A Question of Theft

William B. Fairley
Analysis and Inference, Inc.
Boston, Massachusetts

Jeffrey E. Glen
New York City Law Department
New York, New York

ABSTRACT

Several employees of Brink's, Inc., were convicted of theft from collections of the New York City's parking meters. The city sued Brink's for negligence in supervision of its employees and asked for compensation in the amount of the stolen coins. This chapter first discusses the legal background and criteria for proof in court that shaped the statistical question asked, and the precision required in its answer. It then lays out the statistical analysis presented on behalf of the city and a conflicting analysis presented on behalf of Brink's. The basis for the award given by a jury to the city in district court and its successful defense in the court of appeals is discussed.

1. Introduction

Brink's v. City of New York is a lawsuit in which statistical analysis was crucial to obtaining a verdict in a civil damages action involving frequently litigated questions of negligence and breach of contract (546 F.Supp. 403 (1982)). It is unusual for statistical analysis to be used by attorneys in "ordinary" cases like *Brink's;* the discipline of statistics has most frequently been called on in such specialized areas as antitrust law, where questions of market share have proved susceptible to statistical evaluation, and civil rights law, where proof of disparate racial or sexual impact often requires statistical analysis (Finkelstein, 1978; Baldus and Cole, 1980; Sugrue and Fairley, 1983; Fairley, 1983).

Lawyers rarely have enlisted statistical experts to prove damages in more familiar situations; our experience in *Brink's* indicates both that attorneys should consider such an approach to contract and tort cases and that statisticians should begin to ponder recasting their methods of analysis and presentation so as to make their findings usable, and persuasive, in a broad range of legal matters.

The facts in the *Brink's* case were summarized in the decision on appeal by federal Judges Oakes, Cardamone, and Winter of the United States Court of Appeals for the Second Circuit (717 F.2d 700 (1983), 702–718):

> The City of New York owns and operates approximately 70,000 parking meters, mostly on-street, but some in metered parking lots. Daily collections average nearly $50,000. After public bidding in March of 1978, Brink's was awarded the contract to collect coins from these meters and deliver them to the New York City Department of Finance Depository. Brink's was reimbursed at the rate of 33 cents per day for each meter it collected. Operations under the contract commenced in May of 1978.
>
> The process of collection involved collectors working in teams, with one individual going to each parking meter on a city-prescribed schedule. To collect a meter, the collector would insert a meter key, open the bottom portion of the meter head, and remove a sealed coin box. This coin box would then be placed upside down onto a gooseneck protruding upward from a large metal canister on wheels, resembling a dolly, that the collector rolled along from meter to meter. By twisting the coin box, an unlocking device called a "cannister key," attached to the canister by City personnel and theoretically not reachable by the collector, would allow the coins from the coin box to drop into the canister without the collector's having access to the coins. After the coin box was emptied the collector would replace it and lock the meter. Upon completion of a specific route the canister would be placed in a collection van. At the end of a given team's work on a particular day, the van was driven back to the PMD [Parking Meter Division] and the canisters were turned over to City personnel.
>
> As can be seen from the above, collectors were not to have contact with the coins in the meters. City personnel were to check each collection canister daily to ensure that it was in good working order; upon receipt of a canister the Brink's people signed a receipt attesting that it was in good working order. Under the contract, at the end of the day any canister malfunction or broken or uncollectible meters were to be reported to the City. Brink's was to provide ten three-person collection crews daily, the crews to be rotated in the discretion of the City Department of Finance as a security measure. The City directed that the rotation should be accomplished by a lottery system to prevent the formation of permanent teams, but the City's evidence was that the rotation system was ignored. Daily assignments were frequently made by the collectors themselves and the management of Brink's was aware of this but did nothing to correct it. As the trial judge stated, the jury "could readily have found that the planned rotation system was honored more in the breach than in the observance." 546 F.Supp. at 409. The contract also provided that Brink's would provide supervisory personnel to oversee the proper performance of its obligations. Although the contract called for two supervisors and one field inspector, the full complement was never assigned.
>
> In response to an anonymous tip, the City's Department of Investigation, in conjunction with the Inspector General's Office of the Department of Finance, began an investigation of parking meter collections. Surveillance of Brink's collectors revealed suspicious activity violative of both the City's and Brink's rules and procedures. The investigators then "salted" parking meters by treating coins with a fluorescent substance and inserting them into specific meters. "Salted" meters were checked after the coin boxes had been emptied by Brink's employees to make sure that all of the treated coins were collected; collections from the meters were then scanned to see if any treated coins were missing. The "salting" process indicated that a substantial percentage of coins collected by Brink's personnel were not being returned to the City. Surveillance at the 42 Franklin Street depository revealed that at the end of a day Brink's employees would often arrive in personal vehicles following their assigned collection vans, indicating some kind of "drop-off." Brink's employees were also seen entering a parking lot in Manhattan in Brink's van, placing them in private automobiles and then returning to the vans to continue to the City depository. The City presented video tapes of these transfers at trial; Brink's collectors were shown straining to lift heavy bags into their cars. Brink's employees were also followed to a private residence and again observed carrying heavy bags from their vehicles into the building and emerging empty handed.
>
> James Gargiulo, Trevor Fairweather, Richard Florio, Michael Solomon and John Adams were arrested and charged with grand larceny and criminal possession of stolen property when on April 9 they had in their possession over $4,500 in coins stolen *that* day from parking meter collections. Anthony DeNardo was arrested and charged with petit larceny and criminal possession of stolen property. A charge against Jorge Olivari was dismissed on motion by the district attorney, but the remaining six defendants were either convicted after trial (Florio, Solomon and Adams) or pleaded guilty before trial (Fairweather, Gargiulo, DeNardo). They were sentenced to varying jail terms and fines ranging from $1,000 to $5,000.

This chapter in outline proceeds as follows. Section 2 discusses legal criteria that governed proof in court of the amount of damages suffered by the city and incurred by Brink's. Section 3 defines the choice of a statistical comparison to yield legally admissible evidence of the amount of theft. Section 4 discusses aspects of the statistical analysis: issues of causal attribution; treatment of trend, seasonality and choice of comparison periods; and a linear model for revenues per meter-day. Section 5 gives the Brink's view of the facts in the case as presented by their statistical expert. Finally, Sections 6 and 7 tell the outcome of the case in the trial and appeals courts.

2. Law of Proving Damages in *Brink's*

The city faced three hurdles in its quest for a substantial verdict in this case. First, it had to persuade the judge that numbers of Brink's employees had stolen coins on numbers of occasions, and that the thefts could be found by a jury to be part of a pattern. If the judge were not persuaded of this, under the applicable law, damages could only have been awarded for thefts as to which there was individualized evidence, such as videotapes of bags being transferred or short-counts from the "saltings."

Fortunately, the judge allowed the city attorneys to demonstrate the pattern of theft as a legal matter through the introduction of evidence, such as videotapes of six days of observations and the results of the "saltings," before the jury. As the jury had heard and evaluated the evidence of each occasion of alleged theft, it is not surprising that the judge allowed the jury to conclude, if it so chose, that the thefts were repeated and in all likelihood occurred when unobserved by the cameras. To allow the jury to decide whether to infer continued misconduct—here theft—from the evidence is well supported in law. See *McFarland v. Gregory,* 425 F.2d 443 (2d Cir. 1979), and *Conner v. Union Pacific,* 219 F.2d 799 (9th Cir. 1955). But such an inference would not, under the case law, permit the jury to decide the amount of the damage caused by the misconduct; additional proof would be necessary.

Second, the city had to persuade the judge that the Brink's company could legally be held liable for the thefts of its employees. In general, an employer is responsible for the negligence or other dereliction of its employees, under a legal doctrine called *respondeat superior,* a Latin phrase that translates loosely as "let the boss be held responsible." This doctrine, under the law of New York that governed this case, permits a court to hold an employer liable even for the intentional "torts" (civil wrongs) of his employees. For example, an employer can be compelled to compensate a visitor to a factory who is punched for no reason by a plant security guard. But this doctrine does not, the trial judge in *Brink's* indicated, go so far as to hold an employer liable for thefts by his employees; as lawyers put it, theft is "outside the scope of employment" of an employee.

There are, however, two other principles that the city invoked in *Brink's*. The city argued that since Brink's had agreed in its contract to collect all the money in the meters and deliver it to the city, Brink's should be considered an insurer. Had the judge agreed with this theory, the city would have been entitled to collect the difference between the money put in the meters and the money that arrived at the counting facility, whether that difference was due to theft, carelessness, or any other reason. The statistical inquiry in that case would have been simplified, but the judge held, before the trial, that this theory would not be used in this case. The court permitted a contract claim to go to the jury, but required that the city prove willful or negligent failure to meet the supervision requirements written in the contract in order to prevail. The jury in fact found for the city on this claim, which had the result of starting interest on the judgment running from the date the contract was breached; under the common-law negligence claim, discussed immediately below, interest begins running only from the date the judgment is entered.

The city's remaining theory was that Brink's had been negligent in supervising its employees and in failing to fire them after it learned—or should have learned—that they were stealing. This breach of the historic common-law duty of an employer to the public is the theory on which the case was finally tried. Because it was relegated to this theory, however, the city was forced to call numerous witnesses from Brink's middle and upper management to establish the company's negligent supervision and retention. This had the effect of diverting the jury's attention from the problem of ascertaining how much was actually stolen.

As it turned out, the negligence of Brink's management, particularly in ignoring reports of its own security personnel that drivers and collectors were stealing, was deemed so gross by the jury that it awarded the city $5,000,000 in punitive damages. Such damages under the law are not to be calculated with reference to actual loss, and are meant to deter defendants and others in similar positions from such egregious behavior in the future. By definition, statistical evidence about the extent of the loss suffered is not relevant to the calculation of such damages.

Finally, the city had to prove the total amount of revenue it had lost from theft. In the absence of counters on the actual parking meters, the precise amount of money deposited could not be ascertained. The city knew that on the day of the arrests over $4000 had been recovered, but it felt sure that other thieves had gotten away with even more money that day. Without an actual count of the loot on any other day, the city lawyers believed that even if the jury were to find that there had been repeated and massive theft, it would have no way to estimate the total taken. Perhaps it would simply multiply the $4000 actually recovered by some number of days it would come up with on which it thought theft had occurred. From analyzing the reported appellate decisions, it seemed likely that such a result would be set aside by the court as mere guesswork, and thus not a legal basis on which to assess damages.

The city's attorneys decided, therefore, to seek the assistance of a statistical expert to attempt to prove how much money Brink's should have collected during its contract, and then, by subtracting actual revenues Brink's had turned in, to calculate the shortfall attributable to theft. How this was done is the subject of the body of this article. But from a legal point of view, it turned out to be crucial that an expert made that calculation. For both the trial judge and the appellate court in *Brink's* stated that the city had to first show (717 F.2d 700 (1983), at 712):

> evidence independent of the City's experts' testimony establishing that systematic theft had occurred over a long period of time,

and then must show a total revenue loss through competent expert testimony.

Where, as here, the expert's calculations are based on comparing revenues delivered in two different time periods by two different collection companies, that expert must devise and employ a methodology that, with a reasonable degree of certainty and in conformity with established and accepted statistical practices, enables him to state that (717 F.2d 700 (1983), at 711):

> differences in various nonculpable factors had been adequately accounted for and that conditions in the comparison periods were substantially the same.

Of course, the methodology must be shaped by the comparison to be drawn; if, for example, conditions had changed during the comparison period but, taking those changes into account, theft was still statistically demonstrable, the expert's testimony would be admissible and if believed by the jury, would support an award of damages. Revenue comparison over different time periods as a general approach was approved in the United States Supreme Court case of *Bigelow v. RKO Radio Pictures, Inc.*, 327 U.S. 251 (1946). Failure to use appropriate data, or to take varying conditions appropriately into account, yields not only bad statistics but bad legal results; an award of damages is likely in such a case to be reversed (see *Herman Schwabe, Inc. v. United Shoe Machinery Corp.*, 297 F.2d 906 (2nd Cir. 1962, cert. den. 369 U.S. 865)). Both the *Bigelow* and *Schwabe* opinions are well worth reading by any statistician contemplating using a comparison method in analyzing the effect of variables on revenue production.

The statistical evidence presented in the *Brink's* case met the basic test that had been established over 50 years ago for cases where the acts of the malefactor—such as theft of untallied receipts—prevent precise calculation of the extent of monetary loss:

> In such case . . . , while the damages may not be determined by mere speculation or guess, it will be enough if the evidence show the extent of the damages as a matter of just and reasonable inference, although the result be only approximate. The wrongdoer is not entitled to complain that they cannot be measured with the exactness and precision that would be possible if the case, which he alone is responsible for making, were otherwise. (*Story Parchment Co. v. Paterson Parchment Paper Co.*, 282 U.S. 555 (1931))

3. Framing of the Statistical Question: The Lawyering Consideration

In looking for a way to quantify the damages, the city's lawyers knew that there was no way to count the money actually put into the meters by drivers, and that the first actual count occurred after Brink's had delivered the coins. They also knew that the same system that had clearly failed to keep Brink's honest had been in use by the prior collection company, when revenues delivered had been substantially below those delivered by Brink's; in retrospect, city officials had their suspicions over the honesty of the predecessor collectors. Immediately after Brink's had been taken off the contract, however, on the day following the arrests, a new collector, the CDC Company, had been hired, and within six weeks had taken over all collection work. In its first 10 months of operation, CDC had brought in nearly $1,000,000 more than Brinks had in any 10-month period.

This company worked under strict surveillance by city investigators, and under a procedure of both routine and surprise polygraph examinations of its drivers and collectors. The city's faith that the new collection company was honest was strengthened when, about four months into the contract, a collector failed a polygraph and confessed that he had in fact stolen a coin box from a meter and kept some $20 in coins. It thus seemed that the new security system was effective in preventing any systematic theft.

The city lawyers thus went to the statistical expert with the idea that a comparison of Brink's suspect collections with CDC's honest collections would be likely to yield legally admissible conclusions. After all, the city knew it could prove at least the last day of theft during the Brink's period, and that it could introduce persuasive evidence of theft for a number of other days. The statistician asked what factors other than theft might account for an increase in collections after Brink's was terminated, since it was at least arguable that the videotapes and "saltings" had isolated a small ring of occasional thiefs, and the million dollar difference was otherwise explicable. The city's attorneys went to the heads of the Parking

Meter Division of the New York City Department of Finance, and learned that at least one high-ranking official had made the off-the-cuff comment—duly reported in the *New York Times*—that he thought most of the increase was due to better meter repair. Other supervisory-level personnel pointed to the conversion of some meters from a dime to a quarter and to possible increased car use by commuters because of the end of the 1979 gas shortage as possible causes for increased revenues.

The statistician, William Fairley, played a crucial role at this point, because he questioned whether the suggested factors would account for the dramatic climb in revenues between the last month of Brink's collections and the first month of CDC's. Brief investigation indicated that meter repairs and rate changes in any given month were insignificant, while calculations that had been done in the course of the preparation for trial of the arrested Brink's employees seemed to show extraordinary revenue differentials between Brink's and CDC's collection in certain collection areas frequented by the arrested men. The statistician recommended an in-depth study of comparative collections and a detailed analysis of the changes in revenue potential of the meter plant over as long a period as possible. He also urged the city to hire an expert on parking meter theft, because he could only evaluate the impact of factors that might affect revenue over time; he could not identify these factors.

In what for a budget-conscious municipality was, we believe, a startling leap of faith, the City of New York through its elected Board of Estimate hired the statistician, realizing that his extensive analysis might show that theft was one, but not the only, explanation for Brink's level of revenue production. The city also hired a nationally known expert in the analysis and design of security systems for parking meter plants and parking garages, Laurence Donoghue. When he identified over 40 possible causes of change in parking meter revenues, the analysis seemed doomed. When he was able to conclude that many of these variables either were not present in the New York area during the comparison period, such as the opening or closing of a unique shopping site, were equally present during the Brink's and CDC periods, such as the incidence of legal holidays, or were incapable of making a significant difference, such as the fact that snow emergency streets had parking bans for one more day in the Brink's period than in the CDC period, the picture brightened considerably.

As it turned out, and as the remainder of this chapter demonstrates, the analysis chosen enabled the statistician to conclude that theft accounted for the intercompany revenue difference, and that but for the theft Brink's collections would have exceeded those of CDC for a 10-month comparison period by some $1,400,000. And while one cannot probe the minds of jurors, the jury in fact returned a verdict for the city of $1,000,000 in compensatory damages, and this verdict was affirmed by both the trial judge and the court of appeals.

4. The City's Factual Case

Goals of Statistical Analysis

Theft of unknown dimensions had occurred over an indefinite period of months or years prior to the arrest 6 April 1980 of seven Brink's collectors. The factual question to be answered in determining the amount of damages suffered by the city was how much had been taken.

The law does not here, or in other contexts, spell out precisely the criteria it seeks for estimates in terms of the theories of statistical inference. However, the fact-finding goals of a trial and the roles assigned to the judge and jury suggest the following requirements for the precision and bias of estimators. As discussed above, in determining damages the law does not require any specified level of precision in the estimate, but it does require the use of available data and appropriate methods of estimation as opposed to pure conjecture or speculation as to the amount.

The law looks for an unbiased estimator in the subjective sense that the maker of the estimate have no more reason to believe the estimate too large than too small (see *Sargent v. Mass. Accident Co.* 307 Mass. 246 (1940)). Translated into the language of Bayesian inference, the maker of the estimate could if he chose report the median of a posterior distribution for the quantity. The goals of the statistical analysis could be summarized nontechnically as determining a reasonable estimate of the amount of theft, given available data and practical limitations on the depth of the investigation determined by time and money budgets.

Causal Attribution

The approach chosen to estimate the amount of theft was to compare revenues delivered by Brink's prior to the arrests with revenues delivered by the subsequent contractor, CDC, after the arrests. In making such a before-and-after comparison to determine the size of effect of a known causal intervention, it is useful to distinguish two types of threats to a causal interpretation. These threats to causally interpreting the observed difference as most reasonably attributable to the intervention—as opposed to other factors—are (a) a difference before and after could be due to *general time effects* of trend or seasonality, namely, a general trend in revenues for whatever reasons, or seasonal effects in the periods before and after; or (b) a difference could be due to *specific causes* that could account for the change before and after.

A two-step approach was therefore taken to investigate the validity of interpreting a before-and-after difference as a theft amount estimate. The first step was to investigate the existence of the general time factors of trend and seasonality. The second step was to consider whether other causes were reasonable candidates to explain the kind of difference observed.

Seasonality and the Choice of Periods of Comparison

The effect of seasonal differences on a difference between the two periods was controlled both by choice of periods and by seasonal adjustments.

The two comparison periods of 10 months each were chosen to be the same 10 calendar months, namely, June 1979 through March 1980 for Brink's and June 1980 through March 1981 for CDC. Since the same months were compared, seasonal differences in the two periods were eliminated when comparing the full two periods. Seasonal adjustment was still of interest for the purpose of comparing revenues in the months immediately before and after the transition. The two-month gap in April and May of 1980 between the two comparison periods was necessary because CDC was not fully collecting the same routes as Brink's until June 1980.

A period of 10 months was chosen because in the spring of 1981, when data collection for the study was undertaken, only data for 10 months up through March 1981 was available for CDC, the successor contractor to Brink's. The choice of these 10 months for CDC then suggested matching these with the same 10 months for Brink's one year earlier. Data earlier than this for Brink's was not obtained because the quality of the data was uncertain and a more distant period would likely be subject to more change in other factors that would be difficult to study.

Several adjustments for the purpose of comparing months' revenues immediately preceeding and succeeding the transition were made by a model for the data, as discussed below.

Trend

The question of a general trend over the 22-month period June 1979–March 1981, consisting of 10 months of Brink's collections, a 2-month gap, and 10 months of CDC collections, was investigated graphically and analytically.

Graphs of average monthly revenues delivered per meter-day of operation for the entire city and for each of the five boroughs, either without seasonal adjustments or with monthly seasonal adjustments, did not indicate a trend over time but rather indicated level revenues per meter-day within each 10-month period. Seasonally adjusted monthly revenue per meter-day data were displayed by dividing each city (or borough) actual average for the month by the linear model-predicted monthly revenue (by model defined below) and multiplying by the period average for the city (or borough).

Piecewise-linear regressions were fitted to data on monthly revenues over the 22-month period, one allowing a change in level between the two periods but fitting the same slopes, and the other a change in level as well as a change in slopes between the two periods. Both regressions estimated a positive jump in level and neither estimated an upward trend over both periods.

Fitting a piecewise-linear regression has the following advantage over fitting a single regression line. If there is a trend over the entire 22-month interval, the piecewise-linear regression will estimate it, and if there is a change in level, it will also estimate that. A single trend line, however, would not estimate a change in level if it were present. Thus, a single trend line fit provides no way to measure the amount of theft.

Figure 1 plots the residuals from a single line fitted to all of the data. The single trend line tends to underestimate the early months and to overestimate the later months in both periods, with this effect being most pronounced in the second period. The pattern of the residuals strongly indicates misspecification in the fit by a single line.

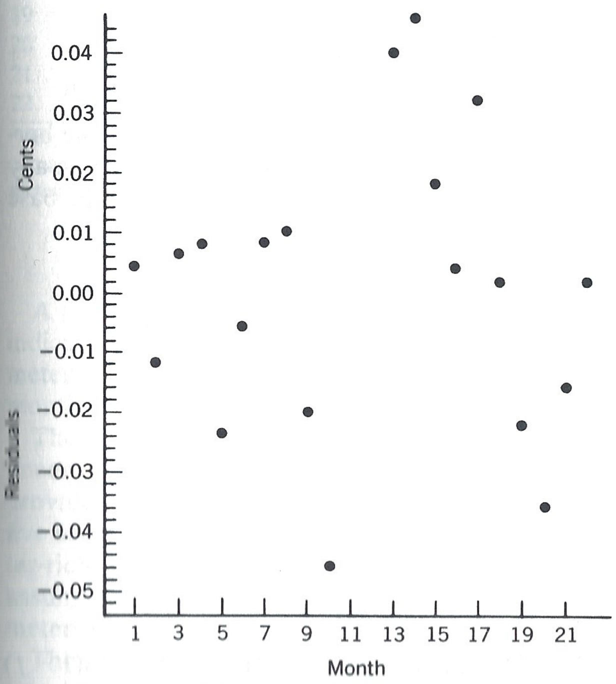

Figure 1. Residuals from trend line fitted to citywide seasonally adjusted revenues per meter-day.

The conclusion both from graphical inspection and from regression analysis is that there is no evidence of a rising general upward trend over the 22-month period that would account for the higher average revenues found in the second 10-month period versus the first 10-month period. Rather, there is evidence for a change in level at the transition.

Linear Model for Revenues Per Meter-Day

A model was specified to explain the dependent variable of average monthly revenues per meter-day for each contractor as a function of an intercontractor differential rate and of month-by-borough factors. The dependent variable monthly average revenues per meter-day by borough for the ith month, jth borough, and ath contractor is denoted by $w_{ij}^{(a)}$ and is defined as

$$w_{ij}^{(a)} = \frac{\sum_k c_{ijk}^{(a)}}{\sum_k (m_{ijk}^{(a)} d_{ijk}^{(a)})}$$

where

$c_{ijk}^{(a)}$ = reported revenue for the kth collection
$m_{ijk}^{(a)}$ = number of meters collected in the kth collection
$d_{ijk}^{(a)}$ = number of days since the last collection for the kth collection.

The model posits a base revenue per meter-day depending on the borough–month combination t_{ij}, a single average intercontractor rate in any borough and any month f, and an error term $e_{ij}^{(a)}$ proportional to the base rate:

$$w_{ij}^{(a)} = (1 - f)^{Z^{(a)}} t_{ij} e_{ij}^{(a)},$$
$$i = 1, \ldots, 10;$$
$$j = 1, \ldots, 5;$$
$$a = 1, 2 \text{ (1 for Brink's, 2 for CDC)}$$

where

$$Z^{(a)} = \begin{cases} 1 & \text{if } a = \text{Brink's month} \\ 0 & \text{if } a = \text{CDC month} \end{cases}$$

and

$$\ln e_{ij}^{(a)} \sim N(0, \sigma^2).$$

After linearizing the model by taking logs, the 51 parameters, $\ln(1 - f)$ and $\ln t_{ij}$, are estimated by least squares using the 100 observations for borough–month combinations.

TABLE 1. Residuals from Trend Line Fitted to Citywide Seasonally Adjusted Revenues per Meter-Day[a]

X	Y	YHAT	RESID
1	0.863	0.858351	0.004649
2	0.852	0.863208	−0.011208
3	0.875	0.868065	0.006935
4	0.881	0.872922	0.008078
5	0.854	0.877779	−0.023779
6	0.876	0.882636	−0.006636
7	0.895	0.887493	0.007507
8	0.903	0.892350	0.010650
9	0.878	0.897207	−0.019207
10	0.857	0.902064	−0.045064
13	0.956	0.916636	0.039364
14	0.968	0.921493	0.046507
15	0.944	0.926350	0.017650
16	0.936	0.931207	0.004793
17	0.968	0.936064	0.031936
18	0.943	0.940921	0.002079
19	0.923	0.945778	−0.022778
20	0.914	0.950635	−0.036635
21	0.939	0.955492	−0.016492
22	0.962	0.960349	0.001651

[a] Key: X = month; Y = revenues per meter-day, actual; YHAT = revenues per meter-day, predicted; RESID = residuals.

A constant intercontractor rate model is consistent with the observed indication of proportionality in the amounts by which CDC revenues per meter-day exceeded Brink's in every borough and in 47 of 50 borough–month combinations.

There are two principal advantages of this model over the piecewise-linear regressions for estimating an intercontractor rate. First, the model provides more degrees of freedom for estimating the parameters than do five separate regressions for each borough. Second, the model is parameter-rich, fitting all borough–month interactions, thereby making minimal assumptions about the way these two factors relate to revenues per meter-day.

The regression R^2 was 0.9937 and $s_E = 0.0438$, thus giving approximately a 4.4% prediction error for $w_{ij}^{(a)}$. Brink's predicted versus actual values shows good fit over the entire range of values of y (the dependent variable). The predicted values of average monthly revenues per meter-day for both Brink's and CDC closely match the actual values in each borough and in each month.

The borough–month parameters were almost all highly significant, and the intercontractor parameter $\hat{\beta} = \ln(1 - \hat{f})$ was highly significant:

$$\hat{\beta} = -0.0815$$
$$\hat{\sigma}_{\hat{\beta}} = 0.0088$$
$$t = -9.29.$$

The estimated intercontractor proportion $\hat{f}$ is

$$\hat{f} = 1 - e^{\hat{\beta}} = 0.0782$$

which, unless other factors indicate a different rate for diversion (see discussion below), is estimated diversion as a rate of expected revenue without diversion. The rate of expected revenue with diversion is

$$\hat{f}/(1 - \hat{f}) = 0.0849 \approx 8.5\%.$$

The total amount estimated diverted in the 10-month Brink's period can therefore be determined as

(Diversion rate from expected revenues)(Expected total revenues in 10 months)

$$= \hat{f}\Big(\sum_{ij} \hat{w}_{ij}^{(2)} u_{ij}^{(1)}\Big)$$
$$= 0.0782 \times \$17{,}674{,}608$$
$$\approx \$1.4 \text{ million}$$

where

$$u_{ij}^{(1)} = \text{number of meter-days collected by Brink's in month } i \text{ and borough } j$$
$$= \sum_k m_{ijk}^{(1)} d_{ijk}^{(1)}.$$

Equivalently, total amount estimated diverted in the 10-month Brink's period can be calculated as

(Estimated amount diverted per meter-day)(Number of meter-days)

= (Diversion rate from expected revenues)

× (Expected revenues per meter day)(Number of meter-days)

= 0.0782(0.9452)(18,698,946)

≈ (7.4¢ per meter day) × (18,698,946 meter-days)

≈ \$1.4 million.

The model estimate of diversion rate is smaller than would be determined by first looking at the difference in revenues over 1, 2, 3 months.

For example, the jump in revenue in the gap between contractors between March 1980 and June 1980 was 13.9%. By adjusting for number of meter-days in each month and for seasonal effects in March and June the predicted revenue per meter-day for Brinks is 88.2¢ in March and for CDC is 98.4¢ in June, for an increase of only 11.6% instead of the observed 13.9%. The model therefore is useful for getting a better estimate of diversion by controlling for other factors.

Other Specific Causal Factors

The general time factors of long-term trend and seasonality do not explain a jump in level that occurred at the transition between Brink's and CDC. But perhaps there are specific causal factors other than theft that can explain the jump. Given the facts in the case, such factors should have the following four properties:

1. **Suddenness.** Be capable of causing an upward shift in level over a two-month gap.
2. **Sizableness.** Vary in magnitude or effect enough to be capable of causing a shift of about 8.5% over Brink's level over a two-month gap.
3. **Uniqueness.** Vary in a way that explained the shift in level at the two-month gap and the absence of a shift in level elsewhere.
4. **Uniformity.** Be capable of explaining a shift in level over the two-month gap in every borough.

Price changes of meters was a factor that could potentially have all four required properties. It was thought worthwhile therefore to investigate this factor. Six so-called Area Description tapes containing information on price changes for all 70,000 meters in the city were examined for changes that would increase or decrease the effective price. There were five types of such changes: size of coin accepted by meter; maximum allowable time limit; hourly rate; active days; and active hours. A count of changes expected to decrease and changes expected to increase revenues over the two-month transition period showed 73 increases and 158 decreases, so that the direction of effect on revenues in terms of the frequencies of such changes would be to decrease expected revenues in the month CDC took over below that expected from Brink's at the end of its tenure. The total of all changes in the two-month transition period was 231 or 0.33% of the 70,000 meter population. Over the entire 22-month period the number of changes expected to decrease revenues slightly exceeded the number expected to increase them, and the excess was greater in the CDC period. Therefore, again, the expected change in revenues was a decrease for CDC versus Brink's.

Thus, the net expected direction of effect of the factor of price changes was downward both at the transition and over the entire 22-month period. This factor in fact fails every one of the four tests of an explanatory factor.

A number of other factors were considered as potential explanations of the jump in level, including meter maintenance, meter back-up, city collections, installation and removal of meters, and the gas shortage. For some factors, pertinent data were available that threw light on whether changes in the factor satisfied the four properties of causal factors. The expert on parking and meter operations was asked for his judgment on the likely revenue effects of a list of factors that were not explicitly controlled for in the statistical model. He testified that he had no more reason to believe that the uncontrolled factor would cause an increase than a decrease, either over the transition or over the entire 22-month period.

In sum, no evidence was presented that demonstrated that specific causal factors other than theft would account for the jump in level between the two periods. Of course it is impossible to say that there were no such factors, for there could be. The jury, however, had to determine the most reasonable estimate it could of the amount of theft. Given the complexity of factors in the world that influences meter use, no complete causal model could ever be expected, certainly none within the time frame of the trial. A general statistical model, specified and estimated in such a way as to explicitly control for, or to estimate, the net effects of time-varying factors, is the only way that such an estimate could be made.

5. The Brink's Factual Case

Brink's had a different view of the facts. Their principal statistical witness was Bruce Levin, a statistician and professor at Columbia University. He testified after the city's experts, William Fairley and Laurence Donoghue.

Levin's conclusion was that other factors besides theft *could* explain the greater revenues delivered to the city by CDC in the 10-month period, June 1980 through March 1981, than delivered by Brink's in the same 10-month interval a year earlier, June 1979 through March 1980. He stated that on statistical grounds alone it was not possible to attribute either all or any part of the difference in revenues to theft.

These conclusions were reached not only by an analysis of the same monthly revenue data used by the city's experts, but also by using additional monthly revenue data going back to the beginning of the Brink's contract in May 1978. The analysis was done separately on data over the period of the Brink's contract for what is apparently the only area of parking meters in the city whose meters had been consistently collected by city personnel and not by either contractor. This area was coded in city records as Area 1A.

The analysis of revenue data was directed at the question of whether a "trend" or a "trendline" existed in the two areas consisting of (a) the city as a whole excluding Area 1A and (b) Area 1A by itself. Two approaches to this question were taken. The first was to calculate five-month moving averages of monthly revenues over the period of the Brink's contract and the CDC contract, May 1978 through March 1981, in the two areas.

The graph of moving averages for the city excluding Area 1A showed a generally rising series of points with some small jags downward and a fall in the last several months of the CDC contract. The graph for Area 1A showed some indication of rising level but with considerable up and down variability. The conclusion reached was, first, that there was an upward trend operating over the period of the Brink's contract, and second, that the increase in revenues in Area 1A in the CDC period in excess of the revenues in the Brink's period showed to a reasonable degree of certainty that other factors besides theft were accounting for the increase in revenues observed citywide.

The second approach to the data was to fit a regression line separately to monthly city revenue excluding Area 1A, and to monthly Area 1A revenue, in the period of the Brink's contract, May 1978 through March 1980. The line fitted to the monthly city revenue excluding Area 1A had a statistically significant positive slope of $17,960 per month revenue increase and a correlation coefficient of 0.69. When this line was extrapolated to the CDC period, the predicted excess of revenues in the 10-month CDC period over the corresponding 10-month Brink's period a year earlier was $2,155,000, compared to an observed excess of approximately $1,000,000. The significant positive slope of the fitted line established the existence of a trendline, and this trend could account for the existence of excess revenue in the CDC period over the corresponding Brink's period. The line fitted to monthly Area 1A revenue also had a positive slope. The conclusions from this regression analysis paralleled those from the moving average analysis. Trend accounted for by other factors than theft could explain the excess revenue delivered by CDC in its 10-month period.

Levin did not testify that he believed *specific* other causal factors explained the observed difference in revenues between the CDC and Brink's periods. He did testify that each of a list of several specific factors raised in questions by counsel to Brink's were potential biasing factors between the two periods, leading him to believe in a "strong possibility" that the two periods compared were "not comparable."

On cross-examination Levin agreed that there was no evidence of trend within either of the two comparison periods, the CDC 10-month period and the corresponding Brink's 10-month period.

Cross-examination also brought out the fact that there were some 45 meters in total in Area 1A, the total meter plant having some 70,000 meters. Since the analysis of monthly revenues was not on a per meter-day basis, this implied that a change of only one in the meter population of Area 1A due to disrepair, addition, or removal could be expected to cause about a 2% change in revenues.

6. Jury Verdict and Judges' Ruling and Opinion

The Federal District Court jury found Brink's liable for negligence with respect to supervising its employees and inspecting the process of collection. It awarded the city compensatory damages of $1,000,000 for loss of revenues due to theft and $5,000,000 punitive damages.

The trial judge, Judge Edward Weinfeld, ruled that the evidence sustained the jury's findings of negligence and the amount of damages, although he did reduce the amount of punitive damages awarded to $1,500,000. In his opinion he compared the methods of analysis presented by the two sides that produced conflicting evidence at trial and explained that such conflict was the jury's role to resolve, not by guess or speculation, but by finding a fair and reasonable estimate based on the evidence presented.

7. Appeal on the Factual Case

Brink's appeal of the case to the U.S. Court of Appeals on the factual aspect of the case contended that the trial judge had erroneously permitted the jury to hear expert testimony about the amount of theft, which was wrong because the CDC and Brink's periods were not comparable; that other factors the City's experts did not consider might well have accounted for the difference in revenues; and that the evidence of a trend in revenues could explain the difference. Brink's argued that the damage estimate was too speculative and the jury finding of a damage amount for the city was based not on a rational consideration, but on a delusive impression of exactitude generated by an array of figures.

In rejecting this basis for appeal, the court made several points. First, while Brink's advanced the *possibility* of other factors explaining the difference, it nowhere introduced evidence that these factors did explain the difference. On the other hand, the city's experts studied differences in various nontheft factors and Fairley concluded that these had been adequately accounted for, and that there was no apparent basis for thinking that factors left out biased the estimate made of difference attributable to theft. The jury believed the city's experts.

Second, the fact that the city's experts could not rule out the possibility that other factors explained some of the estimated theft amount and that Fairley could only offer what he thought was the most reasonable estimate, did not imply that the estimate presented required the jury to guess or engage in mere speculation to arrive at damages.

Third, the estimate was not used to establish the *existence* of damages due to theft. Systematic theft over a long period of time had been independently established, and an estimate was required of the amount of these damages.

Acknowledgments

The authors thank contributors to the project: Patricia Kruger, Wendy Love, Linda Lutz, Linda Scharff, JoAnn Tankard Smith, and Edward Tuozzo. Peter Kempthorne proposed the linear model discussed in Section 4.

References

Baldus, D. C., and Cole, J. W. L. (1980). *Statistical Proof of Discrimination,* New York: McGraw-Hill.

Fairley, W. B. (1983). Statistics in law. In S. Kotz and N. Johnson (eds), *Encyclopedia of Statistical Sciences,* New York: John Wiley & Sons, Vol. 4, pp. 568–577.

Finkelstein, M. O. (1978). *Quantitative Methods in Law,* New York: Free Press.

Sugrue, T. J., and Fairley, W. B. (1983). "A Case of Unexamined Assumptions: the Use and Misuse of the Statistical Analysis of *Castaneda/Hazelwood* in Discrimination Litigation," *Boston College Law Review,* **34,** 925–960.

Comment

Bruce Levin
Division of Biostatistics
Columbia University School of Public Health
New York, New York

Fairley and Glen (FG) have discussed a two-period comparison of gross revenues that raises two problems I wish to discuss in this comment. The first is their assumption that the data are best described by two horizontal lines without trend, separated by a discontinuity at the April–May 1980 juncture. The second is that the dollar size of the gap (or the intercontractor rate parameter in the FG model) is entirely attributable to theft. It cannot be too strongly emphasized that such assumptions are not to be granted without careful scrutiny in an observational study, for it is well known that potentially large biases can result from uncontrolled confounding factors. In my view the justifications advanced by FG are unconvincing on both methodological and substantive grounds, rendering their estimate of theft unpersuasive. I believe that, contrary to FG's conclusion, the data at hand do not support more than a speculative estimate of the amount of revenue differential attributable to theft.

Consider first the question of trend. FG address this issue via a hypothesis test of zero trend versus upward trend in a broken-line model (with and without a parallelism assumption) based on data limited to the two 10-month periods (with and without "seasonal adjustments"). Figure 1 presents the average daily revenue received per collection day for the *entire* period of Brinks' contract (May 1978–April 1980) together with the average daily revenues in the subsequent CDC period. The data exclude revenues received from Area 1A that was at all times collected by city employees and not exposed to theft by Brinks employees. (The revenue per meter-day unit used in FG is a figure representing the average revenue

per meter per day. The unit of revenue per collection day used in this comment indicates monies returned by the collection crews on an average collection day. The latter unit seems more closely related to the measurement of theft. While it does not adjust for variations in the size of the meter plant or in the number of days between collections, it does not appear that either factor changed significantly between the two 10-month periods used by FG. Neither measure adjusts for other variables discussed in the sequel.) The trend line superimposed on the graph was produced with Cleveland's robust locally weighted regression algorithm for smoothing scatterplots (Cleveland, 1979). The smoothing fraction used was 0.3, the local weight function was bicubic, and one robust itera-

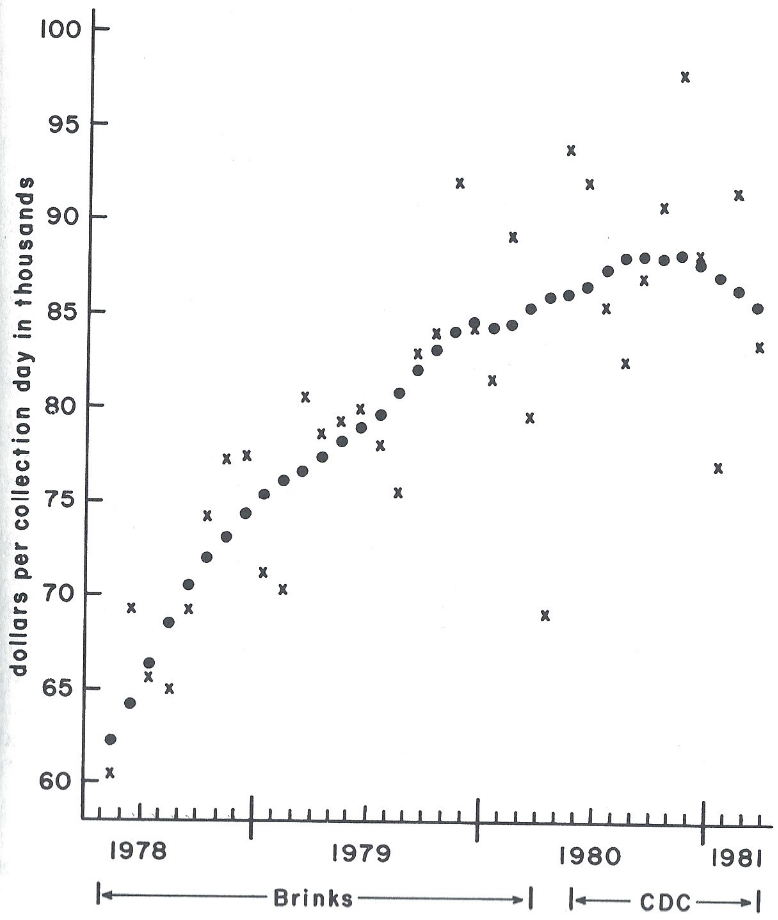

Figure 1. Average monthly revenue per collection day.

tion was used with a bisquare residual weight function (see the reference for definitions of these terms). The picture shows evidence of a general upward trend throughout the Brinks period tapering off into the CPC period. It is apparent that the total effect of all factors influencing revenues was *not* constant during the Brinks period.

How can FG have overlooked this trend? In the first place, a time base of only 10 months in the Brinks period is too narrow to establish the trend clearly. The reasons FG advance for failing to consider the previous year's worth of data are weak; one cannot ignore a major portion of relevant data simply because its "quality" was "uncertain" or affected by factors more "difficult to study." In fact Fairley has forcefully and correctly argued elsewhere that effects are not reliably estimated from a brief span of experience, and that the widest related base of experience should be surveyed in statistical estimation problems (see Fairley, 1979a, especially pp. 335–336).

In technical terms the statistical power of their hypothesis test was apparently very low (i.e., there was little likelihood of detecting the trend). To demonstrate the lack of power when looking through only the two 10-month windows, I simulated 1000 data sets of 20 points each with the underlying trend indicated by circles in Figure 1, with error terms generated by sampling with replacement from the set of 20 actual residuals obtained from the data smoothing (a bootstrap procedure). Of the 1000 slope coefficients obtained from a parallel broken-line model fitted by least squares to the simulated data, only 11.6% had t-statistics exceeding 2.0. Of an additional 1000 slope coefficients corresponding to the Brinks period from a nonparallel broken-line model, only 28.3% had t-statistics exceeding 2.0. It is thus no surprise that FG could not detect the trend in the Brinks period.

In the second place, FG appear to have made a methodological error in their seasonal adjustments of revenue per meter-day that formed the basis of conclusions drawn from FG Figure 1 and Table 1. The method is apparently a multiplicative version of one described by Fairley for adjusting data for seasonal effects to make possible further checking into the presence of a time trend (see Fairley, 1979b, discussing trend and seasonality). FG calculated seasonally adjusted revenues by dividing each month's revenue figure by a "seasonal effect," which is for each pair of months an average of the two months' revenues divided by the grand mean revenue. By taking out the seasonal effect, FG assume what is left should display a time trend, if present.

But this is wrong. In fact what remains is exactly the opposite: a detrended representation of the data in each time period! The error, of course, is that the seasonal effect reflects and incorporates the trend, so that when removed from the unadjusted data, the trend disappears. To

see this clearly, suppose that the revenue figures W_{ij} for month $i = 1, \ldots, 10$ in period $j = 0, 1$ displayed a single log-linear time trend with arbitrary deterministic monthly effects denoted by d_i. (We ignore random error, which is here beside the point.) Thus

$$X_{ij} = \ln W_{ij} = a + b(i + 12j) + d_i.$$

Note there is no break assumed in the time trend between the two periods. The "seasonal effect" in Fairley's method is, for month i,

$$\begin{aligned} X_{i\cdot} - X_{\cdot\cdot} &= [a + b(i + 6) + d_i] - [a + b(5.5 + 6) + d.] \\ &= b(i - 5.5) + (d_i - d.), \end{aligned}$$

where we have used simple averages in the log scale. Then the seasonally adjusted figures become the antilogs of

$$X_{ij} - (X_{i\cdot} - X_{\cdot\cdot}) = (a + 5.5b + d.) + 12bj,$$

which takes the form of two perfectly flat lines separated by a gap between the time periods. The appearance of the data values in column 2 of FG's Table 1 may be a result of the defective deseasonalization. To properly deseasonalize a time series, the overall trend must be estimated and subtracted out first, and the seasonal effects calculated from the *residuals,* before subtracting from the unadjusted data. The result will then be a clearer picture of underlying trend.

Summarizing the first problem: to accept the null hypothesis of no trend in the Brinks period and no trend in the CDC period with a dollar gap in between, against the alternative model of positive trend (with or without a gap), may be an error of type II, due to low power, misleading adjustment, or both. On substantive grounds, to conclude that there is no trend in the Brinks period is to assert that the factors contributing to the trend line in our Figure 1 abruptly abated in May 1979. In my view this is not a reasonable inference to draw.

Turning to the second problem, the assumption that the difference between the two periods is primarily due to persistent theft must compete with another explanation that plausibly would account for a higher revenue level in the CDC period. There was an unusual gasoline shortage that prevailed in New York City from May to December 1979 that undoubtedly reduced automobile use and with it probably reduced metered parking. There was odd–even gasoline rationing during the four-month period June–September 1979, and automobile toll bridge and tunnel revenues showed marked drops during the shortage. In addition, there was higher revenue from New York City buses and rapid transit from June to December 1979 compared with CDC's corresponding period in 1980.

The gasoline shortage and rationing appear to have suppressed automobile usage in the Brinks period. There was an additional factor that may have increased automobile usage during the CDC period: an employee strike of the Port Authority Trans-Hudson subway line into New York City from June to August 1980, during which time automobile traffic between New York City and New Jersey increased by 10.7% over the previous year's period.

FG's model assumes a constant percentage increase in revenue from each month in the Brinks period to the corresponding month in the CDC period, and a constant percentage increase in all boroughs. The data do not support the first of these assumptions, and the second seems implausible. Over the two 10-month periods, the percentage increase in average revenue per collection day was 5.9%. However, in the first four-month subperiod (June–September 1979 and 1980, respectively) the percentage increase was 9.9%. This greater percentage increase appears plausibly related to the presence of odd–even gas rationing in the first period. By contrast, in the later six-month period (October 1979–March 1980 and similarly in 1980–1981) the percentage increase was only 3.5%. The change in percentage increase is thus fully consistent with a causal explanation based on the gasoline shortage and the transit strike, and is inconsistent with FG's model. As for the second assumption of their model, I think it unreasonable to assume that the same percentage of receipts were stolen in every borough in every month.

With regard to factors affecting the meter plant itself, FG show in their article that meter price changes were not responsible for period revenue differences. Additional factors unrelated to theft were introduced at trial that are not discussed by FG, however, such as a reduction in the number of parking meters with "revenue-affecting defects" in the CDC period, and the continuation of the program of relocating meters from low-revenue, high-vandalism areas to high-revenue, low-vandalism areas. We have only the testimony of the city's expert on parking meters that *he* could not determine the direction of these biasing effects. This seems surprising, since as a matter of common sense any effect of these changes would be in the direction of increasing revenue. The expert did state that his consulting firm had not made any in-depth studies of New York City's parking meter system in order to make such a determination.

There was another line of evidence concerning Area 1A. This metered parking area in the vicinity of the federal courthouse in lower Manhattan was always collected by city employees and therefore formed a natural "control" area for comparative purposes. The reasoning was that if one found percentage increases and trends in this area comparable to those in the areas collected by Brinks, then an attribution of theft would be less compelling. In fact, in the 10-month period comparison, the average revenue per collection day received by the city from Area 1A increased by 6.9%, an even larger percentage increase than the 5.9% increase in the city overall, excluding Area 1A. A simple linear trend line fitted to the entire period as in Figure 1 showed a highly significant slope coefficient ($P < 0.001$). While numerically smaller than the citywide slope coefficient, the effect is still present and clearly is not attributable to theft. As FG point out, one should be cautious about conclusions drawn from Area 1A since it is a nonrandom sample of the city's meter plant. It does raise serious concern about causal attribution, however, and certainly similar caution should be elicited over the estimate of theft.

On the general issue of comparability, each factor that has been cited above might reasonably be expected to cause a difference in meter revenue between the two comparison periods. Whether they *did* cause a difference in the presumed direction, and to what extent, is largely unknown, and I could not therefore testify that these factors did indeed have the presumed effect. The absence of such testimony was apparently interpreted as a showing of period comparability. This is an error of inference. Given the factors as they have been presented, and given the parking meter expert's failure to quantify the effect of all but the most trivial of these, there must be a strong presumption of noncomparability, because to assert the contrary would require a coincidental cancellation of effects that is difficult to justify. This position is consistent with our understanding that uncontrolled observational studies will, in the presence of a preponderance of biasing factors, produce biased results unless exceptional circumstances prevail.

FG note that a causal factor other than theft should be able to account for the occurrence of a sizable, sudden upward shift across a two-month period that is unique to that period and that occurs in every borough. We have already seen that precisely these effects may arise as an artifact of the seasonal adjustments FG used. Beyond this point, however, it appears that sizable, sudden upward shifts are common in these data. Referring to Figure 1, the inherent month-to-month variation in the data produces several instances of such shifts: see, for example, February–March 1979 during Brinks' tenure, or January–February 1981 in the CDC period. In fact, the latter jump is of greater magnitude than the jump from March to June 1980. It is apparent that the scatter of points in the CDC period is greater than the scatter of points in the Brinks period. Under these circumstances the appearance of a sizable jump across a two-month period by chance alone is not improbable. The larger point to be made here is that we are observing a highly noisy process in which the contribution of theft was one possibly small component. When data are highly regular, for example, in well-controlled comparative trials, the ascertainment of even small effects is relatively easy; when the signal-to-noise ratio is low, then estimation of specific effects becomes inaccurate. Because of systematic period noncomparability and relatively large inherent variability, I believe the statistical data in this case can be used to provide little more than speculative estimates of theft.

What statistically valid techniques might have been employed to provide a reliable estimate of the extent of theft? One possibility would have been proper use of the "salting" technique. In this technique coins that are specially marked with an invisible fluorescent dye are deposited directly into the coin boxes and later counted under ultraviolet light in the returned revenue. This technique was in fact used in the case, but there was dispute over the evidentiary status of many of the saltings. Brinks' principal objection was that "checkbacks" had not been made to guarantee that uncounted marked coins were indeed missing. The method might have been used to advantage had sound statistical procedures been used, including randomized salting across temporal, geographic, and collecting-team strata; data quality control, including checkbacks on all salted meters; and periodic replications to gauge the rate of theft at various times. With these procedures the city might have obtained a quantitative estimate, free of confounding biases, and with the authority of a well-made scientific measurement.

I have criticized two aspects of Fairley and Glen's analysis. First, the methodology used to address the question of trend is defective because the time intervals they study are too narrow to accurately gauge the trend effect, and because their method of seasonal adjustment artifactually produces a pattern which is improperly interpreted as a meaningful signal in the data. Second, the attribution of period revenue differences entirely to theft remains largely speculative due to the presence of uncontrolled systematic biasing factors and substantial month-to-month variability. Reasonable alternative explanations for the revenue differences are presented that are unrelated to theft and that cannot be easily dismissed.

Additional References

Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association,* **74**(368), 829–836.

Fairley, W. B. (1979a). Evaluating the "small" probability of a catastrophic accident from the marine transportation of liquefied natural gas. In W. B. Fairley and F. Mosteller (eds.), *Statistics and Public Policy,* Reading, Mass.: Addison-Wesley, pp. 331–353.

Fairley, W. B. (1979b). Accidents on Route 2: two-way structures for data. In W. B. Fairley and F. Mosteller (eds.), *Statistics and Public Policy,* Reading, Mass.: Addison-Wesley, pp. 24–33.

Rejoinder

William B. Fairley / Jeffrey E. Glen

Setting aside several technical issues, which we address below, we find that Levin's basic theme is that factors other than theft "might have" caused most of the observed increase in revenues, so that "the contribution of theft was one *possibly* small component" (italics ours).

This point, while true, is not responsive to the determination of the damage amount that the jury was asked to make. As Judge Edward Weinfeld said in his District Court opinion, concurred in by the U.S. Appeals Court, the question of damages before the court was not what a precise and accurate estimate of the damages was, but rather, what was the *most reasonable* estimate of damages that could be made *given* that it *had already been established* that "Brink's employees engaged over an extended period in concerted action in the pilferage of meter coin deposits" (546 F.Supp. 403 (1982), at 407). Indeed, reading Levin's comment one might not have guessed that substantial investigative evidence had been introduced in court to support an estimate of $1,400,000 theft over 10 months and even more. Judge Weinfeld's opinion noted that

> The evidence offered by the City included surveillances of suspected employees who were observed to have deviated from their assigned routes carrying heavy bags from the transport vehicles into an apartment house where one of the employees resided; the results of a salting test which showed substantial unaccounted for coins at various dates and involving different teams of collectors; videotaped surveillances of occasions prior to and on April 9, 1980, when a number of employees were arrested, which showed them acting in what appeared to be a concerted, clandestine, and secretive manner with respect to heavy bags allegedly containing meter collection proceeds; the arrest on April 9, 1980 of the group of employees who had in their possession almost $5,000 in coins; and the conviction of a number of the arrested employees either upon pleas of guilty or by a jury verdict for stealing meter coin deposits. (at 407)

Therefore, Judge Weinfeld observed

> To argue, as Brink's does . . . that there was a complete absence of evidence to support the verdict on this issue . . . is to ignore the force of the totality of evidence and to adopt an ostrich-like pose. (at 407–408)

In seeking completely comparable periods demonstrably "free of confounding biases," Levin erects a pristine standard for statistical control. While such a goal is our ideal, the real world must sometimes be studied by observational studies where control is less than perfect. In an earlier paper, one of the present authors (Fairley, 1978, pp. 794–795) said:

> The limitations of purely observational data and the parallel limitations on the capacity of statistical methods to control for important confounding effects have not been widely understood.
>
> . . . Yet we should avoid the untenable position that experiments alone can support cause-and-effect inference. Much scientific advance, not to mention practical knowledge, has been based on observational studies—and the "observational" sciences of astronomy and geology are not alone in this . . .

In the same vein Hoaglin *et al.* (1982, p. 74) note:

> Comparative observational studies may offer the only means of ever collecting any data at all about a treatment. Some events, such as earthquakes, are impossible to deliver as designed treatments, and certain treatments are unethical, as in many medical and social investigations. In these extremes, investigators must do their best with data from observational studies.

Levin offers no alternative estimate of theft in the fact of the inherent difficulties. All that he offers is a reference to a possible systematic use of "salting," which, *had* it been used, would have provided a quantitative estimate free of confounding biases. Wonderful! But it had not been done, and so the court had to deal in the real world with the data and evidence at hand.

The root problem is that Levin is importing into a court decision-making context a set of standards and conventions for "evidence" that are *not* the same as those that are required by the applicable law. Specifically, Levin is willing to remain in a *dubitante* position and not offer any alternative estimate that he believes is more reasonable. He clearly doesn't want an inference threatened by real-world possibilities of the error of confounding bias. None of us wants such threats, but the court plainly cannot remain in a *dubitante* position. It must decide. The ruling legal doctrine in a case like this, where the *existence* of theft was not in dispute, and where the jury *had* determined tortious conduct on the defendant's part, is that (546 F.Supp. 403 (1982), at 410, and quoting *Bigelow v. RKO Radio Pictures, Inc.* (327 U.S. 251 (1946)) "when a defendant's tortious conduct is of a nature that precludes precise ascertainment of damages, the jury may make a 'just and reasonable estimate of the damage based on relevant data.' "

As support for the *dubitante* position Levin claims that the estimate of theft that we advanced was "little more than speculative." In the relevant case law "speculation" is likened to a "guess" and is not viewed as legally sufficient evidence to support an estimate of damages. However, the statistical and subject matter study of parking meter theft done by the experts testifying for the city was substantial. It was clearly not a "guess" and therefore it was not mere "speculation." In fact, although no observational study could demonstrate that *no* other explanation than theft was possible, the study did rule out or diminish important alternative explanations. No *more* reasonable estimate of the amount stolen was ever proffered.

Were every situation in which a difficult estimate had to be made deemed a "guess" and no estimate were allowed, then defendants in these situations would go scot-free, despite their admitted wrongful conduct. Better an estimate based on a deliberate and careful attempt to provide the most reasonable figure permitted by the evidence than no estimate at all. At least that is what the law requires, as the District Court and the Appeals Court agreed. Levin's requirements for an estimate differ from the law's, and clearly it is the law's requirements that apply.

We turn now to an examination of Levin's specific criticisms of the statistical estimates of theft described in our paper.

Levin says in the first half of his comment that "trend" could account for the increase in revenues in the CDC period over the Brink's period and that we overlooked "this" trend.

We do not agree that there is a "trend." First, if we look, not at monthly revenues in the period from April 1978 through March 1981 selected by Levin and graphed in his Figure 1, but rather at the longer 48-month period May 1977 through March 1981 as provided in a Brink's exhibit, the picture presented is very different. Figure 1 in this rejoinder graphs the monthly revenues for the longer period. The points in the early period are heading down, not up. Thus the longer period displayed in our Figure 1 does not convey the same impression of "a general upward trend" as does Levin's Figure 1.

However, which figure to use is not the point. While we agree that looking at historical context can be useful, one *can* and should here discuss the question of whether one observes, statistically, a trend within *any* specified interval of time. In particular, we can ask if, within each of the two 10-month periods on either side of the arrest of Brink's employees, there is evidence of trend. The answer, graphically or by test, is clearly no. This answer, furthermore, from the point of view of inference, is not tainted by selection bias because the two 10-month periods were selected before looking at the time series, not after. (Our Figure 2 in this rejoinder gives city-wide revenues per meter-day over the two periods.

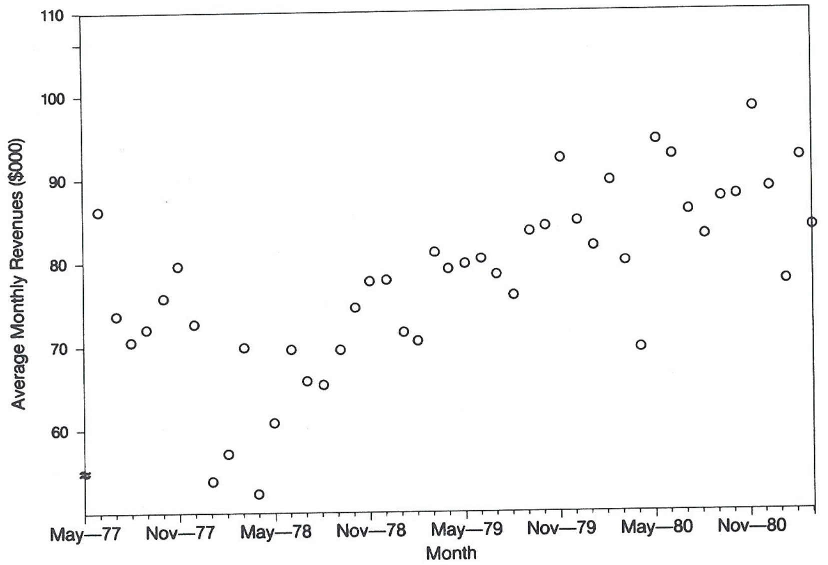

Figure 1. Revenues per collection day.

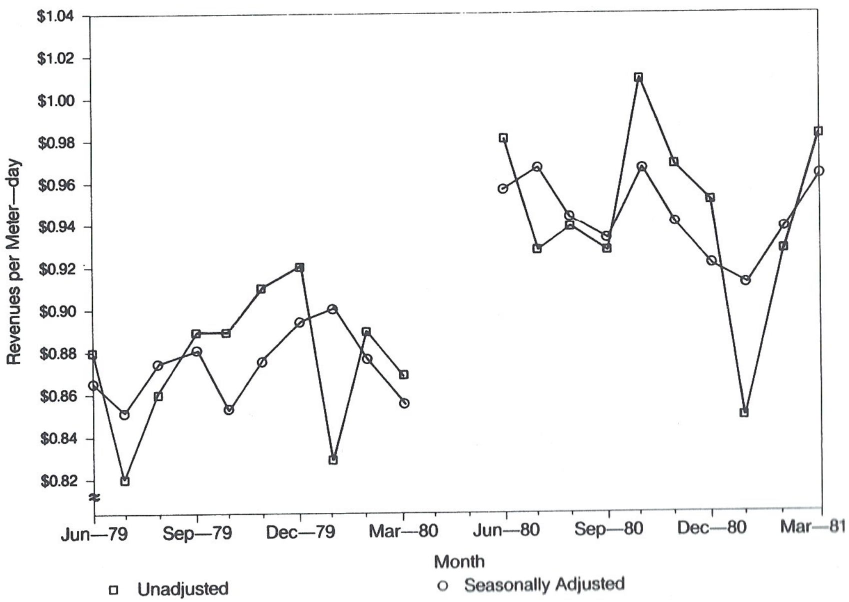

Figure 2. Revenues per meter-day, city-wide.

The absence of clear trend within the periods is also true for each of the five boroughs separately).

Thus, looking at the broadest historical context made available through data at the trial does *not* lead to a conclusion that there was a trend that *would* explain a difference in revenues before and after the arrests of Brink's employees. The possibility that factors other than theft, whether manifesting themselves in a secular trend or not, *could* explain the difference remains, but is not bolstered by any demonstration of a statistical trend purportedly offered by Levin's Figure 1.

Third, demonstration of a *statistically* observed trend, even assuming this were accomplished, is much less weighty as a counter-causal explanation of an observed difference than a demonstration that *specifically identified* causal factors are operating over the 22-month period June 1979–March 1981 to produce that difference. Levin did not establish any such specific factors. Thus, although there is no reason to believe that the total effect of all factors influencing revenues was constant or exactly constant during the 22-month period, the direction of effect of factors other than theft is unknown. That is, no known nontheft factors were shown to have caused the observed increase in revenues. While Levin advances four specific factors that he believes "might reasonably be expected to cause a difference in meter revenue between the two comparison revenues," he acknowledges that, "Whether they *did* cause a difference in the presumed direction, and to what extent, is largely unknown, and I could not, therefore, testify that these factors did indeed have the presumed effect." We are unconvinced that Levin's specific factors are even plausible explanations, as we discuss below.

Fourth, Levin says that the power of our broken-line model "test" for trend was low and therefore if there was a trend in the 22-month period, we would not have detected it. Here again, Levin has a tool for another job, but not for the job at hand. Levin's tool is a test of a single "trend" line fitted to the 36-month period April 1978–March 1981. This test produces a "highly significant slope coefficient ($p < 0.001$)" (Levin, this volume). As we observed in our chapter, fitting such a line can give no estimate of the effect of the *known* causal factor of theft *at* the transition between the Brink's and CDC periods. Such an estimate, however, is what the court required. Further, finding a significant slope coefficient for the line over 36 months, for reasons given above, does not establish that factors other than theft, whether identified or not, caused the observed change in revenues over the 22-month period.

By contrast, our tools, graphical exploratory analysis and a least-squares fit of a broken-line model to the monthly revenues (per meter-day) over the 22-month period, were suited to observe and to estimate a jump, if there was any, and also at the same time to observe and to estimate a statistical trend over the whole period, if there was any. The data do *not* show, as we explained in our chapter, evidence of trend over the entire 22-month period that *cannot be explained* and is not best explained by a jump in the middle. Thus, the claim that our "test" has low power is not a real criticism unless there is another test clearly suited to our job that has interestingly higher power. Levin has not exhibited such a test.

Fifth, Levin says that we "appear" to have made a methodological error in that our method of seasonal adjustment of the monthly revenue data appears not to be appropriate for the data and might have artifactually removed a trend actually present in the data. The method Levin believes we "appear" to have used we did not use. As we stated the seasonal adjustments "were made by a model for the data, as discussed below." Since the linear model described determined the seasonal adjustment, the error that Levin refers to was not committed, because all the linear model parameters, including the season–borough parameters and the intercontractor jump parameter, were determined simultaneously by least squares. We did then, and appropriately so, define an average model-predicted revenue per meter-day in each month for each contractor [defined for each month as the antilog of the sum over boroughs of the model-predicted monthly revenues per meter-day (in log scale) for each contractor and divided by the meter-days pooled over boroughs for the month]. These average model-predicted revenues per meter-day for each month were then used in the standard way to *display* seasonally adjusted revenues by month for each contractor by dividing them into the observed monthly revenues and multiplying by the 10-month period average. Note that this standard display procedure was adopted for a descriptive use long after the existence of a linear trend of the kind Levin postulates was explicitly rejected. ("Graphs of average monthly revenues delivered per meter-day of operation for the entire city and for each of the five boroughs, either without seasonal adjustments or with monthly seasonal adjustments, did not indicate a trend over time, but rather indicated level revenues per meter-day within each 10-month period." Fitted piecewise-linear regressions *also* did not indicate a trend such as Levin postulates).

Finally, Figure 1 in our chapter displays residuals from a trend line fitted to seasonally adjusted revenues per meter-day. The purpose of the fitting done for Figure 1 was not to estimate a trend nor to establish that no trend existed, but to *illustrate* for the reader through a residuals display how city-wide data are not well described by a single linear trend line. The *same* conclusion emerges from examining the residuals from a line fitted to unadjusted data, and the same conclusion emerges for every borough considered separately.

Even had it been true that our choice of seasonal adjustment was inappropriate, the criticism would have no weight because our results would be only trivially changed by using unadjusted data. Figure 2 gives the unadjusted monthly revenues per meter-day for the 22-month period for the entire city. No model is indicated for this data different from that for the seasonally adjusted data. This is also true for each separate borough. None of our conclusions changes. The purpose of the seasonal adjustment was to make a fair comparison between revenues per meter-day in the last month or months of Brink's, beginning March 1980, and the first month or months of CDC, beginning June 1980. For this purpose, the adjustment was correct and appropriate.

The second half of Levin's critique is a discussion of four factors that "might have" caused the observed increase in revenues. Three of these four factors can be immediately dismissed because, even were their existence and importance established, they do not explain the change in level that is observed on either side of the transition between Brink's and CDC. This change in level, measured by the estimated intercontractor parameter in the linear model, can also be described by the difference between one-month, two-month, three-month, etc., averages of revenues per meter-day on either side of the transition time between Brink's and CDC.

The first two of Levin's factors are (a) a presumed reduction in meters with revenue-affecting defects in the CDC period and (b) a presumed program of relocation of meters from low- to high-revenue-generation areas. Brink's did not present an expert who could establish, quantitatively or otherwise, that these presumed programs could be expected to produce a difference in revenues per meter-day of the size observed. Furthermore, these effects, if any, were presumably operating, if at all, over the entire 22-month period and would not then have a discontinuous effect at the transition.

The third factor cited is the gas shortage in the latter part of 1979, which involved odd–even rationing in the period June–September 1979. The effect of this factor on revenues per meter-day was debated at the trial, and in any case it would not explain the change in level at the transition, which occurred several months later.

The fourth factor cited was the strike on the Port Authority Trans-Hudson (PATH) trains between New Jersey and Manhattan in June–August 1980 and an increase in automobile traffic between New Jersey and New York City in that period, presumed to be associated with the strike. The parking meter expert, Laurence Donoghue, could not tell what the effect of these events would be on meter revenue. He noted that an increase in vehicles going into the often saturated metered parking areas of Manhattan could very well *decrease*, not increase, revenues if these vehicles parked for several hours in poorly enforced areas and did not comply as well as the customary parkers in these areas. Also, these vehicles might well have used meters more efficiently than in-and-out shoppers did, thereby decreasing revenues.

It is not possible in these pages to reproduce the discussion of specific causal factors that occupied considerable time at the trial, so we do not believe that any reader should be convinced one way or the other about the likely direction of effect on revenues of the four factors cited. However, at trial the city did present a considerable amount of detailed factual research on specific causal factors, whereas Brink's by and large presented hypothetical possibilities of effects, or rested, in Levin's words, on "common sense" as the arbiter of differences over the direction of effect of these factors on revenues. In sum, we were not, and the jury and judge evidently were not, moved to believe that causal factors other than theft were, on balance, any more likely to cause an increase than a decrease in revenues between the two comparison periods. Furthermore, the conclusion as just stated is the relevant one for determining a reasonable estimate of damages. It was *not* incumbent on the city to show, as Levin believes, that during the 22-month period there was "a coincidental cancellation of effects [of causal factors other than theft]." The legal burden, once tortious conduct had been established, was on Brink's to show that other factors probably did account for the jump that was observed at the transition when the contractors changed.

We close by observing that, while Levin is impressed by the difficulties of controlling for causal factors other than theft, his own analysis of revenues fails to control for two important factors that could be controlled, namely, the number of meters and the number of days they were in operation. Contrary to his assertion (Levin, this volume), we found that the analysis of revenues per meter-day instead of revenues uncontrolled for meter-days in operation was a useful device. To illustrate its use, we note that the jump in unadjusted monthly revenues from \$77,500 in January 1981 to \$92,000 in February 1981, which Levin cites as an example of a jump in revenues larger than that observed at the transition between Brink's and CDC, is reduced, in meter-day units, to a jump of \$0.85 per meter-day in January 1981 to \$0.93 per meter-day in February 1981, when meter-days are controlled for. The latter difference of \$0.08 now is *smaller* than the jump in revenues per meter-day of \$0.11 observed at the transition. This fact is not essential to the conclusions we reached, but illustrates that the conclusions reached in our study were based on a good deal more than "speculation."

Additional References

Fairley, W. B. (1978). Statistics and public policy. In W. H. Kruskal and J. M. Tanur (eds.), *International Encyclopedia of Statistics,* Vol. 2, New York: Free Press.

Hoaglin, D. C., Light, R. J., McPeek, B., Mosteller, F., and Stoto, M. A. (1982). *Data for Decisions: Information Strategies for Policymakers,* Lanham, Md.: Abt Associates.