<td>-0.035*** (-7.32)</td>
<td>-0.050*** (-10.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Low_{SUB, t-1} )</td>
<td>+</td>
<td>0.012** (2.48)</td>
<td>0.023*** (4.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( LEN_{t-1} )</td>
<td>–</td>
<td>-0.337*** (-24.55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( High_{LEN, t-1} )</td>
<td>–</td>
<td>-0.070*** (-14.49)</td>
<td>-0.078*** (-14.68)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Low_{LEN, t-1} )</td>
<td>+</td>
<td>0.046*** (12.70)</td>
<td>0.052*** (10.57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNIQUE(_{i,t} )</td>
<td>+/-</td>
<td>0.012*** (3.10)</td>
<td>0.011*** (2.92)</td>
<td>-0.120</td>
<td>0.016*** (3.19)</td>
<td>0.014*** (2.92)</td>
<td>0.075** (2.35)</td>
</tr>
<tr>
<td>WEIGHT(_{i,t} )</td>
<td>+</td>
<td>0.007*** (3.07)</td>
<td>0.006*** (3.11)</td>
<td>0.007*** (5.40)</td>
<td>0.010*** (4.21)</td>
<td>0.009*** (4.52)</td>
<td>0.009*** (5.58)</td>
</tr>
<tr>
<td>EXPERIENCE(_{i,t} )</td>
<td>–</td>
<td>-0.019*** (-2.61)</td>
<td>-0.020*** (-2.76)</td>
<td>-0.020*** (-3.18)</td>
<td>-0.006 (-5.56)</td>
<td>-0.006 (-5.90)</td>
<td>-0.010 (-1.35)</td>
</tr>
<tr>
<td>WOMAN(_{i,t} )</td>
<td>+/-</td>
<td>-0.027 (-2.65)</td>
<td>-0.030 (-2.76)</td>
<td>-0.033 (-3.18)</td>
<td>0.036 (0.84)</td>
<td>0.017 (0.36)</td>
<td>0.023 (0.38)</td>
</tr>
<tr>
<td>SIZE(_{i,t} )</td>
<td>+/-</td>
<td>0.003 (0.26)</td>
<td>0.004 (0.32)</td>
<td>0.001 (0.04)</td>
<td>-0.033*** (-2.87)</td>
<td>-0.035*** (-2.89)</td>
<td>-0.035** (-1.83)</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>0.136 (1.09)</td>
<td>0.031 (0.23)</td>
<td>0.139 (0.86)</td>
<td>0.355*** (2.81)</td>
<td>0.385*** (2.92)</td>
<td>0.253 (1.26)</td>
</tr>
<tr>
<td>Year Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SOE Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Measure Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of</td>
<td>7,470</td>
<td>7,470</td>
<td>7,470</td>
<td>4,936</td>
<td>4,936</td>
<td>4,936</td>
<td>4,936</td>
</tr>
<tr>
<td>Adjusted R(^2)</td>
<td>0.124</td>
<td>0.121</td>
<td>0.098</td>
<td>0.197</td>
<td>0.170</td>
<td>0.145</td>
<td></td>
</tr>
</tbody>
</table>

T-statistics are reported in parentheses under each estimated coefficient. Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by performance measure. The symbols *, **, and *** correspond to 10 percent, 5 percent, and 1 percent significance levels, respectively, for two-tailed t-tests. Please refer to the paper for a detailed explanation of these tests.

a) This table shows the coefficient estimates of the leniency determinants by using the following equations:

\[
LEN_{jt} = \alpha_0 + \alpha_1 SUB_{Attain, j,t-1} + \alpha_2 UNIQUE_{jt} + \alpha_3 WEIGHT_{jt} + \alpha_4 EXPERIENCE_{jt} + \alpha_5 WOMAN_{jt} + \alpha_6 SIZE_{jt} + \text{Fixed Effects} + \varepsilon
\]  

(1)

\[
LEN_{jt} = \alpha_0 + \alpha_1 High_{SUB, j,t-1} + \alpha_2 Low_{SUB, j,t-1} + \alpha_3 UNIQUE_{jt} + \alpha_4 WEIGHT_{jt} + \alpha_5 EXPERIENCE_{jt} + \alpha_6 WOMAN_{jt} + \alpha_7 SIZE_{jt} + \text{Fixed Effects} + \varepsilon
\]

(2)

where \( i, j, t \) indicate performance measure \( i \), firm \( j \), and year \( t \), respectively.

b) The independent variable, \( Low_{SUB, t-1} (High_{SUB, t-1}) \) is an indicator variable which refers to 1 if the score rate belongs to the lowest (highest) quartile of \( SUB_{Attain} \) in the previous year, 0 otherwise. \( Low_{LEN, t-1} (High_{LEN, t-1}) \) is an indicator variable which refers to 1 if the leniency tendency belongs to the lowest (highest) quartile of \( LEN \) in the previous year, 0 otherwise. See Table 2 for the definitions of other variables.
**TABLE 5**
Regressions of Alternative Measures of Leniency on Prior Performance Score Rate

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Predicted Sign</th>
<th>Dependent Variable</th>
<th>(1) LEN2t</th>
<th>(2) LEN3t</th>
<th>(3) ∆SUB_Attain_t</th>
<th>(4) ∆SUB_Attain2_t</th>
</tr>
</thead>
<tbody>
<tr>
<td>High_SUBt, j</td>
<td>-</td>
<td></td>
<td>-0.035***</td>
<td>-0.030***</td>
<td>-0.114***</td>
<td>-0.038***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-7.34)</td>
<td>(-6.21)</td>
<td>(-18.36)</td>
<td>(-6.70)</td>
</tr>
<tr>
<td>Low_SUBt, j</td>
<td>+</td>
<td></td>
<td>0.012**</td>
<td>0.023***</td>
<td>0.100***</td>
<td>0.025***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.55)</td>
<td>(6.72)</td>
<td>(28.23)</td>
<td>(5.31)</td>
</tr>
<tr>
<td>UNIQUEt</td>
<td>+/-</td>
<td></td>
<td>0.011***</td>
<td>0.010***</td>
<td>0.010***</td>
<td>0.015***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.96)</td>
<td>(2.78)</td>
<td>(2.81)</td>
<td>(2.69)</td>
</tr>
<tr>
<td>WEIGHTt</td>
<td>+</td>
<td></td>
<td>0.006***</td>
<td>0.007***</td>
<td>0.006***</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.11)</td>
<td>(3.15)</td>
<td>(3.46)</td>
<td>(1.57)</td>
</tr>
<tr>
<td>EXPERIENCEt</td>
<td>-</td>
<td></td>
<td>-0.020***</td>
<td>0.009</td>
<td>0.006</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-2.78)</td>
<td>(1.57)</td>
<td>(1.13)</td>
<td>(1.32)</td>
</tr>
<tr>
<td>WOMANt</td>
<td>+/-</td>
<td></td>
<td>-0.029</td>
<td>-0.087***</td>
<td>-0.084***</td>
<td>-0.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-0.69)</td>
<td>(-2.67)</td>
<td>(-2.43)</td>
<td>(-1.45)</td>
</tr>
<tr>
<td>SIZEt</td>
<td>+/-</td>
<td></td>
<td>0.004</td>
<td>0.011</td>
<td>0.012</td>
<td>-0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
<td>(0.88)</td>
<td>(0.88)</td>
<td>(-1.38)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(-0.63)</td>
<td>(-0.64)</td>
<td>(1.50)</td>
</tr>
<tr>
<td>Intercepts</td>
<td></td>
<td></td>
<td>0.001</td>
<td>-0.080</td>
<td>-0.088</td>
<td>0.467</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(-0.03)</td>
<td>(-0.03)</td>
<td>(1.50)</td>
</tr>
<tr>
<td>Year Fixed Effects</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SOE Fixed Effects</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of observations</td>
<td></td>
<td></td>
<td>7,470</td>
<td>7,470</td>
<td>7,470</td>
<td>3,310</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td></td>
<td></td>
<td>0.130</td>
<td>0.118</td>
<td>0.299</td>
<td>0.129</td>
</tr>
</tbody>
</table>

T-statistics are reported in parentheses under each estimated coefficient. Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by performance measure. The symbols *, **, and *** correspond to 10 percent, 5 percent, and 1 percent significance levels, respectively, for two-tailed t-tests. Please refer to the paper for a detailed explanation of these tests.

a This table estimates the impact of prior performance score rate on leniency by using the following equation:

\[
LEN_{ijt} = \alpha_0 + \alpha_1 High_{SUB_{ij,t-1}} + \alpha_2 Low_{SUB_{ij,t-1}} + \alpha_3 UNIQUE_{ijt} + \alpha_4 WEIGHT_{ijt} + \alpha_5 EXPERIENCE_{jt} + \alpha_6 WOMAN_{jt} + \alpha_7 SIZE_{jt} + Fixed Effects + \varepsilon. \tag{2}
\]

where \( i, j, \) and \( t \) indicate performance measure \( i \), firm \( j \), and year \( t \), respectively.

b The dependent variable, LEN2t, is calculated as subjective performance score rate (SUB_Attain) minus modified benchmark (BENCHMARK2) minus change in objective performance score (△OBJ_Attain). The modified benchmark (BENCHMARK2) is driven by the distribution of new measures which is a substitute of forced distribution assumption. LEN3t is LENt plus change in objective performance score (△OBJ_Attain). Change in subjective performance score rate (△SUB_Attain) is the difference between SUB_Attainijt and SUB_Attainij,t-1. The dependent variable in column (4), ASUB_Attain2t, is the gap between SUB_Attainijt and average subjective performance score rate of prior three years. See Table 2 for the definitions of other variables.