Available online at www.sciencedirect.com
SCIENCE DIRECT
Telecommunications Policy 28 (2004) 801–820
TELECOMMUNICATIONS POLICY
www.elsevierbusinessandmanagement.com/locate/telpol

On the design of performance measurement plans in the telecommunications industry

Lisa V. Wood[a,*], David E.M. Sappington[b]

[a]The Brattle Group, Suite 800, 1133 20th Street, NW, Washington, DC 20036-3408, USA
[b]Department of Economics, University of Florida, Gainesville, Florida 32611, USA

Abstract

This paper analyzes the design of performance measurement and remedy plans that have been implemented in the telecommunications industry to ensure competitive local exchange carriers are afforded a meaningful opportunity to compete in the provision of local exchange services. It demonstrates that the plans can impose penalties on incumbent local exchange carriers even when the incumbents provide the same or higher level of wholesale service quality to their competitors than they provide to themselves. Simulations are employed to illustrate the magnitude of these penalties.

Keywords: Performance measurement; Service quality; Standards; Regulation

1. Introduction

The Telecommunications Act of 1996 [1] ("the Act") paved the way for competitive local exchange carriers ("CLECs") to develop a considerable presence in the telecommunications industry.[2] It did so, in part, by requiring incumbent local exchange carriers ("ILECs") to:

*Corresponding author. Fax: +1 202 955 5059.
E-mail addresses: lisa.wood@brattle.com (L. Wood), sapping@ufl.edu (D.E.M. Sappington).

[1]Pub. L. No. 104-104, 110 stat. 56 (codified at 47 U.S.C. Sections 151 *et seq.*).

[2]As of 2001, CLECs were estimated to serve as many as 20% of lines in some ILEC regions (BellSouth, SBC, Qwest, & Verizon, 2002).

0308-5961/$ - see front matter
doi:10.1016/j.telpol.2004.05.014

(i) ‘‘unbundle’’ certain elements of their networks[3] and allow CLECs to purchase these unbundled network elements (‘‘UNEs’’);[4] and (ii) sell their retail services to CLECs at a wholesale discount, thereby enabling CLECs to sell these services to customers at competitive prices (‘‘resale’’). These unbundling and resale requirements were intended to allow CLECs to compete for customers without having to build their own complete, ubiquitous networks.[5]

Because network unbundling and resale operations assist CLECs in their competition with ILECs, an ILEC could benefit financially if these operations did not proceed smoothly. For example, an ILEC could gain if the unbundling process were delayed, or if the level of service quality associated with the delivery of wholesale services to competitors were lower than the corresponding level of service quality enjoyed by the ILEC's own retail affiliate. To limit the likelihood of such outcomes, state regulators have implemented performance measurement and remedy plans (PMPs). As their name implies, PMPs attempt to measure the performance of ILECs in providing wholesale services to CLECs, and to penalize the ILECs for performance that does not meet the requisite standards. In particular, PMPs attempt to assess whether an ILEC is disadvantaging its rivals by delivering a higher level of service quality to ‘‘itself’’ (i.e., to its own retail affiliate) than to its rivals. If the evidence suggests the ILEC is providing such non-parity service to CLECs, the ILEC is penalized financially.

Some argue that the financial penalties imposed on ILECs when they are judged to be providing non-parity service to their competitors are insufficient to deter the ILECs from intentionally disadvantaging their rivals. Indeed, some suggest that ILECs view the penalties they pay for delivering below-parity service simply as a cost of doing business and a cost that is outweighed by the associated benefit. These critics cite the persistent stream of penalties that the ILECs pay under PMPs as evidence in support of their claims.[6]

One purpose of this article is to assess whether, in fact, an ILEC's persistent payment of penalties under a PMP is necessarily evidence that the ILEC is disadvantaging its rivals, intentionally or otherwise. The study concludes this is not the case. Indeed, an ILEC may be penalized persistently even when it consistently provides parity service to its rivals. This anomaly stems from two related factors. First, the provisioning of wholesale services, like most industrial supply relationships, entails some unavoidable randomness. Therefore, an ILEC's measured performance in providing wholesale services will not always reflect perfectly its actual performance. Second, many PMPs incorporate an important asymmetry. The plans typically

[3]The Act directed the Federal Communications Commission to develop an appropriate list of unbundled network elements (UNEs), taking into consideration ‘‘whether (A) access to such network elements as are proprietary in nature is necessary; and (B) the failure to provide access to such network elements would impair the ability of the telecommunications carrier to provide the services that it seeks to offer’’ (47 U.S.C. Section 251 (D) (2)).

[4]For example, ILECs are required to make the loops in their networks available for use by CLECs. A loop is a transmission line that connects a customer's premise with the central office of the telephone company.

[5]The Act anticipated three modes of CLEC competition: (1) CLECs might purchase ILEC retail services at a wholesale discount and resell these services (in a process known as resale); (2) CLECs might purchase unbundled network elements and combine them with their own facilities, or use unbundled network elements to provide an end-to-end service which is commonly referred to as the UNE-platform or UNE-P; and (3) CLECs might employ only their own facilities to serve their customers, and simply interconnect with the ILECs. We will use the term ‘‘wholesale services’’ to refer to all of the wholesale services or elements that CLECs might employ when pursuing either of the first two modes of operation.

[6]See, for example (AT&T Communications of Michigan, Inc. (2002), p. 6 and footnote

impose financial penalties on the ILEC when its measured performance suggests it is delivering a lower level of service quality to its competitors than to itself. In contrast, the plans typically do not provide any financial reward when measured performance suggests the ILEC is delivering a higher level of service quality to its competitors than to itself.

These two factors imply that even though incidents of higher realized service quality for competitors may offset incidents of lower realized service quality for competitors when an ILEC provides the same (stochastic) quality to its competitors and to itself, there is no corresponding balancing of financial rewards and penalties. For this reason, an ILEC may pay penalties persistently under a PMP even though it consistently delivers parity service to its competitors.

The study explains this conclusion more fully and attempts to provide additional insight regarding the design and operation of PMPs as follows. Section 2 describes the PMP that is employed in the state of Michigan ("the Michigan Plan")[7] in order to illustrate the basic structure of PMPs more generally. Section 3 describes the simulations we performed to assess how the financial penalties an ILEC incurs under the Michigan Plan vary according to the (stochastic) level of service quality it delivers to itself and to its competitors. Section 4 presents the results of the simulations. The simulations reveal that an ILEC may incur penalties under the Michigan Plan even when it delivers to competitors a service quality that is, on average, 30% higher than the level of service quality it delivers to itself. Of course, altering a PMP to limit such penalties also can reduce the penalties imposed on an ILEC that delivers a lower level of service quality to its competitors than to itself. Section 4 also analyzes these tradeoffs. Section 5 considers the effects of introducing symmetric rewards and penalties into the Michigan Plan. Section 6 provides concluding observations.

2. The Michigan performance measurement plan

The Michigan Plan was implemented in 2001 to ensure that CLECs receive wholesale service quality from SBC Michigan that provides them with a meaningful opportunity to compete with SBC in Michigan.[8] The Michigan Plan tracks SBC's wholesale service quality performance in the following areas: (1) pre-ordering/ordering; (2) billing; (3) miscellaneous administrative; (4) provisioning; (5) maintenance; (6) interconnection trunks; (7) directory assistance and operator services; (8) local number portability; (9) 911; (10) poles, conduits, and rights of way; (11) collocation; (12) directory assistance database; (13) coordinated conversions; (14) NXX; (15) bona fide request process; and (16) additional measures.[9] The plan employs 148 different measures to

[7]The remedy plan as ordered July 25, 2001 in Michigan Public Service Commission Case No. U-11830. See Ameritech (2001).

[8]Illinois and Ohio established PMPs as a prerequisite for the merger of SBC and Ameritech in 1999. Other states in Ameritech's operating territory (i.e., Indiana, Michigan, and Wisconsin) established PMPs subsequently. The Michigan Plan is modeled after the PMP contained in the Texas 271 Interconnection Agreement ("the Texas Plan"), which is part of a standard interconnection agreement in Kansas, Oklahoma, Arkansas, and Missouri. The other Midwest states served by SBC—Illinois, Indiana, Ohio, and Wisconsin—operate under a PMP that is similar to the Texas Plan.

[9]To provide examples of performance measures in some of these areas, note that when a CLEC places an order, the ILEC's operational support system (OSS) must be available to accept CLEC transactions or data files during scheduled availability. The percentage of the time that the OSS interface is available is an example of a pre-ordering/ordering performance measure. The number of days required to provision a loop is an example of a provisioning measure. The

track performance in these areas. These broad measures are further disaggregated into sub-measures,[10] and the sub-measures are reported in four distinct geographic market areas.[11] This disaggregation implies the Michigan Plan could measure SBC's performance on as many as 963 sub-measures each month for each CLEC in Michigan.[12]

The performance measures (and sub-measures) in the Michigan Plan are of two types: parity measures and benchmark measures. *Parity measures* compare the service that SBC provides to a CLEC with the service it provides to its own retail customers or to its retail affiliate.[13] Thus, with parity measures, the service that SBC provides to "itself" becomes the standard by which the performance it delivers to CLECs is measured. In contrast, *benchmark measures* compare the service that SBC provides to CLECs with a specified standard that is not explicitly linked to the service that SBC provides to itself (typically because SBC does not provide this service to itself).[14]

Almost three quarters (72%) of the *paying measures* tracked in the Michigan Plan are parity measures, while approximately one-quarter (28%) are benchmark measures.[15] Paying measures are measures on which SBC is required to pay penalties if its performance is judged to be inadequate.[16] The judgment process is relatively simple for benchmark measures in the Michigan

(*footnote continued*)
number of trouble reports per 100 access lines is an example of a maintenance measure. The percentage of time that the ILEC misses a due date for collocating CLEC equipment is an example of a collocation measure. The percentage of directory assistance records that are updated inaccurately by the ILEC is an example of a database accuracy measure. See SBC/Ameritech (2001).

[10]Sub-measures are employed to account for the fact that relevant conditions can vary across geographic areas and/or according to the wholesale service being provided, or the nature of the task in question. To illustrate the nature of these sub-measures, consider the percentage of missed due dates caused by SBC/Ameritech in provisioning plain old telephone service (POTS) or a UNE-Platform. This performance measure (#29) is divided into four sub-measures for POTS and four for a UNE-P: residential service requiring fieldwork, residential service requiring no fieldwork, business service requiring fieldwork, and business service requiring no fieldwork. These sub-measures are tracked for each market area. Hereinafter, we will use the terms "measures" and "sub-measures" interchangeably.

[11]The geographic areas are: (1) Detroit Metro; (2) Grand Rapids-Kalamazoo; (3) Saginaw-Lansing-Jackson; and (4) Traverse City-Upper Peninsula. Some measures are tracked, tested, and reported at the state level instead of, or in addition to, the market area level. The empirical analysis in this paper is based on data from the Detroit Metro area.

[12]Fewer measures typically will be tracked in practice because the typical CLEC does not operate in all market areas in Michigan and does not make service requests that relate to the entire set of performance measures. The maximum number of performance measures tested for a CLEC in Michigan in January 2003 was 351.

[13]The service that SBC provides to a CLEC is compared to the better of the service that SBC provides to its retail affiliate(s) and to its retail customers. In Michigan, SBC delivers some services (e.g., data services) to retail customers through affiliates. SBC does not employ affiliates to deliver retail services in Texas. Consequently, under the Texas Plan, the service that SBC provides to a CLEC is always compared to the service that SBC delivers to its retail customers.

[14]As an example, consider the no fieldwork sub-measure for provisioning POTS for performance measure #29 discussed earlier (i.e., the percentage of SBC/Ameritech caused missed due dates). The benchmark for the no fieldwork sub-measure is 97% (for both residential and business) which means that SBC/Ameritech must complete orders by the due date at least 97% of the time. In contrast, the fieldwork sub-measure uses a parity comparison, meaning that the performance provided to the CLEC is compared to SBC/Ameritech retail performance.

[15]These are the percentages for January 2003, which is not an atypical month. For the three-month period from November 2002 to January 2003, for example, 71% of the paying measures in Michigan were parity measures and 29% were benchmark measures.

[16]SBC tracked and tested 7381 paying sub-measures in January 2003: 5311 parity sub-measures and 2070 benchmark sub-measures. This number represents an average of 59 sub-measures for each of the 126 CLECs that operated in Michigan in January 2003. Again, January is not an atypical month. During the three-month period from November

Plan: SBC is judged to have provided adequate performance if and only if the observed performance meets or exceeds the established benchmark standard. Statistical testing is not used for benchmark measures under the Michigan Plan.[17] In contrast, the judgment for parity measures involves a statistical comparison of the mean performance that SBC provides to its competitors and the mean performance that SBC provides to itself on each measure. If the mean performance that SBC provides to a competitor is found to be below the mean performance that SBC delivers to itself by an amount that is statistically significant, SBC is liable for penalty payments on the measure in question. The observed performance difference is deemed to be statistically significant if the probability that the difference could have occurred by chance is 5% or less. Thus, the Michigan Plan is designed to admit a 5% chance of a Type I error on each performance measure that is tested. A Type I error occurs on a parity measure when SBC is judged to have provided lower mean performance to a competitor than to itself on this measure when, in fact, this is not the case. A Type I error in this instance can be viewed as a rejection of the maintained (or ''null'') hypothesis that SBC delivered the same level of service quality to its competitors and to itself when, in fact, this hypothesis is true.

The choice of a 5% probability of a Type I error also affects the probability of a Type II error. In general, a Type II error occurs when the null hypothesis is accepted, even though it is false. In the present instance, a Type II error occurs when SBC delivers lower service quality to its competitors than to itself, but is not judged to have done so. The probability of a Type II error will increase if a test is altered to reduce the probability of a Type I error (holding sample size and variance constant). Therefore, the design of PMPs entails a fundamental tradeoff between the likelihood of Type I and Type II errors.

The design of PMPs is complicated further by the fact that PMPs typically involve the testing of multiple performance measures. As noted in Section 2, the Michigan Plan encompasses 148 different measures and 963 different submeasures. A PMP that admits a 5% chance of a Type I error on each measure that is tested can introduce a large probability that some Type I error will occur when many measures are tested. For example, if there is a 5% chance of a Type I error on each of ten measures, the likelihood that at least one Type I error will occur when ten measures are tested may be as high as 50%.[18]

When multiple tests are conducted, the family wise error (FWE) rate is commonly employed to assess the performance of the tests. The FWE rate is the natural counterpart to the probability of

(*footnote continued*)
2002 to January 2003, SBC tracked and tested an average of 56 sub-measures per CLEC each month. The Michigan Plan also includes some non-paying measures. Such measures often are new measures for which the plan typically requires a diagnostic period, or measures that the ILEC or the CLECs choose to monitor.

[17]The use of statistical testing for benchmark measures varies by state. For example, the Texas Plan employs statistical tests to determine whether SBC's performance meets established benchmark standards.

[18]This upper bound is achieved if the performance measures are mutually exclusive, reflecting the special type of dependency in which a Type I error on one measure implies that a Type I error cannot occur on another measure. To illustrate the likelihood of one or more Type I errors when the performance measures are independent, consider the simple case where there are only two performance measures. The probability of at least one Type I error in this case equals 0.0975 (compared to the corresponding probability of 0.10 when the two measures are mutually exclusive). This 0.0975 probability is the sum of: (1) the probability that a Type I error occurs on the first measure only ($=0.05\times0.95=0.0475$); (2) the probability that a Type I error occurs on the second measure only ($=0.95\times0.05=0.0475$); and (3) the probability that a Type I error occurs on both measures ($=0.05\times0.05=0.0025$).

a Type I error. The FWE rate is the probability that at least one null hypothesis (e.g., that SBC delivered the same level of service quality on a specified measure to its competitors and to itself) is rejected when, in fact, all of the null hypotheses are true (Hochberg & Tamhane, 1987).

The Bonferroni inequality (Snedecor & Cochran, 1989, p. 116) provides a basis for determining the probability of rejecting at least one null hypothesis when all of the null hypotheses are true. The Bonferroni inequality states that when performing multiple hypothesis tests, the probability of rejecting at least one null hypothesis when all null hypotheses are, in fact, true, is less than or equal to the sum of the Type I error probabilities on each of the tests. The Michigan Plan incorporates a K table, which employs the Bonferroni inequality, to control the FWE rate.[19] The K table controls the FWE rate, in part, by designating a specified number of failed measures as statistical aberrations (also known as ''forgiving'' measures).[20] The extent of the forgiveness for each CLEC depends on the number of measures tested for the CLEC.

For measures on which SBC is judged to be non-compliant after application of the K table, the Michigan Plan requires SBC to make penalty payments directly to the affected CLEC. The penalty rate for non-compliance is $75 for each failed occurrence.[21,22] To determine the total penalty imposed on SBC, this penalty rate is multiplied by an estimate of the number of transactions that failed the parity test on the measure. This penalty estimate is the product of: (i) the failed occurrence penalty rate of $75; (ii) the total number of transactions with the CLEC on the measure; and (iii) the extent to which mean performance for the CLEC on the measure falls short of the critical threshold performance level. This critical threshold is the level of quality for which the probability of observing this particular quality level or a lower level of quality is 5% if

[19]The Critical Z-Statistic Table (Ameritech, 2001, Section 9.3) provides the K values and the associated critical Z values. We refer to this table as the K table. A Bonferroni procedure provides one means to ensure that the FWE rate on n tests does not exceed a desired level, α. The Bonferroni procedure involves reducing the probability of a Type I error on each of the n tests to α/n. When the outcomes of individual tests within a family of tests are positively correlated, the Bonferroni procedure may produce an FWE rate that is less than α. The K table does not account for correlations. An alternative approach that accounts for correlations employs adjusted p-values that are calculated using computer-intensive simulations known as resampling methods. (See, for example, Westfall & Young (1993)). The PMP for SBC in California utilizes different p-values for different sample sizes, but does not adjust for correlations. (See California Public Utilities Commission, 2002, Appendix J).

[20]A ''failed'' parity measure is a measure for which the mean performance that SBC provides to a CLEC is below the mean performance that SBC delivers to itself by an amount that is statistically significant. If a failed measure is designated to be a statistical aberration, SBC is not required to pay the penalty associated with the observed failure, and the so-designated measure is said to be a ''forgiven'' measure.

[21]For persistent sub-standard performance to a CLEC, the per-occurrence penalty increases over a 6 month period from $75 in the first month to $600 in the sixth and following months. For measures that are subject to a cap (e.g., many of the pre-ordering/ordering measures), the total monthy penalty payment to a CLEC cannot exceed $10,000 on any particular measure in the first month. The cap increases over a 6 month period from $10,000 in the first month to $60,000 in the sixth and following months. In January 2003, SBC paid $657,250 in penalties to CLECs in Michigan. In addition to CLEC-specific payments, the Michigan Plan requires SBC to make payments to the Michigan State Treasury for persistent non-compliance across all CLECs. These payments are required when SBC delivers sub-standard performance to all CLECs, in aggregate, for three consecutive months.

[22]In the Texas Plan, penalties take on three distinct values—low, medium, or high—depending on the measure in question.

the ILEC is delivering parity quality to the CLEC on the measure, given the maintained assumptions on the distribution of CLEC service quality.[23]

As an example, suppose the ILEC and CLEC each engage in 30 transactions on a measure that reflects percentages (e.g., the percentage of on-time performances). Further, suppose the ILEC delivers quality level 0.9 (e.g., 90% on-time performance) to itself and 0.7 to the CLEC on this measure. (Notice that a higher number represents higher quality in this setting.) Finally suppose the critical threshold quality level on this measure is 0.8. In this example, the ILEC would fail the measure because mean CLEC quality falls below the critical threshold level of quality of 0.8. Consequently, the $75 penalty rate would take effect. The total penalty imposed on the ILEC for failing this measure would be $225. This penalty is the product of: (i) the penalty rate ($75); (ii) the number of transactions (30); and (iii) the difference between the average quality delivered to the CLEC and the critical threshold (0.1 = 0.8–0.7).[24]

3. Simulation exercises

To illustrate how Type I errors, Type II errors, and financial penalties vary with the realized distributions of service quality under a PMP like the Michigan Plan, the authors conducted an exercise based upon actual recent experience under the Michigan Plan. The exercise is designed to approximate a setting in which an ILEC engages in 30 transactions on each of 35 parity-measures with 100 representative CLECs in a given month.[25] The 35 measures consist of binary measures and non-binary measures. Binary measures (e.g., whether an order was completed or not completed) are recorded as either 1 or 0, and are assumed to follow a binomial distribution. Non-binary measures (e.g., the number of days required to complete an order) reflect the actual level of performance, and are assumed to follow a normal distribution.

The initial simulation described next is intended to reflect a setting where the ILEC provides the same (stochastic) service quality to its competitors that it provides to itself. Such parity provision is simulated by assuming the distribution of service quality that the ILEC provides to the CLECs has the same mean and variance as does the service quality that the ILEC delivers to itself. The mean and variance employed in the simulation reflect the actual performance of the ILEC that operates in Michigan (SBC Michigan). Therefore, the exercise permits an estimate of the number of failed performance measures and the resulting penalties that an ILEC will incur under a PMP

[23]The critical threshold level of quality is the level of quality associated with a Z score of 1.645 (or -1.645, depending upon whether a higher value represents a lower or a higher level of quality).

[24]The calculation is similar for measures that reflect average performance levels rather than percentages. To illustrate, suppose the threshold quality level is 0.5 and the (lower) mean CLEC quality is 0.4 on such a measure. In this case, the difference between the threshold quality and the CLEC mean quality, expressed as a fraction, is 0.2 = [0.5–0.4]/[0.5]. Consequently, with 30 transactions on this measure, the penalty amount would be $450 = [0.2] · [30] · [$75].

[25]A transaction is a response by an ILEC to a request for wholesale service provisioning by a CLEC. The simulation considers only parity measures because benchmark measures are not statistically tested under the Michigan Plan. A potential advantage of benchmark measures is that they can limit any incentive an ILEC might have under parity measures to reduce service quality symmetrically to itself and to its competitors (Sappington & Weisman, 2004).

like the Michigan Plan when it delivers to its competitors the same distribution of service quality that it delivers to itself.[26]

This parity exercise consists of the following nine steps. First, 35 performance measures are chosen randomly from among the parity measures that are monitored under the Michigan Plan.[27] Second, the mean and variance of the service quality that SBC Michigan delivered to its retail affiliate or to its own retail customers (hereinafter "to itself") in the Detroit Metropolitan Market Area during January 2003 on each of these measures are calculated.[28] Third, a "CLEC quality distribution" is constructed for each of the 35 measures, using this same mean and variance and the presumed binomial and normal distributions. Fourth, 30 observations (representing 30 transactions in a month) are drawn randomly from each of the 35 CLEC quality distributions.[29] Fifth, the mean of the thirty observations is calculated. Sixth, the difference between this simulated mean and the mean service quality that the ILEC delivered to itself on the measure in question during the sample period is calculated (for each of the 35 measures). Seventh, the statistical test called for in the Michigan Plan is employed to determine whether the difference between the means is statistically significant.[30] If this difference constitutes a significantly lower level of service quality for the representative CLEC than for the ILEC, the ILEC is deemed to have "failed" the relevant parity test. Otherwise, the ILEC is deemed to have "passed" the test.[31] Eighth, the random sampling of the 30 observations and the testing of the 35 measures is repeated 100 times (representing the ILEC's interaction with 100 CLECs), providing a total of 3500 parity tests. Ninth, the penalties associated with the observed failures of the parity tests are calculated according to the terms of the plan.

After completing this parity exercise, the authors undertook exercises intended to reflect settings in which the ILEC delivers systematically higher and systematically lower levels of service quality to its competitors than to itself. Higher (respectively, lower) levels of service quality to competitors were simulated via a CLEC quality distribution that has the same variance but a

[26]Only CLEC-specific penalties are included in this exercise. (These penalties are referred to as Tier 1 penalties in the Michigan Plan.) For simplicity, penalties payable to the Michigan State Treasury for persistent non-compliance are not considered.

[27]Table A1 lists the selected parity measures. On average, 36 parity measures were tested for each CLEC in Michigan in January 2003. Between November 2002 and January 2003, the average number of parity measures tested monthly for a CLEC ranged from 32 to 36. The exercise was limited to 35 measures for computational convenience.

[28]Data from January 2003 were employed because it represented the most recent data to which the authors had access while preparing an affidavit and supplemental filings for SBC. Data from the Detroit Metropolitan area were employed because this is the geographic area in SBC Michigan's operating territory in which SBC received the most CLEC service requests in January 2003.

[29]To construct the CLEC quality distribution for a given performance measure, 30 observations or transactions are drawn from each distribution for each performance measure. One draw of 30 observations represents service quality provided to one CLEC on one measure in 1 month. Thirty transactions are used in the simulation for testing each performance measure because the sampling distribution of the sample mean approaches a normal probability distribution as the sample size becomes large, and classical statistical tests may not be robust for small sample sizes. In January 2003, 29% of all parity measures tested in Michigan included 30 or more transactions, 15% included between 10 and 29 transactions, and 56% included fewer than 10 transactions.

[30]The critical Z value is 1.645 in the Michigan Plan with no K table and 1.68 (corresponding to 35 performance measures) in the Michigan Plan with the K table.

[31]Under the Michigan Plan with no K table, no observed failed measures are forgiven. Under the Michigan Plan with the K table, a maximum of three failed measures are forgiven for each of the 100 simulated CLECs.

Table 1
Simulated performance under the Michigan Plan with no K table[a]

	Mean service quality delivered to CLECs						
	30% lower	20% lower	10% lower	Parity	10% higher	20% higher	30% higher
Number of measures failed	1129	1032	801	223	131	98	84
Number of measures passed	2371	2468	2699	3277	3369	3402	3416
Percent of measures failed	32.3	29.5	22.9	6.4	3.7	2.8	2.4
Probability of a Type I error on a measure (%)				6.4	3.7	2.8	2.4
Probability of a Type II error on a measure (%)	67.7	70.5	77.1				
Number of CLECs with at least one failed measure	100	100	100	88	78	61	56
Number of CLECs with at least one passed measure	100	100	100	100	100	100	100
Probability of at least one Type I error for a CLEC (%)				88	78	61	56
Probability of at least one Type II error for a CLEC (%)	100	100	100				
Total penalty	$443,730	$231,940	$101,523	$9877	$4955	$4103	$3240

[a]Based on 35 tested measures for 100 CLECs with 30 observations per measure.

more favorable (respectively, less favorable) mean than the mean of the service quality that SBC Michigan delivered to itself. The means employed in the CLEC quality distributions represented levels of service quality that, on average, are 10, 20, and 30% higher (and lower) than the mean quality for SBC Michigan. These non-parity exercises paralleled the parity exercise described above, except for the change in the mean of the CLEC quality distribution. Table A1 in Appendix A lists the performance measures employed in all of the simulations,[32] along with the mean performance values for each of the exercises and the variance of the CLEC quality distribution under parity.

The authors also repeated the parity exercise and all non-parity exercises for cases where there were 100 and 500, rather than 30, observations for each of the 35 measures. The intent of doing so was to better simulate the interaction between an ILEC and a large CLEC, which is likely to involve more than 30 monthly transactions.

4. Results

Tables 1 and 2 summarize the results of the simulation for the setting with 30 observations for each of the 35 measures. Table 1 reports outcomes when the K table is not employed. Table 2

[32]Consider the first performance measure listed in Table A1. SBC's mean performance on this measure is 0.05743, and smaller values indicate better performance on this measure. A 20% increase in service quality to competitors on this measure is captured by a 20% reduction in the mean (to 0.04594) of the CLEC quality distribution. Similarly, a 20% decrease in service quality to competitors is simulated by a 20% increase in the mean of the CLEC quality distribution (to 0.06891). In cases where the relevant (e.g., 20%) increase in quality would exceed the feasible upper bound on quality (e.g., 100% of orders completed on time), the upper bound on quality was employed.

Table 2
Simulated performance under the Michigan Plan with the K table[a]

	Mean service quality delivered to CLECs						
	30% lower	20% lower	10% lower	Parity	10% higher	20% higher	30% higher
Number of measures failed	827	729	500	27	3	1	2
Number of measures passed	2673	2771	3000	3473	3497	3499	3498
Percent of measures failed	23.6	20.8	14.3	0.8	0.1	0.0	0.1
Probability of a Type I error on a measure (%)				0.8	0.1	0.0	0.1
Probability of a Type II error on a measure (%)	76.4	79.2	85.7				
Number of CLECs with at least one failed measure	100	100	100	18	2	1	2
Number of CLECs with at least one passed measure	100	100	100	100	100	100	100
Probability of at Least one Type I error for a CLEC (%)				18	2	1	2
Probability of at least one Type II error for a CLEC (%)	100	100	100				
Total penalty	$406,200	$247,641	$84,934	$2683	$152	$848	$145

[a]Based on 35 tested measures for 100 CLECs with 30 observations per measure.

reports the corresponding outcomes when the K table is employed. (The results for the corresponding settings with 100 and 500 observations per measure are reported in Table B1 in Appendix B.) The first three columns of data in Tables 1 and 2 present results for settings where, on average, the ILEC delivers 30, 20, and 10% lower levels of wholesale service quality to its competitors than to itself. The fourth column presents results for the parity exercise where the ILEC delivers the same distribution of service quality to its competitors that it delivers to itself. The fifth through seventh columns of data present results for settings where, on average, the ILEC delivers 10, 20, and 30% higher levels of service quality to its competitors than it delivers to itself.

The first row of data in Tables 1 and 2 records the number of measures on which the ILEC's simulated performance is judged to be below parity, so the ILEC "fails" the measure. The second row of data presents the number of measures that the ILEC "passes" because the simulated quality it delivers to its competitor is judged to be at least comparable to the level it delivers to itself on the relevant measures. The third row of data states as a percentage the fraction of the total number of measures (3500) the ILEC fails. As the title of the fourth row of data indicates, the percentage of measures failed when the ILEC delivers to its competitors at least the level of service quality it delivers to itself corresponds to the probability of a Type I error on a measure. The associated percentage of measures passed when the ILEC delivers to its competitors a lower level of service quality than it delivers to itself corresponds to the probability of a Type II error on a measure. This Type II error rate is presented in the fifth row of data.

The sixth row of data in Tables 1 and 2 presents the number of CLECs (out of the 100 simulated CLECs) that experienced at least one failed measure in the simulation. The seventh row of data reports the corresponding number of CLECs that experienced at least one passed measure in the simulation. These data underlie the statistics in the eighth and ninth rows of data in Tables

1 and 2. The eighth row of data records the probability that at least one Type I error occurs for a CLEC in the simulation. This probability reflects the percent of CLECs for which the ILEC fails one or more measures despite providing to the CLEC, on average, at least the level of service quality it provides to itself. The ninth row of data provides the probability that at least one Type II error occurs for a CLEC. This probability reflects the percent of CLECs for which the ILEC passes at least one measure despite providing a lower level of service quality to the CLEC, on average, than it provides to itself. The last row of data in Tables 1 and 2 records the penalty payments that the ILEC would be required to make under the Michigan Plan, given the simulated quality realizations.

Tables 1 and 2 provide six general conclusions. First, the probability of failure on a measure increases rapidly as the service quality delivered to CLECs declines below parity levels. Notice, for example, that as the service quality delivered to CLECs declines from parity to 30% below parity, the probability of failure on a measure increases more than five-fold (from 6.4% to 32.3%) when the K table is not employed. The corresponding increase (from 0.8% to 23.6%) is almost 30-fold when the K table is in effect.

Second, compared to these rates at which the probability of failure on a measure increases as CLEC quality declines below parity levels, the probability of failure on a measure does not decline as rapidly when the service quality delivered to CLECs increases above parity levels. For example, as the service quality delivered to CLECs increases from parity to 30% above parity, the probability of failure on a measure declines more than 60% (from 6.4% to 2.4%) when the K table is not employed. The corresponding decline is more than 85% (from 0.8% to 0.1%) when the K table is in effect.[33] These declines are substantially smaller than the corresponding increases in the probability of failure on a measure as CLEC quality declines 30% below parity levels.

Third, the financial penalties imposed on the ILEC respond even more asymmetrically to increases in CLEC quality above parity levels and decreases in CLEC quality below parity levels. Notice, for example, that when the K table is not employed, a 30% increase in CLEC service quality above parity levels reduces the penalties imposed on the ILEC by two-thirds (from \$9877 to \$3240). In contrast, a corresponding reduction in CLEC quality from parity to 30% below parity increases the penalties imposed on the ILEC more than 40-fold (from \$9877 to \$443,730). The asymmetry is even more pronounced when the K table is employed. In this case, ILEC penalties decline by 95% (from \$2683 to \$145) as CLEC service quality increases, on average, by 30% above parity levels. In contrast, ILEC penalties increase more than 100-fold (from \$2683 to \$406,200) as CLEC service quality decreases by 30% below parity levels.[34]

Fourth, as it is intended to do, the K table reduces significantly the probability of a Type I error. When the ILEC delivers the same distribution of service quality to CLECs that it delivers to itself, for example, the K table reduces the probability of a Type I error on a measure from 6.4%

[33]When the K table is in effect, the number of failed measures in the simulation is higher when CLEC quality is 30% above parity levels than when it is 20% above parity levels. This outcome reflects random variation in the simulation with 30 observations per measure. This outcome is not observed with a larger number of observations per measure. (See Table B1)

[34]These asymmetries arise in part because realized quality attains its upper bound on some measures (e.g., measures 10, 11, 20–23, and 32 in Table A1) at a level of CLEC quality that is less than 30% above parity. Consequently, simulating an increase in average quality to 30% above parity does not further increase the mean of the CLEC quality distribution.

to 0.8%. The K table also reduces the probability of at least one Type I error for a CLEC from 88% to 18% in this parity setting. When the ILEC delivers service quality to CLECs that is, on average, 30% higher than the quality it delivers to itself, the K table reduces the likelihood of a Type I error on a measure to almost zero, and reduces to 2% the probability of a Type I error for a CLEC.[35]

Fifth, as it reduces the incidence of Type I errors, the K table increases the likelihood of Type II errors. Notice, for example, that when the ILEC delivers to CLECs a level of service quality that is, on average, 30% below the level of service quality it delivers to itself, the K table increases the probability of a Type II error on a measure from 68% to 76%. There is no corresponding increase in the probability of at least one Type II error for a CLEC because at least one such Type II error occurs in the simulation for all CLECs even when the K table is not employed.

Sixth, while the K table reduces the financial penalties imposed on the ILEC, it does so in an asymmetric manner. The K table introduces much smaller proportionate reductions in penalties when the ILEC provides below-parity quality to CLECs than it does when the ILEC provides above-parity quality to CLECs. Notice, in particular, that the K table reduces the ILEC's financial penalty by less than 9% (from \$443,730 to \$406,200) when, on average, the ILEC delivers a 30% lower level of quality to CLECs than it delivers to itself. In contrast, the K table reduces the ILEC's financial penalty by more than 95% (from \$3240 to \$145) when the ILEC provides a 30% higher level of quality to CLECs, on average.[36,37]

In summary, a PMP like the Michigan Plan that imposes financial penalties but offers no financial rewards can obligate an ILEC to pay penalties persistently even when it provides the same or better wholesale service quality to its competitors than to itself. Statistical procedures like the one based on the K table can limit the incidence of such penalties. They do so, however, at the expense of increasing Type II errors. The K table implements a proportionate reduction in financial penalties that is most pronounced when the ILEC delivers the same or a higher level of quality to CLECs.

5. Symmetric rewards and penalties

If it is deemed important to limit the incidence of Type II errors, one might wish to consider alternatives to the K table. One possible alternative might be to employ symmetric rewards and penalties. For example, an ILEC might receive financial rewards at a rate of \$75 per measure

[35]The K table eliminates Type I errors on all measures for all but one CLEC in the simulation when the quality delivered to CLECs is 20% above parity, and for all but two CLECs when the quality delivered to CLECs is 30% above parity.

[36]When the ILEC provides the same level of quality to CLECs on average that it provides to itself, the K table reduces by approximately 73% (from \$9877 to \$2683) the financial penalty imposed on the ILEC.

[37]As is evident from Table B1, these six general conclusions continue to hold as the number of observations per measure increases. A primary effect of a larger number of observations is to reduce the probability of Type I and Type II errors on a measure. An additional effect is to increase the level of aggregate penalties under parity performance (primarily because penalties are paid on more transactions). The larger number of observations also eliminates the anomalous finding of a smaller percentage of failed measures when CLEC quality is 20% above parity than when CLEC quality is 30% above parity (when the K table is employed).

Table 3
Rewards and penalties under a symmetric Michigan Plan[a]

	Mean service quality delivered to CLECs						
	30% lower	20% lower	10% lower	Parity	10% higher	20% higher	30% higher
Total penalty for significantly lower CLEC quality	\$443,730	\$231,940	\$101,523	\$9877	\$4955	\$4103	\$3240
Total reward for significantly higher CLEC quality	\$111	\$239	\$843	\$5500	\$20,082	\$52,928	\$105,997
Net penalty	\$443,620	\$231,701	\$100,681	\$4377	−\$15,127	−\$48,825	−\$102,757

[a]Based on 35 tested measures for 100 CLECs with 30 observations per measure, and no K table.

when CLEC service quality is judged to be above parity, just as the ILEC incurs the \$75 penalty rate under the Michigan Plan when CLEC service quality is judged to be below parity.

The effects of such a symmetric PMP are illustrated in Table 3. The table reflects the simulated performance described in Table 1, but employs symmetric \$75 reward and penalty rates, rather than only the \$75 penalty rate reflected in Table 1. In particular, the ILEC is afforded a \$75 reward rate on a measure if the mean quality the ILEC delivers to a CLEC on the measure is sufficiently above the corresponding mean quality the ILEC delivers to itself. Similarly, the ILEC incurs a \$75 penalty rate on a measure whenever the mean quality the ILEC delivers to a CLEC on the measure is sufficiently below the corresponding mean quality the ILEC delivers to itself. The ILEC neither receives a reward nor incurs a penalty on a measure if the CLEC mean quality on the measure is sufficiently close to the corresponding ILEC mean quality.[38]

The first row of data in Table 3 replicates the last row of data in Table 1, presenting the total penalty the ILEC incurs because CLEC mean quality realizations are sufficiently below the corresponding ILEC mean quality realizations.[39] The second row of data records the total financial reward the ILEC receives because CLEC mean quality realizations are sufficiently above the corresponding ILEC means. The last row in Table 3 presents the net penalty imposed on the ILEC. This net penalty is the difference between the penalties and the rewards incurred by the ILEC (specified in the first and second rows of data in Table 3, respectively).

Table 3 reveals that a symmetric penalty and reward structure of this type provides a substantial net reward (i.e., a negative net penalty) to an ILEC that delivers above-parity quality to CLECs. The net reward for above-parity quality, though, is less than one-fourth of the net

[38]"Sufficiently above", "sufficiently below", and "sufficiently close" here mean "above", "below", and "between" critical threshold quality levels, respectively. Two critical thresholds are required in a PMP that incorporates both rewards and penalties. Under such a PMP, if higher measured performance reflects higher CLEC quality: (i) rewards are provided when measured CLEC quality exceeds the (higher) upper quality threshold; and (ii) penalties are imposed when measured CLEC quality falls below the lower quality threshold. Neither rewards nor penalties arise when CLEC quality falls above the lower threshold and below the upper threshold. In the simulation reflected in Table 3, the \$75 penalty rate is applied to the estimated number of failed transactions, as explained at the end of Section 2. The \$75 reward rate is applied in analogous fashion to the estimated number of transactions for which CLEC quality exceeds the critical upper threshold. This upper threshold is the quality level for which there is a 5% probability of observing this particular level of CLEC quality or a higher level if the ILEC is providing parity quality to the CLEC.

[39]Notice that these penalties reflect the penalties under the Michigan Plan when there are 30 observations per measure and when the K table is not employed.

penalty for the corresponding level of below-parity quality. To illustrate, the net reward for CLEC quality that is 20% above parity is $48,825, which is less than one-fourth of $231,701, the net penalty the ILEC incurs when CLEC quality is 20% below parity. The symmetric penalty and reward structure also reduces slightly the net penalty imposed on the ILEC when it delivers below-parity quality to CLECs. For example, the decline in the net penalty is: (i) less than 1% (from $101,523 to $100,681) when CLEC quality is 10% below parity; and (ii) less than 0.03% (from $443,730 to $443,620) when CLEC service quality is 30% below parity.

Table 3 also reveals that the symmetric penalty and reward structure reduces by more than half (from $9877 to $4377) the positive net penalty imposed on the ILEC that delivers parity quality to CLECs. Although this reduction in net penalties is substantial, the net penalties are not eliminated. This asymmetry, and the asymmetric net rewards and net penalties for above-parity and below-parity CLEC quality, persist in the presence of symmetric penalty and reward rates in part because SBC's mean performance is close to perfect performance on several measures in the simulation. For example, on measure #21 in Table A1 (the fraction of installations completed before the due date requested by the customer), SBC's mean performance exceeds 0.997. Performance on this measure cannot exceed 1.0. Consequently, performance that constitutes statistically significant improvement relative to mean performance on this measure, and on other measures in the simulation is impossible. Therefore, no financial rewards are available in practice on these measures, despite the symmetric penalty and reward rates.

In summary, symmetric reward and penalty rates are not necessarily a panacea for the effects of asymmetric PMPs, especially in settings where the ILEC is providing service quality that is close to the best possible service quality. Alternative possible modifications of PMPs are discussed briefly in the next section.

6. Conclusions

The study employed simulations to examine how Type I and Type II errors, and financial penalties vary with realized quality distributions under PMPs like the Michigan Plan. The authors found that an ILEC can incur substantial penalties under a PMP even when, on average, it delivers significantly higher quality to CLECs than to itself. These penalties can be reduced through the use of a K table, but this reduction comes at the expense of more frequent Type II errors. More symmetric financial rewards and penalties can reduce the penalties imposed on an ILEC that delivers relatively high levels of service quality to CLECs without reducing substantially the penalties imposed on an ILEC that delivers relatively low levels of service quality to CLECs. However, symmetric policies of this sort may not eliminate the net penalties imposed on an ILEC that provides parity quality to CLECs, particularly if the ILEC is providing very high levels of service quality to itself.

Future research should analyze the effects of alternative modifications of PMPs. At least three distinct modifications merit investigation. First, different measures might be afforded different weights in constructing an overall service quality index. Financial penalties (and rewards) might then be based on realized performance on this overall index. Such an index approach has the advantage of placing different emphasis on different dimensions of service quality, some of which may be more important than others to CLECs. An index approach also allows an ILEC to determine how it can achieve any specified level for the index at minimum cost. Thus, if the weights employed in the index reflect how highly CLECs value improved performance on the various dimensions of service quality, an index approach can induce an ILEC to expend its resources efficiently in serving the needs and best interests of CLECs.[40] In practice, the weights employed in the overall service quality index might result from negotiations between ILECs, CLECs, and consumer representatives (e.g., public service commission staff). Since CLECs are not homogeneous, different weights might conceivably be employed for different CLECs.

Second, a PMP might be explicitly designed to minimize a weighted average of Type I and Type II errors. The relevant weights could reflect estimates of the losses that occur when Type I and Type II errors are committed under the plan. As indicated above, a trade-off between Type I and Type II errors is inevitable. A PMP that minimizes a weighted average of these two types of errors recognizes this trade-off explicitly, and has the potential to manage the trade-off efficiently.[41]

Third, nonlinear penalty (and reward) structures might be implemented. A small increment in quality on a particular dimension may be of little value to a CLEC when the CLEC is already receiving high levels of quality on this dimension. However, this same small increment may be of great value to the CLEC when it is receiving a low level of quality on the dimension in question. Penalty (and reward) structures that reflect this differential value may better induce an ILEC to develop a mix of service quality on multiple dimensions that best serves the needs of CLECs. The ideal structuring of such a nonlinear penalty (and reward) structure awaits careful investigation. Future research might consider explicitly the costs of monitoring realized service quality and the possibility that realized performance on different quality measures may not be independent. In addition, future research might account for the possibility that realized CLEC quality can depend upon actions undertaken by the CLEC as well as actions undertaken by the ILEC.[42]

Acknowledgement

This paper is based on work the authors have submitted to the Michigan Public Service Commission on behalf of SBC Communications in Case No. U–11830. In the matter of SBC's submission on performance measurements, reporting, and benchmarks in compliance with the October 2, 1998 order in MPSC Case No. U-11654. The views expressed here are solely those of the authors. The helpful comments provided by James Ehr, Martin Grambow, Mary Mulry, and three anonymous referees are gratefully acknowledged. We also thank Ian Heavers and Justin Bledin for invaluable research assistance.

[40] See Berg and Lynch (1992) and Lynch, Buzas, and Berg (1994) for analyses of the design of retail service quality indexes. The PMP in Georgia (BellSouth, 2003) assesses overall performance on a measure by effectively considering a weighted average of performance on relevant sub-measures, where the weights reflect the relative numbers of transactions. Although this PMP continues to test aggregate performance measures individually, it allows some trade-off of performances on the multiple sub-measures that constitute each measure.

[41] The PMP in Georgia (Bell South, 2003) incorporates an explicit balancing of Type I and Type II errors.

[42] Of course, such "double moral hazard" considerations (e.g., Demski & Sappington, 1991; Gupta & Romano, 1998) can complicate the design of optimal penalty and reward structures.

Appendix A

Means and variance for the 35 measures employed in the simulations are given in Table A1.

Table A1
Means and variance for the 35 measures employed in simulations

	Tracking number	Measure name	SBC variance	Mean of CLEC quality distribution for simulation						
				30% lower	20% lower	10% lower	SBC mean	10% higher	20% higher	30% higher
1	38	% SBC/Ameritech caused missed due dates—POTS—Res—FW	0.05398	0.07465	0.06891	0.06317	0.05743	0.05168	0.04594	0.04020
2	41	% SBC/Ameritech caused missed due dates—POTS—Bus—No FW	0.00571	0.00749	0.00692	0.00634	0.00576	0.00519	0.00461	0.00403
3	50	% Trouble reports w/in 30 days of install—POTS—Res—FW	0.09253	0.13453	0.12418	0.11383	0.10349	0.09314	0.08279	0.07244
4	51	% Trouble reports w/in 30 days of install—POTS—Res—No FW	0.04261	0.05800	0.05353	0.04907	0.04461	0.04015	0.03569	0.03123
5	77	% Repeat reports—POTS—bus	0.08492	0.12231	0.11290	0.10349	0.09408	0.08467	0.07527	0.06586
6	172	% SBC/Ameritech caused missed due dates—UNE—BRI loop with test access	0.08373	0.14034	0.12955	0.11875	0.10795	0.09716	0.08636	0.07557
7	174	% SBC/Ameritech caused missed due dates—UNE—DS1 loop with test access	0.02928	0.04426	0.04085	0.03745	0.03404	0.03064	0.02723	0.02383
8	184	Of installation—UNE—8.0 dB loop w/out test access	0.09144	0.13261	0.12241	0.11221	0.10201	0.09181	0.08161	0.07141
9	303	Mean time to restore—UNE—8.0 dB loop without test access (hrs)—dispatch	367.40972	19.81973	18.29513	16.77054	15.24594	13.72135	12.19676	10.67216
10	372	% Out of service (OOS) <24 h—POTS—Residence	0.08351	0.63662	0.72757	0.81852	0.90946	1.00000	1.00000	1.00000
11	373	% Out of service (OOS) <24 h—POTS—Business	0.07096	0.65040	0.74332	0.83623	0.92915	1.00000	1.00000	1.00000
12	423	% SBC/Ameritech caused missed due dates > 30 days—POTS—Bus—No FW	0.00091	0.00118	0.00109	0.00100	0.00091	0.00082	0.00073	0.00064
13	498	30 days—UNE—8.0 dB loop without test Access	0.00166	0.00216	0.00199	0.00183	0.00166	0.00149	0.00133	0.00116
14	501	% SBC/Ameritech caused missed due dates > 30 days—UNE—BRI loop with test access	0.00578	0.00828	0.00764	0.00701	0.00637	0.00573	0.00510	0.00446
15	541	% SBC/Ameritech missed due dates due to lack of facilities—POTS—Res	0.03015	0.04058	0.03746	0.03433	0.03121	0.02809	0.02497	0.02185

16	573	Lack of facilities—UNE—BRI loop with test access	0.03636	0.05382	0.04968	0.04554	0.04140	0.03726	0.03312	0.02898
17	590	To lack of facilities—UNE—8.0 dB loop without test access	0.00128	0.00167	0.00154	0.00141	0.00129	0.00116	0.00103	0.00090
18	591	To lack of facilities—UNE—8.0 dB loop without test access	0.00016	0.00021	0.00019	0.00018	0.00016	0.00014	0.00013	0.00011
19	596	To lack of facilities—UNE—BRI loop with test access	0.00578	0.00828	0.00764	0.00701	0.00637	0.00573	0.00510	0.00446
20	936	% Installations completed w/in customer requested due date—UNE-P—Res—FW	0.06339	0.65465	0.74817	0.84169	0.93522	1.00000	1.00000	1.00000
21	937	% Installations completed w/in customer requested due date—UNE-P—Res—No FW	0.00301	0.69808	0.79781	0.89753	0.99726	1.00000	1.00000	1.00000
22	938	% Installations completed w/in customer requested due date—UNE-P—Bus—FW	0.03405	0.67800	0.77486	0.87171	0.96857	1.00000	1.00000	1.00000
23	939	% Installations completed w/in customer requested due date—UNE-P—Bus—No FW	0.00811	0.69652	0.79603	0.89553	0.99504	1.00000	1.00000	1.00000
24	941	% SBC/Ameritech caused missed due dates—UNE-P—Res—No FW	0.00403	0.00527	0.00486	0.00446	0.00405	0.00365	0.00324	0.00284
25	943	% SBC/Ameritech caused missed due dates—UNE-P—Bus—No FW	0.00571	0.00749	0.00692	0.00634	0.00576	0.00519	0.00461	0.00403
26	944	% SBC/Ameritech missed due dates due to lack of facilities—UNE-P—Res	0.03015	0.04058	0.03746	0.03433	0.03121	0.02809	0.02497	0.02185
27	955	Lack of facilities—UNE-P—Res—>30 calendar days	0.00127	0.00166	0.00154	0.00141	0.00128	0.00115	0.00102	0.00090
28	956	Lack of facilities—UNE-P—Res—>90 calendar days	0.00010	0.00013	0.00012	0.00011	0.00010	0.00009	0.00008	0.00007
29	968	% Trouble reports w/in 30 days of install—UNE-P—Res—FW	0.09253	0.13453	0.12418	0.11383	0.10349	0.09314	0.08279	0.07244
30	978	% Missed repair commitments—UNE-P—Bus—dispatch	0.08472	0.12217	0.11277	0.10337	0.09397	0.08458	0.07518	0.06578
31	983	Rcpt to clear duration—UNE-P—Res—no dispatch—out of service (hrs)	29.89221	3.40291	3.14115	2.87939	2.61763	2.35586	2.09410	1.83234
32	994	% Out of service (OOS)—<24 h—UNE-P—Bus	0.07096	0.65040	0.74332	0.83623	0.92915	1.00000	1.00000	1.00000
33	1070	% SBC/Ameritech caused missed due dates—>30 days—UNE-P—Res—No FW	0.00004	0.00005	0.00004	0.00004	0.00004	0.00003	0.00003	0.00002
34	1088	Avg installation interval—DSL—with line sharing—without conditioning	0.31864	3.85771	3.56097	3.26422	2.96747	2.67072	2.37398	2.07723
35	1251	30 of installation—UNE—DSL loops—line sharing	0.02100	0.02805	0.02589	0.02373	0.02157	0.01942	0.01726	0.01510

Appendix B

The results of simulations with 100 and 500 observations per measure are given in Table B1.

Table B1
Simulated performance under the Michigan Plan

	Mean service quality delivered to CLECs						
	30% lower	20% lower	10% lower	Parity	10% higher	20% higher	30% higher
(a) *Without the K table*[a]							
Number of measures failed	1279	1125	998	191	108	89	56
Number of measures passed	2221	2375	2502	3309	3392	3411	3444
Percent of measures failed	36.5	32.1	28.5	5.5	3.1	2.5	1.6
Probability of a Type I error on a measure (%)				5.5	3.1	2.5	1.6
Probability of a Type II error on a measure (%)	63.5	67.9	71.5				
Number of CLECs with at least one failed measure	100	100	100	92	69	62	40
Number of CLECs with at least one passed measure	100	100	100	100	100	100	100
Probability of at least one Type I Error for a CLEC (%)				92	69	62	40
Probability of at least one Type II error for a CLEC (%)	100	100	100				
Total penalty	$1,644,546	$859,591	$416,970	$17,139	$5659	$3796	$1812
(b) *With the K table*[a]							
Number of measures failed	978	822	696	9	4	0	0
Number of measures passed	2522	2678	2804	3491	3496	3500	3500
Percent of measures failed	27.9	23.5	19.9	0.3	0.1	0.0	0.0
Probability of a Type I error on a measure (%)				0.3	0.1	0.0	0.0
Probability of a Type II error on a measure (%)	72.1	76.5	80.1				
Number of CLECs with at least one failed measure	100	100	100	8	4	0	0
Number of CLECs with at least one passed measure	100	100	100	100	100	100	100

Probability of at least one Type I error for a CLEC (%)				8	4	0	0
Probability of at least one Type II error for a CLEC (%)	100	100	100				
Total penalty	$1,554,137	$932,798	$371,919	$1301	$412	$0	$0
(c) *Without the K table*[b]							
Number of measures failed	1621	1365	1070	154	43	30	26
Number of measures passed	1879	2135	2430	3346	3457	3470	3474
Percent of measures failed	46.3	39.0	30.6	4.4	1.2	0.9	0.7
Probability of a Type I error on a measure (%)				4.4	1.2	0.9	0.7
Probability of a Type II error on a measure (%)	53.7	61.0	69.4				
Number of CLECs with at least one failed measure	100	100	100	79	37	24	24
Number of CLECs with at least one passed measure	100	100	100	100	100	100	100
Probability of at least one Type I error for a CLEC (%)				79	37	24	24
Probability of at least one Type II error for a CLEC (%)	100	100	100				
Total penalty	$9,597,938	$4,764,747	$2,587,936	$28,009	$3319	$1424	$1123
(d) *With the K table*[b]							
Number of measures failed	1309	1059	764	2	0	0	0
Number of measures passed	2191	2441	2736	3498	3500	3500	3500
Percent of measures failed	37.4	30.3	21.8	0.1	0.0	0.0	0.0
Probability of a Type I error on a measure (%)				0.1	0.0	0.0	0.0
Probability of a Type II error on a measure (%)	62.6	69.7	78.2				
Number of CLECs with at least one failed measure	100	100	100	2	0	0	0
Number of CLECs with at least one passed measure	100	100	100	100	100	100	100
Probability of at least one Type I error for a CLEC (%)				2	0	0	0
Probability of at least one Type II error for a CLEC (%)	100	100	100				
Total penalty	$9,489,535	$5,777,276	$2,244,296	$877	$0	$0	$0

[a]Based on 35 tested measures for 100 CLECs with 100 observations per measure.

[b]Based on 35 tested measures for 100 CLECs with 500 observations per measure.

References

Ameritech. (2001). The Ameritech Michigan performance remedy plan description, September 12.

AT&T Communications of Michigan, Inc. (2002). TCG Detroit, MCIMetro Access Transmission Services, LLC., Brooks Fiber Communications of Michigan, Inc., and MCI WorldCom Communications, Inc., Comments in response to the Michigan Public Service Commission's February 25, 2002 Opinion and Order (Case No. U-11830), In the matter of SBC's submission on performance measurements, reporting, and benchmarks in compliance with the October 2, 1998 Order in MPSC Case No. U-11654, filed April 5, 2002.

BellSouth, SBC, Qwest, & Verizon (2002). UNE Fact Report, 2002. Submitted to the Federal Communications Commission. In the matter of Review of Section 251 Unbundling Obligations of Incumbent Local Exchange Carriers, CC Docket No. 01-338; Implementation of the Local Competition Provisions in the Telecommunications Act of 1996, CC Docket No. 96–98; and Deployment of Wireline Services Offering Advanced Telecommunications Capability, CC Docket No. 98–147, April 2002.

BellSouth. (2003). Georgia self-effectuating enforcement mechanism administration plan, version 1.2, July 22, 2003.

Berg, S., & Lynch, J. (1992). The measurement and encouragement of telephone service quality. *Telecommunications Policy*, *16*(3), 210–224.

California Public Utilities Commission. (2002). *Opinion on the performance incentives plan for Pacific Bell Telephone Company*. Decision 02-03-023, March 5, 2002.

Demski, J., & Sappington, D. (1991). Resolving double moral hazard problems with buyout agreements. *Rand Journal of Economics*, *22*(2), 222–240.

Gupta, S., & Romano, R. (1998). Monitoring the principal with multiple agents. *Rand Journal of Economics*, *29*(2), 427–442.

Hochberg, Y., & Tamhane, A. (1987). *Multiple comparison procedures*. New York: Wiley.

Lynch, J., Buzas, T., & Berg, S. (1994). Regulatory measurement and evaluation of telephone service quality. *Management Science*, *40*(2), 169–194.

Michigan Public Service Commission. (2001). Opinion and Order (Case No. U-11830). In the matter of Ameritech Michigan's submission on performance measurements, reporting, and benchmarks pursuant to the October 2, 1998 Order in Case No. U-11654, July 25, 2001.

Sappington, D., & Weisman, D. (2004). Competition policy, parity regulation, and self-sabotage. *Info*, *6*(1), 52–56.

SBC/Ameritech. (2001). *Performance measurement user guide*, version 1.9. ' 1 to the Ameritech Michigan Performance Remedy Plan, September 12, 2001.

Snedecor, G., & Cochran, W. (1989). *Statistical methods* (8th ed). USA: Iowa State Press.

Telecommunications Act (1996). Pub. L. No. 104-104, 110 stat. 56 (1996) (codified at 47 USC. Sections 151 et seq).

Westfall, P., & Young, S. (1993). *Resampling-based multiple testing*. New York: Wiley.