Motor Carrier Industry Profile Study: Statistical Inference of Safety Performance Measures

Analysis Division,
Office of Information Management,
Federal Motor Carrier Safety Administration
October 23, 2002

FMCSA-RI-02-008

Table of Contents

Motor Carrier Industry Profile Study: Statistical Inference of Safety Performance Measures

Thomas P. Keane, *Analysis Division, Federal Motor Carrier Safety Administration*
William C. Horrace, *Department of Economics, University of Arizona*
Kristine N. Braaten, *Econometrica, Inc.*

Management Summary

Introduction

The primary mission of the Federal Motor Carrier Safety Administration (FMCSA) is to prevent commercial motor vehicle-related fatalities and injuries. The FMCSA contributes to ensuring safety in motor carrier operations through public education and outreach, as well as regulatory enforcement. The motor carrier industry is highly diverse and competitive, comprised of many unique types of operations and hauling many different types of commodities. In an effort to better understand the diverse nature of this industry and explore safety and operational differences among its major segments, FMCSA, with the University of Maryland, College Park, undertook the Motor Carrier Industry Profile Study. The study used the Motor Carrier Information Management System (MCMIS) and the Motor Carrier Safety Status Measurement System (SafeStat) as its sources. The MCMIS and SafeStat are maintained by the FMCSA and are populated with data from roadside inspections, FMCSA and State compliance reviews, crashes, and enforcement cases against motor carriers.

The study examines the recent safety performance of 11 For-Hire and 10 Private segments of the motor carrier industry, using nine driver-related, crash-related, vehicle-related and safety management-related measures for each segment in each motor carrier sector. The 10 Private commodity segments examined were: Building Materials, Bulk Freight, Refrigerated (non-produce), General Freight-Truckload, Household Goods, Intermodal, Large Machinery, Passenger, Produce, and Tank. The 11 For-Hire segments examined in the study included the 10 commodity segments referenced above, plus the Less-Than-Truckload (LTL) segment. The nine safety performance measures were: Driver Safety Evaluation Area (SEA), Driver Out-Of Service (OOS) Rate, Vehicle SEA, Vehicle OOS Rate, Accident SEA, Fatal Crash Rate, Total Crash Rate, Safety Management Review Measure (SMRM), and Enforcement Severity Measure (ESM). The mean safety scores of each For-Hire and Private segment were compared, respectively, to their peer segments for each of the nine safety performance measures using a simple ranking system. For each safety performance measure, a segment received a ranking based on its performance relative to all other segments analyzed in the study. Statistical inference was then performed on the safety rankings for each segment to determine which segments performed best and worst in each safety category. A description of the data and safety measures are contained in the Final Report, entitled: "Motor Carrier Industry Profile Study: Evaluating Safety Performance by Motor Carrier Industry Segment".

Results

Results of the study are contained in Table S1a and S1b for the For-Hire sector and the Private sectors, respectively. Consider Table S1a. The first column contains the nine safety measures as defined in the Final Report. The other columns in the table are the motor carrier segments and are populated by the average segment score for each safety measure (the decimal value), the relative ranks of each sector (the parenthetical values) and sometimes either the designator “B” or “W” for each safety measure. (For example, the Refrigerated segment received a Driver SEA score of 51.03, a rank of 11 out of 11, and the designator “W”.) The designators B and W are used for the purpose of statistical inference and imply the following:

B = Segment was statistically best for the safety measure with 95% confidence.

W = Segment was statistically worst for the safety measure with 95% confidence.

When a cell in Table S1a contains no B or W it implies that the segment was neither best nor worst for the safety measure. When a cell contains both a B and a W, it means that for that measure the segment is both best and worst. This implies that the estimation error (or sampling variability) associated with that cell was too high for the analysis to determine if the segment was best or worst. This was typically caused by too small a sample size for that particular cell.

Notice that for a particular safety measure (any row) that there can be several segments that are best (B) or worst (W). For example, looking at the Vehicle OOS Rate measure, the best segments (those with B’s) are the Passenger segment (score = 18.50, rank = 1 out of 11) and the “General Freight Less-Than-Truckload” segment (score = 20.12, rank = 2 out of 11). This means that the statistical analysis could not differentiate between these two segments in terms of which was the best for Vehicle OOS Rate. This could be caused by the point estimates themselves being similar in magnitude or by the estimates containing a high degree of estimation error (or sampling variability), usually caused by smaller sample sizes. It also means that the segments should be treated as if they were both "best" with 95% confidence even though the scores and ranks are different. For the Vehicle OOS Rate it is also the case that the worst performing segments (those with W's) are Intermodal and Large Machine.

When a particular safety measure (any row) contains many B's and W's, it implies that the statistical inferences for that measure were relatively "less sharp" than for a safety measure with few B's and W's. When a measure contains only one B and one W, the inference was very sharp. Considering the Driver SEA measure (first row), the single best segment is the Passenger segment and the single worst is the Refrigerated segment. This is a strong inference statement because it identifies single segments as the safety extrema at the 95% confidence level; the inference is very sharp. This is also the case for the Driver OOS Rate and the Vehicle SEA measures. Overall, the Passenger and the "General Freight Less Than Truckload" segments perform the best, being "best" 6 times each. The worst segments are the Intermodal and Produce segments, being worst 4 times each.

Turning to the Private sector results of Table S1b, the inferences tended to be less sharp than the For-Hire results. For example, in the Driver SEA category, the best segment is the Tank segment, but the worst set consisted of five segments (Household, Intermodal, Large Machine, Passenger and Produce). Also, there tend to be a lot of "B,W" designations throughout the table. The Fatal Crash Rate and the ESM tend to be the least reliable safety measures in terms of discerning a best and worst segment. However, the Safety Management Review Measure (SMRM) is also fairly unreliable. All three measures have a multiplicity of segments that are both in the best and in the worst subsets. The sharpest overall inference is for the Driver OOS Rate, Vehicle SEA, Vehicle OOS rate and the Accident SEA. The only safety measures that produce a single best or worst segment are the Driver SEA, which determined that the Tank segment was the single best; the Accident SEA, which determined that the Refrigerated segment was the single worst; and the Total Crash Rate, which determined that the Passenger segment was the single best. For the Private sector there is no clear overall best or worst segment.

Conclusions

This inferential study is an excellent start to a larger analysis of the measures obtained from the SafeStat data. It has highlighted the limitations and the strengths of measures in various motor-carrier segments insofar as the inferences of Tables S1a and S1b tended to have variable levels of "sharpness" for different safety measures. Without rehashing specific results, some strong conclusions have surfaced. First and foremost, the segment rankings by themselves need to be interpreted with caution. This is not to say that the segment rankings are wrong, but it is to say that inferential procedures are necessary to get a true sense of which segments are best (B) and worst (W). Future research efforts should focus on ways to sharpen the inference for those safety measures where there were many segments that were "best" and/or "worst". This could include increasing the number of observations of each safety measure (collecting new data), collecting more informative data (more details), or exploring alternative ways to analyze the data.

This research also implies a comprehensive firm-level survey study of the best and worst segment performers to uncover why they do so well or so poorly in each safety category. This has been started with the design of a "Best Practices Survey" which surveys the individual performance of 250 carriers. Finally, the results of this research may have strong Federal policy implications. Perhaps in the future, Federal regulations could be tailored to improve the performance of the worst segments based on the results of subsequent studies. As new and improved data become available and are analyzed, the results of future research will yield improved estimates and inferences suitable for policy decision-making.

Table S1a. For-Hire Trucking Segment-by-Segment Results: Relative Ranks

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Less Than Truck Load	General Freight Truck Load	Househld Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related											
Driver SEA	40.41 (7)	37.08 (4)	51.03 (11) W	36.52 (3)	43.67 (8)	44.89 (9)	39.04 (5)	39.06 (6)	21.36 (1) B	48.57 (10)	36.49 (2)
Driver OOS Rate	9.36 (6)	8.86 (5)	12.15 (10)	4.76 (1) B	11.18 (8)	13.78 (11) W	8.08 (4)	9.64 (7)	6.93 (2)	12.06 (9)	7.32 (3)
Vehicle-Related											
Vehicle SEA	50.23 (9)	47.30 (6)	46.02 (5)	41.33 (2)	47.53 (8)	44.07 (3)	57.23 (11) W	51.33 (10)	30.87 (1) B	47.49 (7)	44.51 (4)
Vehicle OOS Rate	27.04 (9)	25.53 (8)	23.82 (5)	20.12 (2) B	25.25 (6)	23.28 (3)	28.61 (10) W	28.95 (11) W	18.50 (1) B	25.35 (7)	23.33 (4)
Crash-Related											
Accident SEA	11.67 (7)	9.96 (3)	12.85 (8)	30.80 (11) W	11.23 (6)	7.86 (2) B	14.28 (9)	9.96 (3)	7.67 (1) B	10.86 (5)	15.82 (10)
FAT_CR_D (Fatal Crash Rate)	0.015 (7)	0.022 (11) W	0.017 (9) W	0.003 (1) B	0.015 (7)	0.013 (4) W	0.014 (5) W	0.014 (5)	0.007 (2) B	0.020 (10) W	0.009 (3)
TOT_CR_D (Total Crash Rate)	0.342 (7)	0.411 (10) W	0.363 (9)	0.091 (1) B	0.352 (8)	0.302 (5)	0.241 (4)	0.329 (6)	0.219 (3)	0.418 (11) W	0.205 (2)
Others											
Safety Management Review Measure	20.45 (4)	22.88 (9)	21.26 (6)	9.44 (1) B	22.17 (8)	29.32 (11) W	20.85 (5)	21.53 (7)	17.09 (3)	23.31 (10) W	12.66 (2) B
Enforcement Severity Measure	2.50 (5)	2.29 (2)	3.12 (10) W	2.50 (5) B,W	2.60 (8)	2.51 (4)	2.62 (9) W	2.43 (3)	1.57 (1) B	3.14 (11) W	2.51 (7)

Numbers in parentheses are the relative ranks of the segments for each safety measure.
Definitions of safety measures are contained in the Final Report.
B = Segment is best with 95% confidence.
W = Segment is worst with 95% confidence.
For Enforcement Severity Measure, “General Freight Less Than Truckload” only had 19 observations, so inference is less sharp.

Table S1b. Private Trucking Segment-by-Segment Results: Relative Ranks

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related										
Driver SEA	25.89 (3)	25.63 (2)	28.53 (4)	29.40 (6)	30.05 (7) W	28.69(5) W	31.17 (8) W	34.60 (10) W	31.58 (9) W	20.70 (1) B
Driver OOS Rate	9.98 (5)	8.33 (2)	9.11 (4)	12.24 (9) W	13.84 (10) W	8.75 (3) B	11.61 (7)	11.75 (8) W	10.54 (6)	6.43 (1) B
Vehicle-Related										
Vehicle SEA	47.32 (7)	46.53 (6)	31.58 (2) B	38.16 (4)	28.67 (1) B	50.50 (8) W	56.03 (10) W	51.05 (9) W	37.16 (3)	42.72 (5)
Vehicle OOS Rate	27.90 (6)	28.56 (8)	16.41 (2) B	21.45 (4)	15.18 (1) B	28.25 (7)	33.66 (10) W	28.81 (9) W	19.96 (3)	24.17 (5)
Crash-Related										
Accident SEA	4.81 (5)	5.32 (7)	7.28 (10) W	3.81 (3)	2.74 (1) B	4.82 (6) B	4.13 (4)	2.74 (1) B	5.83 (9)	5.57 (8)
FAT_CR_D (Fatal Crash Rate)	0.013 (7) W	0.016 (9) W	0.011 (5) W	0.018 (10) W	0.003 (2) B	0.011 (5) B,W	0.014 (8) W	0.000* (1) **	0.010 (3) B	0.010 (3) W
TOT_CR_D (Total Crash Rate)	0.277 (5)	0.359 (9) W	0.224 (4)	0.372 (10) W	0.310 (8) W	0.199 (2)	0.286 (6)	0.080* (1) B	0.286 (6)	0.207 (3)
Others										
Safety Management Review Measure	31.53 (4) W	32.73 (5) W	35.05 (9) W	33.89 (8) W	28.22 (3) B,W	39.34 (10) B,W	33.84 (7) W	26.67 (2) B,W	33.43 (6) W	20.52 (1) B
Enforcement Severity Measure	1.48 (7) W	1.30 (4)	2.40 (10) W	0.98 (2)	1.34 (5) B,W	1.08 (3) B,W	1.45 (6)	0.14 (1) B,W	2.19 (9) W	1.50 (8) W

Numbers in parentheses are the relative ranks of the segments for each safety measure.
Definitions of safety measures are contained in the Final Report.
B = Segment is best with 95% confidence.
W = Segment is worst with 95% confidence.
** Inference could not be performed due to a complete lack of variability in the data (all observations the same).

Motor Carrier Industry Profile Study: Statistical Inference of Safety Performance Measures

Thomas P. Keane, *Analysis Division, Federal Motor Carrier Safety Administration*
William C. Horrace, *Department of Economics, University of Arizona*
Kristine N. Braaten, *Econometrica, Inc.*

1.0 Introduction

The primary mission of the Federal Motor Carrier Safety Administration (FMCSA) is to prevent commercial motor vehicle-related fatalities and injuries. The FMCSA contributes to ensuring safety in motor carrier operations through public education and outreach, as well as regulatory enforcement. The motor carrier industry is highly diverse and competitive, comprised of many unique types of operations and hauling many different types of commodities. In an effort to better understand the diverse nature of this industry and explore safety and operational differences among its major segments, the FMCSA, with the University of Maryland, College Park, undertook the Motor Carrier Industry Profile Study. Results of this study are found in Keane, Corsi and Braaten (2001), henceforth the KCB study.

The KCB study examines the recent safety performance of 11 For-Hire and 10 Private segments (e.g. Household Goods, Passenger, and Produce) of the motor carrier industry. The safety performance differences between individual segments are examined, as well as the differences between the For-Hire and Private sectors as a whole. The study used the Motor Carrier Information Management System (MCMIS) and the Motor Carrier Safety Status Measurement System (SafeStat) as its sources. The MCMIS and SafeStat are maintained by the FMCSA and are populated with data from roadside inspections, FMCSA and State compliance reviews, crashes, and enforcement cases against motor carriers.

The KCB study estimates safety performance for nine driver-related, crash-related, vehicle-related and safety management-related measures for each segment of the For-Hire and Private motor carrier sectors. It then ranks each segment of the motor carrier industry in terms of each safety measure estimate. For example, they find that the “General Freight Less-Than-Truckload” and Passenger segments are safest in terms of fatal accident rates within the For-Hire segment. They also conclude that For-Hire Produce and Bulk Freight segments are the least safe in terms of fatal accidents.

While the KCB study is thorough and compelling, the authors correctly point out one limitation of their results: that their safety estimate rankings lack rigorous statistical inference. The purpose of this paper is to revisit the KCB study and apply the theory of Ranking and Selection to perform inference on their results. This paper also proposes an alternative estimator of each safety measure in each segment using an Ordinary Least Squares (OLS) estimator of a General Linear Model that accounts for multiple-segment firms in the sample; the KCB safety measures do not. The paper is organized as follows. The next section summarizes the results of the KCB study and proposes an alternative OLS safety estimator based on the same data. Section 3 describes the theory behind ranking and selection inference procedures. Section 4 provides the ranking and selection results for the KCB study and for the proposed OLS safety measures, and Section 5 summarizes and concludes.

2.0 The KCB Study

The KCB study examines the recent safety performance of 11 For-Hire and 10 Private segments of the motor carrier industry. The safety performance differences between individual segments are examined, as well as the differences between the For-Hire and Private sectors as a whole. In their analysis, safety performance is evaluated according to two driver-related safety measures, two vehicle safety measures, three crash-related measures, and two safety management performance measures (nine measures in all). The study uses the Motor Carrier Information Management System (MCMIS) and the Motor Carrier Safety Status Measurement System (SafeStat) as its sources. The MCMIS and SafeStat are maintained by FMCSA and are populated with data from roadside inspections, FMCSA and State compliance reviews, crashes, and enforcement cases against motor carriers.

In the KCB study, the 10 Private commodity segments examined were: Building Materials, Bulk Freight, Refrigerated (non-produce), General Freight-Truckload, Household Goods, Intermodal, Large Machinery, Passenger, Produce, and Tank. The 11 For-Hire segments examined in their study included the 10 commodity segments referenced above, plus the Less-Than-Truckload (LTL) segment. The mean scores of each For-Hire and Private segment were compared, respectively, to its peer segments on each of the nine safety performance measures using a rudimentary ranking system. For each safety performance measure, a segment received a ranking based on its performance relative to all other For-Hire or Private segments, respectively, analyzed in the study. A description of the data and safety measures are contained in the Data Appendix. The next section defines the KCB estimation algorithm.

2.1 KCB Safety Measurement

Let Y_j^k represent an observation j (j = 1, …, J) of safety measure k (k = 1, …, K) in no particular motor-carrier segment. Let d_{ij}^k be an indicator variable equal to 1 if the j^{th} observation is from segment i (i = 1, …, M), equal to "0" otherwise. Even though safety measure k is made at the motor-carrier firm level, it should be noted that the unit of observation j does not index motor-carrier firms. Indeed, firms can operate within multiple segments, so for a given firm, its value of safety measure k can occur in the data set multiple times, if it operates in multiple segments. (Unfortunately, indexing Y by firm *and* by segment is not practical, since the data are not collected at such a disaggregate level; they are collected at the firm level. Additionally, the proposed notation is useful for the General Linear Model that is proposed in the sequel.)

Based on this notation, $N_i^k = \sum_{j=1}^{J} d_{ij}^k$ (k = 1, …, K) is the number of observations of safety measure k in segment i. Then the KCB study calculates:

$$\overline{Y}_i^k = \frac{1}{N_i^k} \sum_{j=1}^{J} d_{ij}^k Y_j^k \text{ , } k = 1, \ldots, K, i = 1, \ldots, M$$

as a measure of the safety performance of segment i for safety measure k. This implies the sample variance measure:

$$s_{ik}^2 = \frac{1}{(N_i^k - 1)} \sum_{j=1}^{J} d_{ij}^k (Y_j^k - \overline{Y}_i^k)^2 \text{ , } k = 1, \ldots, K, i = 1, \ldots, M.$$

The segment safety measures $\overline{Y}_i^k$ (i = 1, …, M), imply rank statistics for each safety measure k:

$$\overline{Y}_{(1)}^k \leq \overline{Y}_{(2)}^k \leq \ldots \leq \overline{Y}_{(M)}^k, k = 1, \ldots, K.$$

For the safety measures in the KCB study a smaller value of $\overline{Y}_i^k$ indicates better safety performance, so the results imply that segment (1) is the best segment and segment (M) is the worst. Unfortunately, the measure $\overline{Y}_i^k$ is potentially flawed, because it does not account for the effects of multiple segment firms in the data. Consider the following scenario. Suppose a particular firm operates in two segments: A and B. Suppose further that the firm performs well in segment A but poorly in segment B. Since the safety data are collected at the firm level (and not at the segment level within firms), an aggregate measure of the firm's performance is observed, and this single value is imputed to both segments A and B. The implication is the firm's performance in segment A will be understated (the performance will be worse than it actually is). Similarly, its performance will be overstated in segment B. Therefore, the measures $\overline{Y}_i^k$ may be biased. This is not to say that the KCB measures are useless, particularly if a firm's true safety performance is not segment specific, but more of a firm-level phenomenon. Unfortunately this is not testable within the available data, however future studies may want to take this fact into consideration.

2.2 KCB Results

Results of the KCB study are reproduced in Table 1a and 1b for the For-Hire sector and the Private sectors, respectively. Consider Table 1a. (Table 1b, the Private sector, is read is similar fashion to Table 1a and will not be discussed. Tables 1c and 1d contain the sample sizes N_i^k, used in the KCB study for the For-Hire and Private sectors, respectively. These are provided for informational purposes and are self-explanatory.) The first column of Table 1a contains the nine safety measures as defined in the Data Appendix. The other columns in the table are the motor carrier segments and are populated by the KCB measure $\overline{Y}_i^k$ and the relative ranks of each sector (the parenthetical values). For example, based on the Driver SEA measure, the best segment in the sample was the Passenger segment (Driver SEA equals 21.36 and rank equals 1) and the worst segment was the Refrigerated segment (Driver SEA equals 51.30 and rank equals 11). The second best segment for Driver SEA was the Tank segment (Driver SEA equals 36.49 and rank equals 2); the third best was the "General Freight Less Than Truckload" segment (Driver SEA equals 36.52 and rank equals 3). Therefore the three best segments in terms of Driver SEA had estimates of 21.36, 36.49 and 36.52. Based on these point estimates alone, it is unclear whether these three values are significantly different (in a statistical sense), so there is no way of determining if the Passenger segment is truly the best in the population or if, perhaps, the "General Freight Less Than Truckload" segment is not third best but is actually tied with the Tank segment for second best in the population. These are potentially interesting inferential statements that can only be made with rigorous statistical techniques like Ranking and Selection which will be introduced in section 3. However, we now address the shortcomings of the KCB measure by introducing an alternative model that accounts for the multiple segment nature of the firms.

2.3 An Alternative Safety Measure Model

Regardless of the adequacy or inadequacy of the KCB measure, it is still a potentially useful statistic. However, the data set did contain quite a number of multiple-segment firms. See Table 2, which gives counts of the firms in the data set that operated in multiple segments. In the For-Hire sector there were about 94,000 firms of which 60,125 (64%) operated in a single segment. In the Private sector 42,478 of the approximately 55,000 firms operated in a single segment (about 78%). The rest operated in multiple segments. For example in the For-Hire sector, 16,251 firms (17%) operated in two segments while 8,930 (9%) operated in three. These numbers imply that multiple-segment firms may be a significant percentage of the data set and need to be addressed in the assessment of safety across segments.

Therefore, consider the general linear model:

$$Y_j^k = \alpha + \sum_{i=1}^{M} \beta_i^k d_{ij}^k + \varepsilon_j ,$$

where ε_j is a random variable with zero mean and constant variance, uncorrelated with the d_{ij}^k, and the α and β_i^k are parameters for estimation. Then Ordinary Least Squares (OLS) estimation of this model produces estimates $\hat{\alpha}$ and $\hat{\beta}_i^k$. The estimate $\hat{\beta}_i^k$ is the marginal effect on safety measure *k* of a firm being in segment *i*. Based on these estimates, a segment specific safety measure that captures the effects of a single firm being in multiple segments is $\hat{Y}_i^k = \hat{\alpha} + \hat{\beta}_i^k$. These estimates imply the alternative safety rank statistic:

$$\hat{Y}_{(1)}^k \le \hat{Y}_{(2)}^k \le \ldots \le \hat{Y}_{(M)}^k, k = 1, \ldots, K.$$

In general, the $\hat{Y}_i^k$ may be different in magnitude than the $\overline{Y}_i^k$, as will the associated rank statistic. (However, as will be seen later, at the extreme ends on the rank statistic the associated rankings tended to be the same for the $\hat{Y}_i^k$ and the $\overline{Y}_i^k$.) The $\hat{Y}_i^k$ specifically control for effects of multiple-segment firms in the data. Additionally, the multiple segment nature of the data implies that safety measure estimates may themselves be correlated across segments. That is, any estimates of safety measure *k* should be correlated across segments through the multiple segment firms in the data, so that the covariance of $\hat{Y}_s^k$ and $\hat{Y}_t^k$, Cov($\hat{Y}_s^k$, $\hat{Y}_t^k$), should be non-zero for $t \neq s$, and these covariances may have profound effects on any inferences that uncover statistically significant differences across segments. To see this, one only has to realize that Var($\hat{Y}_s^k$ - $\hat{Y}_t^k$) = Var($\hat{Y}_s^k$) + Var($\hat{Y}_t^k$) – 2Cov($\hat{Y}_s^k$, $\hat{Y}_t^k$). These types of covariances are readily estimated for the alternative estimates $\hat{Y}_i^k$, but there is no sense in which these covariances can be estimated for the KCB estimates, $\overline{Y}_i^k$. This is a limitation of the SafeStat data set that should be revisited to improve any future statistical analyses such as this.

3.0 Ranking and Selection Procedures

To motivate the discussion of ranking and selection consider the following hypothetical example. For simplicity of exposition, suppose there are only four motor carrier commodity segments under consideration: A, B, C and D. Suppose further that for each particular safety measure (A, B, C, D) the estimated safety scores within these segments are calculated as: 10.5, 12.0, 11.5 and 9.0, respectively. If a lower safety score is considered better, then the ranked safety scores are 9.0, 10.5, 11.5 and 12.0 and the ranked commodity segments are: D, A, C and B, respectively. From this simple analysis D would be considered the safest segment and B the least safe segment. However, the safety scores are statistics and, as such, are random variables. Indeed, if new safety data were collected for these four segments and new averages calculated, the numbers might change and, hence, the rankings might change. These changes across different samples of data can be thought of as *sampling variability* or *estimation error*. Without accounting for this variability or error, it may be incorrect to conclude that segment "D is best" and segment "B is worst". In fact, it may be the case that (in truth) "D and A are equally best" or "C is best".

Ranking and Selection procedures are designed to preclude these types of inferential mistakes by accounting for the sampling variability implicit in the statistical exercise. Specifically, these procedures control for inferential error rates and yield probability statements concerning the relative magnitude of the estimates in the rank statistics. To continue the example, these procedures yield statements such as: "segment D is the safest with 90% probability" or "segments D and A are safest with 95% probability" or "segment B is the least safe with 95% probability". Clearly these are much more powerful inferential statements than the rankings reported in the KCB study. The purpose of the present study is to make these types of probabilistic statements for each safety measure in each sector analyzed in the KCB study.

3.1 Ranking and Selection Theory

Ranking and Selection Theory is a subset of a larger body of statistical inference procedures called "Multiple Comparison Procedures". An excellent reference on the subject of

Multiple Comparison Procedures (and to a lesser extent on Ranking and Selection Theory) is Hsu (1996). A related body of literature is the "Multiple Comparisons with the Best" literature. For an excellent explanation of these procedures, which are closely related to Ranking and Selection Theory, see Horrace and Schmidt (2000). The Technical Appendix provides some background information on what follows, however the interested reader is referred to this literature for a complete understanding of the theory and proofs.

Let λ be a value on the unit interval (0,1), then λ can be thought of as an inferential error rate and (1-λ)*100% a confidence level. For example, for this study we select λ = 0.05, so that our inferential confidence level is 95%. Define the set of all segment index numbers $\{1,\ldots, M\}_k$ for safety measure *k*. Ranking and Selection Theory defines two subsets from the set $\{1,\ldots, M\}_k$, denoted S_+^k and S_-^k, each containing some subset of segment index numbers from $\{1,\ldots, M\}_k$. Call S_+^k the "subset of the **best**", and S_-^k the "subset of the **worst**". Then,

> *with probability at least 1-λ, the set S_+^k contains the* ***best*** *motor-carrier segments for safety measure k,*
>
> *and with probability at least 1-λ, the set S_-^k contains the* ***worst*** *the motor-carrier segments for safety measure k.*

Therefore, motor carrier segments that fall into the subset of the best can be deemed "best", and those in the subset of the worst can be deemed "worst" at the (1-λ)*100% confidence level for each safety measure.

3.2 Discussion

Continuing the example from section 3.0, if we let the segment designations A, B, C and D represent segment indices, then the set $\{1,\ldots, M\}_k$ = {A,B,C,D}. Ranking and selection procedures may indicate that the 95% subset of the best is S_+^k = {D, A}, and the 95% subset of the worst is S_-^k = {B}. That is, with 95% confidence segment A or D is the best. The statistical analysis has not revealed a single best segment, but of all the segments analyzed, segments A and D are better than B and C with 95% probability. Additionally, with 95% confidence segment B is the single worst performing segment.

The aforementioned probability statements, concerning the subsets of the best and worst, are extremely powerful. They allow us to better understand the significance of the segment-by-segment rankings for each safety measure. It should be noted that these subsets will generally contain more than one segment (but not always). It should also be noted that as λ decreases, 1-λ increases and so does the probability associated with each statement. For this study we select the standard λ = 0.05 implying that the probability statements are made at the 95% confidence level, but for smaller λ the confidence level increases and we should expect the number of motor-carrier segment in the subsets of the best and worst to grow. At larger values of λ, the confidence level decreases and the inference procedure can produce a single segment in the each of the subsets. That is, with lower confidence levels we could single out one segment as the best and one segment as the worst, but this was not a goal of the study, so λ was fixed at 0.05, implying a fixed 95% confidence level. Confidence level experiments could be conducted in the future to single out one best and one worst segment for each safety measure, *k*.

4.0 Results

The results for the best and worst subsets for both the KCB measure ($\overline{Y}_i^k$) and the OLS measure ($\hat{Y}_i^k$) for both the For-Hire and Private sectors are contained in Tables 3, 4, 5 and 6. Tables 3 and 4 contain the KCB measures with KCB For-Hire in Table 3 and KCB Private in Table 4. Tables 5 and 6 contain the OLS measure with OLS For-Hire in Table 5 and OLS Private in Table 6. All the tables have the same format. The columns contain the segments (*i*) and the rows contain the safety measures (*k*). The cells are populated for each safety measure with: "B" if the segment was contained in the subset of the best with 95% confidence, "W" if the segment was contained in the subset of the worst with 95% confidence, and a "B,W" if it was contained in both. Blanks imply that the segment was neither "best" nor "worst" for that measure.

4.1 KCB Results

The KCB results are contained in Tables 3 and 4. In Table 3 (For-Hire), for the Driver SEA measure (first row) the single best segment was the Passenger segment and the single worst was the Refrigerated segment. This is a strong inference statement because it identifies single segments as the safety extrema at the 95% confidence level. This is also the case for the Driver OOS Rate and the Vehicle SEA measures. In Table 3 the only occurrence of a "B,W" classification is for the Enforcement Severity Measure (ESM) in the "General Freight Less Than Truckload" segment. This occurred because the measure was only based on 19 observations, so the precision of the inference for that measure is suspect; there is not enough data for the ESM measure to be statistically meaningful. In Table 3 the results are generally good because the sample sizes tended to be large. (See Tables 1c and 1d for the sample sizes used in this analysis.) Overall, the Passenger and the "General Freight Less Than Truckload" segments performed the best, being in the Best subset 6 times each. The Worst segments were the Intermodal and Produce segments, being in the worst subset 4 times each. The least reliable measure for ranking the segments was the Fatal Crash Rate and the ESM, which tended to have multiple segments in the best and worst subsets.

Turning to the Private sector results of Table 4, the inference is less sharp, due to generally smaller sample sizes than for the For-Hire analysis. (Again, the reader is referred to Tables 1c and 1d for sample sizes.) For example, in the Driver SEA category, the best segment was the Tank segment, but the worst subset consisted of five segments (Household, Intermodal, Large Machine, Passenger and Produce), but again this may be due to small sample sizes: Intermodal only had 166 observations. Again the Fatal Crash Rate and the ESM tended to be the least reliable safety measure in terms of discerning a best and worst segment. However, the Safety Management Review Measure (SMRM) was also fairly unreliable. All three measures had a multiplicity of segments that were both in the best and in the worst subsets, and often suffered from small sample sizes within certain segments. For example, in the ESM category the Passenger, Intermodal and Household segments only had 7, 12 and 29 observations, respectively. The sharpest overall inference was for the Driver OOS Rate, Vehicle SEA, Vehicle OOS rate and the Accident SEA. The only safety measures that produced a single best or worst segment were the Driver SEA, which determined that the Tank segment was the single best, the Accident SEA, which determined that the Refrigerated segment was the single worst, and the Total Crash Rate, which determined that the Passenger segment was the single best. The Household Goods segment was best 6 times but worst 5 times. The Passenger segment was best 4 times but worst 6 times. The Intermodal segment was best 5 times, but worst 5 times.

4.2. OLS Results

Since the present study focuses on inference of the ranking statistics, the OLS estimation results are not reported. For comparative purposes to the KCB results, only the ranking and selection results for the OLS estimates are contain in Tables 5 and 6. The results of Table 5, are provided for comparative purposes with the KCB results of Table 3; the results of Table 6 are provided for comparative purposes with the KCB results of Table 4. Consider Tables 3 and 5. In general, the segments that were best and worst in Table 3 were also best and worst in Table 5.

Hence, the OLS results tend to confirm the results of the KCB measure with the added feature of controlling for multiple segment firms. However, the inference of the OLS results of Table 5 tended to be less sharp, insofar as they added many segments to the subsets of the best and the worst for each measure. Overall, Table 5 adds 17 segments to the subsets of the best and only 6 segments to the subsets of the worst. Insofar as interest centers on identifying and potentially improving the worst performing sectors, the addition of only 6 segments to the subsets of the worst implies that the KCB estimator is effective at identifying the worst segments.

In only two cases were segments dropped from subsets in moving from Table 3 to Table 5. Passenger, Accident SEA was best in Table 3 but dropped from that subset in Table 5. Refrigerated, Driver SEA was worst in Table 3 but dropped from that subset in Table 5. In only one case (Refrigerated, Fatal Crash Rate) did a segment switch subsets: from worst to best. Therefore, it is safe to conclude that the KCB measure is effective for identifying the worst segments but less so for identifying the best segments. These general conclusions of the For-Hire KCB and the OLS measure comparisons are confirmed in comparing the Private sector results of Tables 4 and 6.

4.3 Overall Findings

Considering Tables 3, 4, 5 and 6 simultaneously, it is clear that the OLS generally confirms the KCB results. To summarize, the KCB results do not control for the multiple segment firms in the data set, while the OLS results do. The bulk of the firms do operate in single segments, so it may be safe to conclude that the KCB results are not as problematic as one might have thought. (It might be interesting to redo the KCB study using only single segment firms and compare the results to the OLS analysis; this should be a part of the future research agenda for this project) There are some clear consistencies across the KCB and OLS results. Indeed, if a segment falls into the subset of the best or the worst across the two analyses it is safe to conclude that for this sample those firms are best or worst (respectively), regardless of which estimates (KCB or OLS) are considered. Therefore, Tables 7 and 8 report inferential findings that were consistent across studies for the For-Hire and Private sectors, respectively. These tables are provided for informational purposes; they tend to confirm the KCB results.

In general, the best safety measures for comparative purposes are the three SEAs and the two OOS rates; the two crash rates, the SMRM and the ESM are less effective for comparative purposes (although the Total Crash Rate and the SMRM were somewhat effective for the KCB For-Hire analysis in Table 3). The For-Hire results are more compelling than the Private sector results because the Private sector uses fewer observations in estimating safety performance in all the segments for all the safety measures. Perhaps future studies should try to increase the sample sizes in the Private sector analysis.

Finally, the “General Freight Less Than Truckload” segment seems to present some additional challenges. Notice that the accuracy of the estimates obtained is suspect, because the segment has many B’s, W’s and B,W’s in it. Consider Table 5. Here the “General Freight Less Than Truckload” segment was best in Driver OOS Rate but worst in Driver SEA. Perhaps these results seem contradictory, but, perhaps, they highlight the fact that the data collection procedure is not uncovering the true nature of safety in this segment. Future safety studies should be designed to sharpen the inference results in this segment.

5.0 Conclusions and Future Research

This inferential study is an excellent start to a larger analysis of the KCB estimates obtained from the SafeStat data. It has highlighted the limitations and the strengths of the KCB measures in various motor-carrier segments, and, perhaps more importantly, it has highlighted the strengths and weaknesses of the SafeStat data itself. Without rehashing the specific results of section 4, some strong conclusions have surfaced. First and foremost, KCB rank statistics by themselves need to be interpreted with caution. This is not to say that the segment-by-segment ranks are wrong, but it is to say that inferential procedures are necessary to get a true sense of which segments are best and worst. These inferential results are a function of a pre-selected confidence level; for this study the “standard” 95% was chosen. However, one could envision extending the results to other confidence levels (e.g. 99% or 90%) to see how sensitive these results are to this choice.

Second, concerns over the effects of multiple-segment firms on the validity of the KCB results have, in part, been mitigated, because the OLS results confirm the KCB results. However, it is extremely important to re-perform the KCB study for single-segment firms only. This is a relatively simple way to confirm or not confirm the KCB results for all firms in the data. Alternatively, research should be performed to develop ways to quantify and control for “multi-segment firm effects” in the data themselves. As collected the data do not admit this quantification for the KCB results, so perhaps the data collection process could be altered to make this feasible.

Third, this research implies a comprehensive firm-level survey study of the best and worst segment performers to uncover why they do so well or so poorly in each safety category.

This has been started with the design of a “Best Practices Survey” which surveys the individual performance of 250 carriers. Fourth, it would be useful to more closely examine the crash data results (Fatal and Total Crash Rates), since crashes are clearly important safety measures. These data tend to have relatively smaller sample sizes than other measures. While increasing the sample sizes of crash data is clearly **not** feasible, perhaps the information provided in the SafeStat data set for these measures should be expanded as greater informational content may well improve the accuracy and sharpen the inference for these measures. One obvious improvement would be to provide information on what the trucks were carrying when the crashes occurred; this would tend to disaggregate the data, mitigate the multiple-segment carrier problem and improve the accuracy of the KCB results.

Finally, the results of this research may have strong Federal policy implications. Perhaps in the future, federal regulations could be tailored to improve the performance of the worst segments based on the results of subsequent studies. As new and improved data become available and are analyzed, the results of future research will yield improved estimates and inferences suitable for policy decision-making.

References

Horrace, W.C. and P. Schmidt (2000). Multiple Comparisons With The Best, With Economic Applications. *Journal of Applied Econometrics*, 15, 1-26.

Hsu, J.C. (1996). Multiple Comparisons, Theory and Methods. Chapman & Hall, New York.

Keane, T., T. Corsi and K. Braaten (2001). Motor Carrier Industry Profile Study: Evaluating Safety Performance by Motor Carrier Industry Segment. Unpublished Manuscript, U.S. Department of Transportation. Available under Current Analysis Results on the Analysis and Information Online website at: ai.volpe.dot.gov.

TABLES AND APPENDICES

Table 1a. For-Hire Trucking Segment-by-Segment Results: $\overline{Y}_i^k$ (KCB) and Relative Ranks

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Less Than Truck Load	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related											
Driver SEA	40.41 (7)	37.08 (4)	51.03 (11)	36.52 (3)	43.67 (8)	44.89 (9)	39.04 (5)	39.06 (6)	21.36 (1)	48.57 (10)	36.49 (2)
Driver OOS Rate	9.36 (6)	8.86 (5)	12.15 (10)	4.76 (1)	11.18 (8)	13.78 (11)	8.08 (4)	9.64 (7)	6.93 (2)	12.06 (9)	7.32 (3)
Vehicle-Related											
Vehicle SEA	50.23 (9)	47.30 (6)	46.02 (5)	41.33 (2)	47.53 (8)	44.07 (3)	57.23 (11)	51.33 (10)	30.87 (1)	47.49 (7)	44.51 (4)
Vehicle OOS Rate	27.04 (9)	25.53 (8)	23.82 (5)	20.12 (2)	25.25 (6)	23.28 (3)	28.61 (10)	28.95 (11)	18.50 (1)	25.35 (7)	23.33 (4)
Crash-Related											
Accident SEA	11.67 (7)	9.96 (3)	12.85 (8)	30.80 (11)	11.23 (6)	7.86 (2)	14.28 (9)	9.96 (3)	7.67 (1)	10.86 (5)	15.82 (10)
FAT_CR_D (Fatal Crash Rate)	0.015 (7)	0.022 (11)	0.017 (9)	0.003 (1)	0.015 (7)	0.013 (4)	0.014 (5)	0.014 (5)	0.007 (2)	0.020 (10)	0.009 (3)
TOT_CR_D (Total Crash Rate)	0.342 (7)	0.411 (10)	0.363 (9)	0.091 (1)	0.352 (8)	0.302 (5)	0.241 (4)	0.329 (6)	0.219 (3)	0.418 (11)	0.205 (2)
Others											
Safety Management Review Measure	20.45 (4)	22.88 (9)	21.26 (6)	9.44 (1)	22.17 (8)	29.32 (11)	20.85 (5)	21.53 (7)	17.09 (3)	23.31 (10)	12.66 (2)
Enforcement Severity Measure	2.50 (5)	2.29 (2)	3.12 (10)	2.50 (5)	2.60 (8)	2.51 (4)	2.62 (9)	2.43 (3)	1.57 (1)	3.14 (11)	2.51 (7)

Numbers in parentheses are the relative ranks of the segments for each safety measure.
Definitions of safety measures are contained in the Data Appendix.

Table 1b. Private Trucking Segment-by-Segment Results: $\overline{Y}_i^k$ (KCB) and Relative Ranks

Safety Measure	Building Materials	Bulk Freight	Refrig- erated	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related										
Driver SEA	25.89 (3)	25.63 (2)	28.53 (4)	29.40 (6)	30.05 (7)	28.69 (5)	31.17 (8)	34.60 (10)	31.58 (9)	20.70 (1)
Driver OOS Rate	9.98 (5)	8.33 (2)	9.11 (4)	12.24 (9)	13.84 (10)	8.75 (3)	11.61 (7)	11.75 (8)	10.54 (6)	6.43 (1)
Vehicle-Related										
Vehicle SEA	47.32 (7)	46.53 (6)	31.58 (2)	38.16 (4)	28.67 (1)	50.50 (8)	56.03 (10)	51.05 (9)	37.16 (3)	42.72 (5)
Vehicle OOS Rate	27.90 (6)	28.56 (8)	16.41 (2)	21.45 (4)	15.18 (1)	28.25 (7)	33.66 (10)	28.81 (9)	19.96 (3)	24.17 (5)
Crash-Related										
Accident SEA	4.81 (5)	5.32 (7)	7.28 (10)	3.81 (3)	2.74 (1)	4.82 (6)	4.13 (4)	2.74 (1)	5.83 (9)	5.57 (8)
FAT_CR_D (Fatal Crash Rate)	0.013 (7)	0.016 (9)	0.011 (5)	0.018 (10)	0.003 (2)	0.011 (5)	0.014 (8)	0.000* (1)	0.010 (3)	0.010 (3)
TOT_CR_D (Total Crash Rate)	0.277 (5)	0.359 (9)	0.224 (4)	0.372 (10)	0.310 (8)	0.199 (2)	0.286 (6)	0.080* (1)	0.286 (6)	0.207 (3)
Others										
Safety Management Review Measure	31.53 (4)	32.73 (5)	35.05 (9)	33.89 (8)	28.22 (3)	39.34 (10)	33.84 (7)	26.67 (2)	33.43 (6)	20.52 (1)
Enforcement Severity Measure	1.48 (7)	1.30 (4)	2.40 (10)	.98 (2)	1.34 (5)	1.08 (3)	1.45 (6)	.14 (1)	2.19 (9)	1.50 (8)

Numbers in parentheses are the relative ranks of the segments for each safety measure.
Definitions of safety measures are contained in the Data Appendix.
*Private Passenger Carrier Segment has only 13 valid observations. One carrier's data (Coach Leasing, Inc.) was removed because researchers felt it had been erroneously entered (e.g., 23 total crashes for 1 power unit within the last 30 months).

Table 1c. For-Hire: Sample Sizes Used in Analysis for Each Segment/Safety Measure Combination, N_i^k

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Less Than Truck Load	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related											
Driver SEA	11,236	15,767	8,983	91	47,329	3,296	2,686	8,866	1,589	11,865	4,377
Driver OOS Rate	15,962	23,455	11,688	89	66,961	5,107	3,477	13,212	2,826	16,100	5,575
Vehicle-Related											
Vehicle SEA	10,243	13,651	8,127	91	42,494	3,015	2,513	8,090	1,612	10,605	4,159
Vehicle OOS Rate	15,962	23,455	11,688	89	66,961	5,107	3,477	13,212	2,826	16,100	5,575
Crash-Related											
Accident SEA	15,572	22,989	11,391	91	50,719	4,627	3,399	12,884	2,406	15,661	5,453
FAT_CR_D (Fatal Crash Rate)	4,364	5,190	3,442	85	13,751	1,005	1,276	3,211	589	3,899	2,367
TOT_CR_D (Total Crash Rate)	4,364	5,190	3,442	85	13,751	1,005	1,276	3,211	589	3,899	2,367
Others											
Safety Management Review Measure	1,783	1,816	1,591	19	5,390	399	510	1,251	444	1,861	893
Enforcement Severity Measure	1,783	1,816	1,591	19	5,390	399	510	1,251	444	1,861	893

Larger sample sizes typically imply sharper inference.

Table 1d. Private Sector: Sample Sizes Used in Analysis for Each Segment/Safety Measure Combination, N_i^k

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related										
Driver SEA	7,524	4,867	2,314	4,912	1,125	166	8,402	61	1,811	5,866
Driver OOS Rate	14,315	10,061	3,564	10,345	2,666	261	17,187	114	3,256	9,730
Vehicle-Related										
Vehicle SEA	7,067	4,259	2,176	4,492	1,026	156	7,722	58	1,663	5,645
Vehicle OOS Rate	14,315	10,061	3,564	10,345	2,666	261	17,187	114	3,256	9,730
Crash-Related										
Accident SEA	14,060	9,866	3,487	9,815	2,571	257	16,817	105	3,142	9,510
FAT_CR_D (Fatal Crash Rate)	1,923	1,378	820	1,004	204	46	2,018	13*	539	1,715
TOT_CR_D (Total Crash Rate)	1,923	1,378	820	1,004	204	46	2,018	13*	539	1,715
Others										
Safety Management Review Measure	542	309	171	248	29	12	782	7	107	530
Enforcement Severity Measure	542	309	171	248	29	12	782	7	107	530

Larger sample sizes typically imply sharper inference.

*Private Passenger Carrier Segment has only 13 valid observations. One carrier's data (Coach Leasing, Inc.) was removed because researchers felt it had been erroneously entered (e.g., 23 total crashes for 1 power unit within the last 30 months).

Table 2. Multiple Segment Firms

FOR-HIRE		PRIVATE	
Number of segments within firm	Number of firms within data set	Number of segments within firm	Number of firms within data set
1	60,125	1	42,478
2	16,251	2	8,829
3	8,930	3	2,368
4	5,021	4	658
5	2,151	5	183
6	1,203	6	63
7	615	7	32
8	237	8	8
9	81	9	4
10	14	10	1
11	0	11	0
Firm Totals	94,628		54,624

Table 3. For-Hire Trucking Subset of the Best and Worst, Based on $\overline{Y}_i^k$ (KCB) and Numerical 95% Critical Values

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Less Than Truck Load	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related											
Driver SEA			W						B		
Driver OOS Rate				B		W					
Vehicle-Related											
Vehicle SEA							W		B		
Vehicle OOS Rate				B			W	W	B		
Crash-Related											
Accident SEA				W		B			B		
FAT_CR_D (Fatal Crash Rate)		W	W	B		W	W		B	W	
TOT_CR_D (Total Crash Rate)		W		B						W	
Others											
Safety Management Review Measure				B		W				W	B
Enforcement Severity Measure			W	B, W			W		B	W	

B = Segment is in the subset of best with probability at least 95%.
W = Segment is in the subset of the worst with probability at least 95%.
For Enforcement Severity Measure, "General Freight Less Than Truckload" only had 19 observations, so inference is less sharp.

Table 4. Private Trucking Subsets of the Best and Worst, Based on $\overline{Y}_i^k$ (KCB) and Numerical 95% Critical Values

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related										
Driver SEA					W	W	W	W	W	B
Driver OOS Rate				W	W	B		W		B
Vehicle-Related										
Vehicle SEA			B		B	W	W	W		
Vehicle OOS Rate			B		B		W	W		
Crash-Related										
Accident SEA			W		B	B		B		
FAT_CR_D (Fatal Crash Rate)	W	W	W	W	B	B, W	W	**	B	W
TOT_CR_D (Total Crash Rate)		W		W	W			B		
Others										
Safety Management Review Measure	W	W	W	W	B, W	B, W	W	B, W	W	B
Enforcement Severity Measure	W		W		B, W	B, W		B, W	W	W

B = Segment is in the subset of best with probability at least 95%.
W = Segment is in the subset of the worst with probability at least 95%.
In all cases where a segment was designated “B, W”, the segment suffered from a small sample size, making inference less sharp.
** Inference could not be performed due to a complete lack of variability in the data (all observations the same).

Table 5. For-Hire Trucking Subsets of the Best and Worst, Based on $\hat{Y}_i^k$ (OLS) and 95% Simulated Critical Values

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Less Than Truck Load	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related											
Driver SEA				W	W				B		
Driver OOS Rate				B		W	B		B		B
Vehicle-Related											
Vehicle SEA				W			W		B		
Vehicle OOS Rate				B			W	W	B		
Crash-Related											
Accident SEA				W		B					
FAT_CR_D (Fatal Crash Rate)	B	W	B	B	B, W	B, W	B, W	B	B	W	B
TOT_CR_D (Total Crash Rate)		W		B					B	W	B
Others											
Safety Management Review Measure				B		W			B	W	B
Enforcement Severity Measure		B	W	B, W	W	B, W	B, W	B	B	W	B

B = Segment is in the subset of best with probability at least 95%.
W = Segment is in the subset of the worst with probability at least 95%.

Table 6. Private Trucking Subsets of the Best and Worst, Based on $\hat{Y}_i^k$ (OLS) and 95% Simulated Critical Values

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related										
Driver SEA		B	W		B	W		B, W	W	B
Driver OOS Rate				W	W	B		B, W		B
Vehicle-Related										
Vehicle SEA					B	W	W	W		W
Vehicle OOS Rate			B		B		W	W		
Crash-Related										
Accident SEA			W	B	B	B		B		
FAT_CR_D (Fatal Crash Rate)	B, W	B, W	B, W	B, W	B, W	B, W	B, W	B, W	B, W	B, W
TOT_CR_D (Total Crash Rate)		W	B	W	B, W	B, W		B, W	W	B
Others										
Safety Management Review Measure	W	W	W	W	B, W	B, W	W	B, W	W	B
Enforcement Severity Measure			W	B	B, W	B, W		B, W	B,W	

B = Segment is in the subset of best with probability at least 95%.
W = Segment is in the subset of the worst with probability at least 95%.

Table 7. For-Hire Trucking Subset of the Best and Worst That Were Common to Both KCB and OLS Measures

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Less Than Truck Load	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related											
Driver SEA									B		
Driver OOS Rate				B		W					
Vehicle-Related											
Vehicle SEA							W		B		
Vehicle OOS Rate				B			W	W	B		
Crash-Related											
Accident SEA				W		B					
FAT_CR_D (Fatal Crash Rate)		W		B		W	W		B	W	
TOT_CR_D (Total Crash Rate)		W		B						W	
Others											
Safety Management Review Measure				B		W				W	B
Enforcement Severity Measure			W	B, W			W		B	W	

B = Segment is in the subset of best with probability at least 95%.
W = Segment is in the subset of the worst with probability at least 95%.

Table 8. Private Trucking Subsets of the Best and Worst That Were Common to Both KCB and OLS Measures

Safety Measure	Building Materials	Bulk Freight	Refrig-erated	General Freight Truck Load	Household Goods	Inter Modal	Large Machine	Passenger	Produce	Tank
Driver-Related										
Driver SEA						W		W	W	B
Driver OOS Rate				W	W	B		W		B
Vehicle-Related										
Vehicle SEA					B	W	W	W		
Vehicle OOS Rate			B		B		W	W		
Crash-Related										
Accident SEA			W		B	B		B		
FAT_CR_D (Fatal Crash Rate)	W	W	W	W	B	B, W	W		B	W
TOT_CR_D (Total Crash Rate)		W		W	W			B		
Others										
Safety Management Review Measure	W	W	W	W	B, W	B, W	W	B, W	W	B
Enforcement Severity Measure			W		B, W	B, W		B, W	W	

B = Segment is in the subset of best with probability at least 95%.
W = Segment is in the subset of the worst with probability at least 95%.
In all cases where a segment was designated "B, W", the segment suffered from a small sample size, making inference less sharp.

Data Appendix

Data Description

The Motor Carrier Management Information System (MCMIS) was the primary source of data used in the KCB study and in this study. The MCMIS is used by FMCSA to maintain a comprehensive safety record of For-Hire and Private property and passenger carriers subject to the Federal Motor Carrier Safety Regulations (FMCSR). The MCMIS Census File contains records from over 500,000 entries (e.g., motor carriers, hazardous materials shippers, and registrants), and contains information on each company's identity (name, address), operations classification (type of business), cargo classification (type of cargo carried), and numbers of trucks and drivers within the company. A motor carrier's identifying information is originally collected when the carrier registers with FMCSA when preparing to operate in interstate commerce.

The MCMIS data set also contains an Inspection File, which contains the results of roadside inspections (submitted by States), enforcement actions (taken by Federal personnel against a motor carrier), and compliance reviews (conducted by FMCSA and State safety investigators.) Enforcement actions may include civil penalties or out-of-service (OOS) orders placed against a carrier. Compliance reviews are on-site reviews of a motor carrier's operations, conducted by FMCSA and State personnel to determine the level of compliance with the FMCSRs.

The MCMIS data set also includes a Crash File, which contains data from State police crash reports electronically transmitted to FMCSA. The census, inspection, and crash data are reviewed and updated as new information is collected by FMCSA on a motor carrier, whether through inspections, compliance reviews, enforcement action, or reportable crashes.

This study also uses data produced by the Motor Carrier Safety Status Measurement System, or "SafeStat", a powerful analytical tool developed by FMCSA and the Volpe National Transportation Systems Center to accurately identify and monitor high-risk motor carriers within the overall motor carrier population. The SafeStat became operational in 1995 and uses MCMIS data as input to evaluating a motor carrier's relative safety fitness. The SafeStat system incorporates current on-road safety performance data for each carrier with on-site compliance review data collected by FMCSA and State safety investigators. All the data are run through an algorithm, with the result being a comprehensive evaluation of a motor carrier's safety performance using four types of data: accident, driver, vehicle, and safety management. Safety event data are normalized to account for a carrier's size or amount of exposure using carrier-descriptive data such as number of power units or the number of roadside inspections. Data are also time-weighted, with the most recent events having greater weight than older ones. Crash data are also severity-weighted, based on the number and type of injuries sustained by victims involved in the crashes.

All the SafeStat data serve to measure a carrier's relative safety fitness and assess its risk of having future crashes. It should be noted that not all motor carriers contained in the MCMIS data set are analyzed by SafeStat. To be assessed by SafeStat, a motor carrier has to have experienced at least three inspections within the past 30 months. As such, the number of firms with SafeStat data (e.g., those analyzed in this study) is significantly smaller than the total population of carriers contained in MCMIS.

Safety Performance Measures

The nine specific safety performance measures examined in this analysis include two driver-, two vehicle-, three crash-, and two safety management-related measures.

Driver Safety Evaluation Area (SEA) is a SafeStat composite value calculated from Driver Inspection Indicator (DII), Driver Review Indicator (DRI), and Moving Violation Indicator (MVI). The DII is based on driver roadside out-of-service (OOS) inspection violations, the DRI is based on violations of driver-related acute and critical regulations discovered during a compliance review, and the MVI is based on serious moving violations recorded in conjunction with roadside inspections. Each inspection is weighted by its age and the number of driver OOS violations found, and then normalized by the number of driver inspections within the last 30 months. A lower Driver SEA value indicates better safety performance.

Total Driver OOS Rate is derived from MCMIS data. The total number of driver OOS violations, divided by the total number of driver inspections experienced by the motor carrier. A lower Driver OOS rate indicates better safety performance.

Vehicle Safety Evaluation Area (SEA) is a SafeStat composite value calculated from the Vehicle Inspections Indicator (VII) and the Vehicle Review Indicator (VRI). The VII is based on the number of vehicle roadside OOS inspection violations and the VRI is based on violations of vehicle-related acute and critical regulations discovered during compliance reviews. Each inspection is weighted by its age and the number of vehicle OOS violations found, and then normalized by the number of vehicle inspections within the last 30 months. A lower Vehicle SEA value indicates better safety performance.

Total Vehicle OOS Rate is derived from MCMIS data. The total number of vehicle OOS violations divided by the total number of vehicle inspections experienced by a motor carrier. A lower Vehicle OOS rate indicates better safety performance.

Accident Safety Evaluation Area (SEA) is a SafeStat composite value calculated based on Accident Involvement Indicator (AII) and the Recordable Accident Indicator (RAI). The AII uses measures derived from state-reported crash data normalized by the number of power units owned/leased by the motor carrier from MCMIS. The RAI uses measures based on recordable crashes and annual vehicle miles traveled (VMT) data gathered at the most recent compliance review. A lower Accident SEA value indicates better safety performance.

Fatal Crash Rate is derived from MCMIS data. The number of fatal crashes experienced by a carrier divided by the number of power units owned or leased by that carrier. A lower fatal crash rate indicates better safety performance.

Total Crash Rate is derived from MCMIS data. The number of total recordable crashes experienced by a carrier divided by the number of power units owned or leased by that carrier. A lower total crash rate indicates better safety performance.

Safety Management Review Measure (SMRM) is a compliance review measure that uses the number of safety management-related acute and critical violations of FMCSA regulations discovered during a compliance review. A lower value indicates better safety performance.

Enforcement Severity Measure (ESM) is derived from MCMIS data. The number of past (closed) enforcement cases brought against a carrier by FMCSA. This is a general indicator of the commitment to safety by the motor carrier over time (within the last six years). A lower value indicates better safety performance.

These nine measures were selected from a larger list of 23 safety performance measures for which data were collected in this study. Many of the 23 measures are closely correlated, and several serve as direct inputs to others. In an effort to reduce the potential effect of this correlation on the results, researchers selected nine "core" measures in this analysis. These nine measures are fairly representative of the four safety areas of interest in this study: driver-, vehicle-, crash-, and safety management-related areas.

Technical Appendix

Ranking and Selection Results

Assume that the data for safety measure *k* are generated by an *M*-variate normal distribution with *M* finite means, μ_i, and unknown covariance structure. (Suppress the superscript *k* in what follows for notational parsimony.) The data can be standardized in the usual way (by demeaning and dividing by the square root of the variance) to produce an *M*-dimensional standard normal distribution. Let the covariance structure of standard normal distribution be given by the *M*-dimensional covariance matrix Ω, implying an *M*-dimensional correlation matrix, *R*. Let realizations of the *M*-dimensional standard normal distribution be given by vector $Z = [Z_1, \ldots, Z_M]$. Define *U* as a Chi-squared random variable with *v* degrees of freedom, independent of the random variable *Z*. Let $T_i = Z_i(U/v)^{1/2}$. Then $T = [T_1, \ldots, T_M]$ is an *M*-dimensional Student *t* distribution with correlation matrix *R* and *v* degrees of freedom. The joint density of *T* is:

$$f(t_1,\ldots,t_M, R, v) = \frac{\Gamma((M+v)/2)}{(v\pi)^{M/2}\Gamma(v/2)}(\det R)^{-1/2}(1+T'R^{-1}T/v)^{-(M+v)/2}$$

where Γ is the usual Gamma function. Define the critical value $T^{\lambda}_{M,v,R}$ as the solution in *t* of the equation:

$$P\{\max_i |T_i| \leq t\} = \int_{-t}^{t}\ldots\int_{-t}^{t} f(t_1,\ldots,t_M, R, v)\,dt_1\ldots dt_M = 1-\lambda\,.$$

When the Z_i are independent, Ω and *R* are identity matrices, and the critical values $T^{\lambda}_{M,v,R}$ are said to be "equicorrelated" (with correlations everywhere equal to zero) and can be numerically solved and tabulated for various values of *M*, λ and *v*. Then as $v \to \infty$, f becomes a normal distribution, then the equicorrelated critical values become normal critical values and can tabulated for various values of values of *M* and λ. Tabulations for both equicorrelated cases can be found in Hsu (1996). If the equicorrelated structure does not hold or if M is prohibitively large, the critical values cannot be calculated numerically, but they can be simulated. A simulation algorithm used in the study is provided below. For this study the degrees of freedom are generally large enough to assume normality for the critical values. Additionally, due to the Central Limit Theorem, for purposes of ranking and selection, the degrees of freedom are large enough so that the data need not be generated from an *M*-dimensional normal distribution; any distribution with finite moments will suffice for inference to be valid using a normality assumption for the critical values.

Let $\hat{Y}_i^k$ $i = 1, \ldots, M$ be any unbiased estimates of the unknown population means. Let $\hat{\Omega}^k$ be an $M \times M$ estimate of the covariance structure of the *M* populations for safety measure *k*. The individual elements of $\hat{\Omega}^k$ are $\hat{\omega}^k_{st}$, $s,t = 1, 2, \ldots, M$. Define the "Gupta subsets", S^k_+ and S^k_-, such that:

$$S^k_- = \{s : \hat{Y}^k_s - \hat{Y}^k_i + T^{\lambda}_{M-1,v,R}(\hat{\omega}_{ss} + \hat{\omega}_{ii} - 2\hat{\omega}_{si})^{1/2} \geq 0\} \text{ and}$$

$$S^k_+ = \{s : \hat{Y}^k_i - \hat{Y}^k_{si} + T^{\lambda}_{M-1,v,R}(\hat{\omega}_{ii} + \hat{\omega}_{ss} - 2\hat{\omega}_{si})^{1/2} \geq 0\}\,.$$

For the aforementioned theory, the proposed OLS measure $\hat{Y}_i^k$ was used because it admits simple (OLS) estimates of $\hat{\Omega}^k$ and hence $\hat{\omega}_{st}^k$. In this case, the inference is complicated, but tractable. However, using the proposed measure has the slight disadvantage of not allowing numerically calculated critical values, $T_{M,v,R}^{\lambda}$. As mentioned earlier, numerically calculated critical values are only forthcoming when the equicorrelated case holds. Since $\hat{Y}_i^k$ admits non-zero $\hat{\omega}_{st}^k$, it also admits a non-equicorrelated structure and, hence, requires simulated critical values for inference. Numerically calculated critical values are always preferred over simulated critical values. However, Horrace and Schmidt (2000) argue that for economic applications (such as the present study) simulated critical values are a reasonable alternative to numerical ones.

Alternatively we can also use the KCB measure, $\overline{Y}_i^k$, but this measure only admits an estimate $\hat{\omega}_{ii}^k = s_{ik}^2 / N_i^k$ (variance), but not $\hat{\omega}_{st}^k$ (covariance). Therefore, to use the KCB measure we must assume that the measures are not correlated ($\hat{\omega}_{st}^k = 0$) to make the inference tractable. This assumption may be dubious, given the fact that there are many multiple-segment firms in the data, but, again, this is less of a problem if it can be assumed that safety is at the firm-level and not at the segment-level within firms. One advantage of the KCB measure is that we can use numerically calculated critical values. In summary, this study performs inference on the safety measures two ways:

1. *using the KCB measure $\overline{Y}_i^k$ with **numerically** calculated critical values, $T_{M-1,v,R}^{.05}$, assuming an **equicorrelated** structure, and*
2. *using the proposed OLS measure $\hat{Y}_i^k$ with **simulated** critical values, $T_{M-1,v,R}^{.05}$, and a **non-equicorrelated** structure.*

Simulated Critical Value Algorithm

To generate the simulated critical values used in this study:

1. Perform a Choleski decomposition of $\hat{\Omega}$ into H, such that $H'H = \hat{\Omega}$.
2. Generate M independent standard normal variates: $Z_q = [Z_{1q}, \ldots, Z_{Mq}]$.
3. Generate an independent Chi-squared variate, U_q, with v degrees of freedom.
4. Calculate $T_q = H'Z_q(U_q/v)^{-1/2}$.
5. Find $T_q^* = \max | T_q |$, the maximum element of T_q.
6. Perform steps 2, 3, 4 and 5 for $q = 1, \ldots, Q$.
7. Calculate a (1-λ)*100 percentile from T_q^*, $q = 1, \ldots, Q$. This simulated value serves as a consistent estimates of $T_{M,v,R}^{\lambda}$.

As $Q \to \infty$, the simulated critical value converges in probability to the numerically calculated critical value. Critical values were simulated using the GAUSS programming language random number generator. For the simulation Q was set to 100,000.