Design Considerations for Experiments in Major Regulatory Change

Rolf R. Schmitt, National Transportation Policy Study Commission, Washington, D.C.

Major changes in the regulation of transportation have been frequently proposed and occasionally tried on a nationwide scale. Beyond theoretical and poorly substantiated evidence, the effects of proposed and trial changes are generally unknown or inadequately understood. This paper argues that field experiments provide the most sound basis for evaluating major regulatory changes. It also examines the conditions that must be met for experimentation to be effective. After briefly surveying the variety of experimental designs used in nontransportation research, criteria are developed for selecting designs to evaluate regulatory changes. The paper concludes with a recommended evaluation strategy and some caveats.

Major changes in the regulation of transportation are being proposed with increasing frequency. These changes range from modifications to the criteria for setting rates to total deregulation. Several changes have been implemented on a trial or permanent basis, yet surprisingly little validated information on the impacts of these changes exists to guide future regulatory policy.

Existing information on major regulatory change comes from two sources: (a) systemwide changes in federal policy toward specific modes and (b) comparisons between differing federal and state regulatory environments. Information on systemwide policy changes is easily distorted by many concurrent events and extraneous factors that affect measured outcomes but are not germane to the policy being evaluated. Information is also distorted by unanticipated or indirect impacts of the policy that are not measured. These problems are exacerbated by national data-collection efforts, which currently tend to be myopic, disjointed, or inaccessible (1). Knowledge gained from comparative studies of federal and state regulatory policies is similarly degraded by infrequently considered variations in local conditions that affect the performance and impacts of transportation. The policymaker is thus left with theoretical speculation and poorly substantiated empirical evidence to evaluate proposed regulatory changes.

When faced with the possibility of adverse consequences, a frequently advocated solution to inadequate information is to experiment. Although field experiments have been used to evaluate proposed housing, income maintenance, and other social programs, this method has not been applied to transportation subjects other than structural engineering, vehicle safety, and small-scale traffic control. Carefully controlled field experiments in major regulatory change are virtually nonexistent.

The utility and design of field experiments for evaluating major regulatory change in transportation are explored in this paper. The reasons for experimentation and the conditions that must be met for field experiments to be useful are examined first. The wide range of experimental designs developed in other fields and applicable to transportation are surveyed. Criteria are developed for selecting appropriate designs. The paper concludes with a recommended evaluation strategy and some caveats.

FIELD EXPERIMENTS AS AN EVALUATION TOOL

Evaluations of transportation policy are largely molded by their sources of information, which include predictive models, case studies, and field experiments. A currently popular source of information is the predictive model, which includes the simulation experiment (2). Predictive models are useful for evaluating proposed actions with potentially traumatic and irrevocable consequences, and simulation experiments can indicate the sensitivity of the model's estimates to its assumptions and to stochastic events; however, this approach ultimately requires faith in the model's theoretical underpinnings and in the temporal and geographic stability of its parameters. Case studies are the second and most common source of information on the effects of regulatory change. They include narrative histories, dockets filed with regulatory commissions, and demonstration projects. Although based on observations of actual (rather than predicted or simulated) events, case studies do not comprehensively analyze the concurrent events and extraneous factors that can distort empirical observations. This is not true of field experiments, which are the third and least used source of information on the effects of regulatory change. In evaluation research, field experiments are the measurement of the effects of a policy or program implemented in a physically or analytically controlled environment. Unlike case studies (particularly demonstration projects), field experiments are designed explicitly to cancel out or reveal extraneous factors so that the measured effects can be attributed with confidence to the policy or program being evaluated.

Predictive models and case studies have been considered to be acceptable sources of information on regulatory change largely because many transportation analysts are content with reliable measurements. This contentment is myopic because reliability is a necessary but not sufficient condition to establish validity. Reliable measures address unknown and nebulous phenomena in a reasonable and uniform manner; however, the results may misrepresent actual outcomes either by incorrectly labeling actual elements of change or by being sensitive to extraneous factors. To assure valid results, information must be collected and analyzed in a way that certifies proper labeling of measures and that reveals or controls for extraneous factors. Field experiments are explicitly designed to do this and thus provide superior evidence that a policy does or does not work and indicate how the policy might work better.

Like any evaluation technique, field experiments are not always appropriate and occasionally may be misused. Even the term is often used incorrectly as a label for demonstration projects that are at best primitive forms of experimentation and lack adequate controls for extraneous influences. More importantly, actual field experiments can be used to postpone major reforms, to avoid responsibility, to create public relations cannon fodder, to divert attention from serious issues, or merely to fulfill requirements (3). Even if the policymaker desires a careful evaluation of a regulatory change, field experiments require adequate time, financial resources, and professional talent to yield useful findings.

Beyond requisite time and financial support, the feasibility and appropriateness of field experiments depend on the following considerations (4):

1. Political considerations: Has the policy been irrevocably committed? Has adequate flexibility for post-evaluation alterations of the policy been allowed? Is the cost of delaying full-scale implementation less than the cost of a full-scale faux pas?
2. Ethical considerations: Are potentially adverse effects of the experiment harmful to individuals? Is experimentation just an excuse for delaying action or for distributing actions unevenly?
3. Technical considerations: Can substantive questions asked in the experiment be answered without resorting solely to "black box" explanations?
4. Administrative-managerial considerations: Can a working, knowledgeable team be gathered or trained to execute the study? Can they develop credibility with their clients (the policymakers)?

If any questions are answered "yes" in the first two considerations and "no" in the remainder, then experimentation will most likely be a fruitless or counterproductive exercise that should probably not be attempted.

Although political and related conditions vary among types and timing of proposals for regulatory change, there is nothing intrinsic to transportation that precludes the use of field experiments to evaluate major regulatory policies. The difficult task is to design the field experiment in a way that creates both timely and generalizable results. Fortunately, the transportation analyst can draw on past experience with the design of field experiments for social psychology, agriculture, and other areas. To properly utilize this experience, however, the analyst must master at least some of the terminology of these fields and understand the differences between evaluation and transportation. In particular, the spatial characteristics of transportation can affect the selection of experimental designs used previously in social policy research.

EXPERIMENTAL DESIGNS USED IN PAST EVALUATIONS

Most of the experimental designs that have been developed to evaluate social policies and programs can be applied to the evaluation of major regulatory changes in transportation. These designs have been inventoried in detail and classified by Campbell and his associates (5-7), whose works were placed in the transportation context by Charles River Associates (8) and Schmitt (9). The 28 experimental designs are grouped as follows:

1. True experimental designs: pretest-posttest control group design, posttest-only control group design, Solomon four-group design;
2. Quasi-experimental designs: (a) classic one-treatment before-and-after designs—nonequivalent control group design, nonequivalent dependent variable design; (b) separate-sample one-treatment before-and-after designs: separate-sample pretest-posttest design, multiple separate-sample pretest-posttest design, separate-sample two-pretest-one posttest design, separate-sample pretest-inclusive-posttest design, separate-sample pretest-posttest control group design, expanded separate-sample pretest-posttest control group design; (c) basic time-series designs: interrupted time-series design, interrupted time-series design with nonequivalent dependent variables, interrupted time-series design with nonequivalent control group, interrupted time-series design with switching replications; (d) one-group multiple-treatment designs: repeated treatment design, removed-treatments pretest-posttest design, equivalent time samples design, equivalent materials sample design; (e) multiple-group multiple-treatment designs: recursive separate-sample pretest-posttest design, reverse-treatment nonequivalent control group design, counterbalanced design, institutional cycle design; and (f) regression-correlation designs: regression discontinuity analysis design, quantified multiple control group posttest-only design, posttest-only design with predicted higher-order interactions, path analysis correlation design, cross-lagged panel correlation design.

Each of the experimental designs is a specific plan that includes procedures for (a) selecting the group to be exposed to the policy change (the "treatment group") and, for most designs, selecting a "control group" that is not exposed to the change for subsequent comparisons; (b) implementing the policy change (the "treatment"); and (c) making observations (called "pretests" when made before implementation of the change and "posttests" when made afterwards).

Campbell (10) has shown that an experiment can be designed after the change is implemented, provided that relevant conditions were adequately monitored and that the policy implementation was distinct (to act as a treatment). Whether devised before or after the regulatory policy is changed, the experimental design is a process of monitoring and analysis, carefully tailored around the implemented policy to counter problems of research validity.

The 28 designs are classed as true or quasi experiments. True experimental designs are generally the more powerful and desirable because the random assignment of individuals, organizations, or areal units to treatment and control groups assures that differences between the groups, which are not relevant to the policy change, can be statistically removed. Measured changes are then attributable solely to the policy being evaluated. Unfortunately, the ability to randomly select individuals or other units for exposure to a new policy is limited in transportation research. The random exclusion of individuals or other units from the affected transportation service is generally less practical for political, ethical, and geographic reasons (9). (How can potential patrons along a bus line be randomly allowed to use or not use the service?) Equivalence between treatment and control groups must therefore be assured by other means. These varied means that do not use random assignment techniques are quasi-experimental designs. (These designs may use randomization techniques for drawing sample observations from—but not for assigning membership to—treatment and control groups.) Both true and quasi-experimental designs are explained and their strengths and weaknesses documented in several inventories (5-9).

CRITERIA FOR SELECTING AN EXPERIMENTAL DESIGN

No one experimental design is a panacea for the problems of evaluation research, nor are all designs appropriate for the evaluation of a particular regulatory change. Because many variations of and additions to the list of designs given here are possible, design specifications for evaluating a particular regulatory change can and should be tailor-made to deal with the most important problems at hand. This flexibility in coping with diverse situations makes evaluation "cookbooks" impractical or inappropriate (11), so the transportation analyst must rely on experience, the previously cited inventories, and four basic criteria to select or modify an experimental design.

Criteria for the selection or modification of an experimental design to evaluate a regulatory change include the control of validity threats, timeliness, analytical complexity, availability of data-collection instruments, and applicability to recursive innovation development. These interacting criteria simultaneously establish the costs and constraints under which the evaluator selects an experimental design to provide the most credible findings possible.

Control of Validity Threats

The central purpose of an experimental design (and the most complex criterion for its selection) is to expose or eliminate validity threats so that the measured outcomes of a regulatory change can be attributed with confidence solely to the policy being evaluated. Validity threats are the generic labels for the concurrent events and extraneous factors to which the evaluation measures and their interpretation may be sensitive. These threats have been classified by four conditions for validating measured outcomes of an experimental policy and the relevance of those measurements to other situations (6).

One condition is that the action taken and the condition to be ameliorated actually covary. Covariance is not always obvious when sampling is involved. "Statistics are used for testing whether there is covariation... (and) function as gatekeepers. Unfortunately, there are fallible gatekeepers even when they are properly used, and they fail to detect both true and false patterns of covariation" (6). Problems associated with correctly determining covariation are called threats to statistical validity.

Another condition is that the field experiment itself is not a source of bias. Did the observed groups react unnaturally because they knew that they were part of an experiment? Did changes in the data-collection procedure—rather than changes in the condition being monitored—create measured differences? As proposed in the last century (12), questions such as these are answered by considering rival hypotheses to establish internal validity.

In social settings, change is rarely measurable directly. The data must be interpreted, generally with the aid of a theoretical construct or explanatory model. Problems of interpretation are called threats to construct validity, and these should be understood as threats to correct labeling of the cause and effect operations in abstract terms that come from linguistic usage or formal theory. Actually, the problem of construct validity is broader than this and obviously applies to attempts to label any aspect of an experiment including the nature of the setting, the nature of persons participating, and so forth (6).

For purposes of evaluating regulatory change and other transportation policies, a field experiment has little value if the results cannot be generalized beyond the specific time and place of the policy change. The conditions for generalizability are addressed as threats to external validity.

Listed below are 34 validity threats that have been identified from experience in evaluating experimental programs in education, criminal justice, and industrial management (6):

1. Threats to statistical validity: statistical power, fishing and error rate problem, reliability of measures, reliability of treatment implementation, random irrelevancies in the experimental setting, random heterogeneity of respondents;
2. Threats to internal validity: history, local history, maturation, testing, instrumentation, statistical regression, selection, mortality, interaction with selection, ambiguity about the direction of causality, diffusion or imitation of treatment, compensatory equalization of treatment, compensatory rivalry, resentful demoralization of respondents receiving less desirable treatment;
3. Threats to construct validity: inadequate preoperational explication of constructs, mono-operation bias, mono-method bias, hypothesis-guessing within experimental conditions, evaluation apprehension, experimenter expectancies, confounding levels of constructs and constructs, generalizing across time; and
4. Threats to external validity: interaction of the treatment and treatments, interaction of the treatment and testing, interaction of the treatment and selection, interaction of the treatment and setting, interaction of the treatment and history, and generalizing across effect constructs.

Very little of this experience includes research in which geographic space plays a significant role. Geographic space creates an additional set of validity threats, such as those listed below, particularly when observations are made on areal units (e.g., counties) rather than individuals or organizations:

1. Boundary distortions—overextension, truncation;
2. Partition distortions—spurious location and diffusion, spatial autocorrelation, excessive heterogeneity within zones, density bias;
3. Scale distortions;
4. Interaction of scale and constructs;
5. Interaction of scale and statistical validity;
6. Generalizability across scales;
7. Interaction of space and time; and
8. Confusion of spatial and aspatial issues.

As defined and explained elsewhere (13), these threats are commonly relevant to transportation research.

The types of validity threats and the degree of their control vary among experimental designs. Selection is largely a problem of matching a design's strengths and weaknesses with the threats that are most important to the policy or program being evaluated. For example, time-series designs are very useful for evaluating policy changes with slowly consummated effects because these designs control the internal validity threat of history rather well. The threat of history is a label for the likelihood that an event occurred in between measurements that affected the conditions being monitored but had nothing to do with the implemented policy. The potential for this threat obviously increases with greater lengths of time between observations. To evaluate major policy changes such as the removal of railroad exit restrictions, substantial lengths of time are necessary for monitoring to precede anticipatory reactions and to capture long-term adjustments; therefore, time series are most appropriate. Policy changes entailing shorter-term adjustments, such as the increased enforcement of safety regulations affecting intercity truck drivers, can be evaluated with more timely and less complicated before-and-after designs. These designs do not control for the validity threat of history as well as do time-series designs, but this threat is far less important because the periods between observations can be short.

Beyond the nature of the policy and its effects, the importance of a design's inherent strengths and weaknesses is largely determined by the evaluation's purpose. Designs that emphasize control of threats to external validity, for example, are very desirable for field experiments in federal regulations attempting to provide generalized information for a nationwide spectrum of affected parties and environments. This emphasis is far less important for a locally sponsored effort to solve a local problem. The former could include a regional experiment in changing federal entry restrictions on common carrier trucking; the latter could include a municipal government's experimental change in awarding taxicab medallions.

Although the relationships among experimental designs and most validity threats are well documented, the control of threats inherent to geographic space are little studied and potentially the most difficult to control. These threats are particularly important in the present context because experiments in regulatory change range in scale from local jurisdictions to the entire globe.

One effect of geographic space on the selection of experimental designs has already been cited: geographic proximity generally precludes the use of true experimental designs to evaluate network-based or areawide transportation services. It is rarely possible to allow randomly selected individuals or organizations to use a transportation service while excluding others.

Of the quasi-experimental designs, those that use either separate treatment and control groups or use multiple treatments with separate groups are particularly susceptible to validity threats related to geographic space. If groups are defined by areal units, these designs require the units to be far enough apart to eliminate social or economic contact; otherwise, the policy change or knowledge of its existence may diffuse to the control group, which would no longer provide a valid base-line comparison. On the other hand, greater spatial separation increases the possibility that local differences along social, economic, and political dimensions may distort the comparisons. For example, deregulation of a particular commodity by the federal government could be evaluated by implementing the policy change only within one region and comparing any subsequent differences in ton-kilometers between counties in the treatment (deregulated) region and the control (unaffected) region. Because the treatment and control counties would probably be in different states, comparisons would be distorted by different state regulations (e.g., truck size and weight restrictions) and by other factors that are extraneous to the removal of the rate restriction. Either the effects of these factors must be removed analytically, or the policymaker should be thoroughly warned of their potential existence.

The designs that suffer least from threats related to geographic space are those that allow all groups to be exposed to the policy change. In these designs, diffusion of the treatment (the policy change and its effects) to a control group is irrelevant. The difficulties in finding separated but similar locales is thus eliminated. For these reasons, the nonequivalent dependent variable design; most of the separate-sample, one-treatment, before-and-after designs; the one-group, interrupted time-series designs; and the one-group, multiple treatment designs are particularly useful for transportation research.

Timeliness

Although basic research can be afforded the luxury of long-term data collection and analysis, the evaluation of transportation services that are already on the street requires more immediate results. This criterion may preclude time-series and cross-lagged panel designs for short-term evaluations of just-implemented policies. Any of the designs that use more than one preimplementation measurement, such as the separate-sample, pretest-posttest design with two before measures, may also be precluded if the regulatory change is implemented before the pretests can be made.

Analytical Complexity

The criterion of analytical complexity is primarily an issue of interpretability by policymakers who are not technically disposed. As in the case of large-scale predictive models critiqued by Lee (14), Carver (15) notes that the credibility of an evaluation is diminished if its measures or design cannot be explained in lay terms. Regression-correlation designs are particularly difficult to portray simply, other than as a black box, and will probably be viewed with more suspicion than a simpler design.

Complexity is also a problem for the analyst if large quantities of data must be processed without computer assistance. Both the probability of error and the requisite staff hours are increased by increasingly complicated designs.

Although complexity is an important criterion, it is the least critical. If the other criteria are reasonably met, the selection of a design can finally be resolved by Occam's razor: the least complicated, viable design to implement and explain is the best.

Data-Collection Instruments

In contrast to the criterion of analytical complexity, the availability of data-collection instruments is crucial to the range of experimental designs that can be used to evaluate a particular regulatory change. These instruments, such as personal interviews, mailback questionnaires, tabulations of waybill entries, and mechanically recorded field observations, involve varying degrees of accuracy, respondent reactivity (i.e., falsifying reports), requisite personnel time and expertise, set-up time, processing capability, and monetary expense. Each attribute is magnified by the sampling rate (16), and the instrument's use may be constrained by public reactions to invasion of privacy, reporting burdens, and similar real or perceived issues that should be addressed by a comprehensive federal information policy. In general, the evaluator should select the most cost-effective sampling rate and the most accurate available data-collection instrument that is affordable, and then select a design from among those that the resulting data base will support.

Applicability to Recursive Innovation Development

Because policymakers are rarely omniscient when devising innovative policies, it is generally desirable to apply experimental designs in a recursive process of policy implementation, evaluation, policy refinement, reevaluation, and so on (9). The multiple-treatment designs are particularly well suited for this approach, although reusing one-treatment designs is also possible. If space and population size permit, the application of one-treatment designs in different locales for each policy adaptation is desirable. This alternative lessens the threats of patron reactivity and testing-induced change that are inherent to the multiple-treatment designs.

RECOMMENDED EVALUATION STRATEGY AND SOME CAVEATS

Given the speculative nature of much evidence on the effects of major regulatory change, it is important to learn as much as possible from any policy action, be it a small-scale experiment or a systemwide trial. When the effects are potentially adverse and widespread, the small-scale field experiment is preferable because it allows maximum control of validity threats and minimizes the consequences of inappropriate policy changes; however, major regulatory changes are occurring and the knowledge that could be gained should not be lost or subverted. In order to consistently improve the information base for future regulatory policymaking, the following evaluation strategy is proposed.

Whenever the political, ethical, technical, and administrative-managerial considerations indicate that experimentation is viable, proposed regulatory changes should be evaluated and refined through a four-step process:

1. Objectives of the policy change (treatment) are defined as measurable conditions.
2. The policy change is implemented under a carefully designed monitoring procedure that may or may not be initiated prior to the treatment, depending on the specific experimental design used. (Implementation may be in limited areas or nationwide, which will affect the design's selection.)
3. Controls on and results of the monitoring procedure are analyzed to determine whether the measured impacts are attributable solely to the policy change.
4. If the measured impacts are considered desirable, modifications of the new policy are implemented with concurrent experimentation until the objectives are maximized within the constraints of available resources. If the measured impacts are considered to be undesirable, policy objectives are reviewed to determine whether they were properly defined. New or modified objectives, policies, or both, are developed, and the process restarts at Step 2.

This process should be reiterated until doubts about the policy's outcomes are removed and the policy has itself evolved with the experience to provide the most desirable transportation service.

When political or other factors require immediate systemwide regulatory change, field experiments need not be abandoned as a source of validated information. If steps are taken to collect adequate data beforehand, then an experiment can be designed after the new policy is implemented. Although the results of a post-hoc experiment will probably be less robust than their premeditated counterparts, such results are better than the speculative hypothesizing and poorly validated evidence that are now commonplace.

Current arguments concerning the effects of transatlantic air travel are a case in point. Simple before-and-after comparisons of ridership levels neither confidently explain changes in these levels nor suggest directions for further regulatory change. An interrupted time-series design with nonequivalent dependent variables could control for the concurrent effects of changing seat capacity, leisure time, levels of disposable income, and similar factors that are not components of the new policy. It may turn out, for example, that diversions from other carriers and skyjacking frequency explain many changes in ridership. If such a finding were reasonably validated, then a pricing policy might be supplanted by a policy of intensified countermeasures against skyjacking to effect desired ridership changes.

Since regulatory changes are often imminent or already in effect by the time their evaluations have been mandated and financed, the maintenance of a comprehensive, ongoing transportation data base is critical. Without adequate data on transportation demand, stocks and flows, performance, and impacts, neither public nor private evaluators will be able to learn fully from the regulatory changes that are occurring now. If this experience is not to be lost, the federal regulatory and other transportation agencies must become comprehensive and anticipatory, rather than remain myopic and reactive, in designing their reporting requirements, periodic censuses and surveys, and information storage-and-retrieval systems.

Even when robust field experiments can be implemented in a timely and adequate fashion, whether before or after the regulatory change, they cannot be used without cognizance of some significant caveats. Already mentioned are the abusive purposes to be avoided and the preconditions to be met for experimentation to be worthwhile. More importantly, experimentation should not be seen as a guarantee of totally unbiased objective results. Even if an objective reality exists from all perspectives, the process of designing and interpreting an experiment, especially one outside of the laboratory, requires much professional judgment to deal with the rampant uncertainties of evaluation research. In dealing with subjectivity, experimentation differs from other approaches in its requirement that the evaluator be highly self-critical and seek all alternative explanations for any research finding. It is not enough to predict the present with historical data or to demonstrate a measurement's goodness-of-fit with someone's pet theory. All validity threats must be explained or otherwise be published as explicit grains of salt for the policymaker's consideration.

Perhaps the greatest difficulty with experimentation is the need to specify objectives and impact measures. Although attention to construct validity may reduce some of the problems in specifying measures, it is very difficult to determine appropriate goals in the byzantine world of transportation policymaking. For example, was Amtrak created to provide an economically efficient rail passenger service, to serve other societal ends, to provide a hidden subsidy to freight railroads, or to provide a politically expedient method of abolishing passenger service from the rails? Each purpose embodies a very different set of appropriate measures and their valuation may be in conflict.

A final caveat is the need for a change in the attitudes and strategies among policymakers and their analysts. When public officials advocate a policy change as a solution rather than as an attempt to solve a major problem, then a careful evaluation exposes them to failure. Campbell (17) suggests an alternative strategy: Emphasize that the status quo is untenable and so important that several tentative solutions should be tried. A specific disappointment can thus be rationalized as only one in an ongoing series of efforts to deal with a difficult and multifaceted condition. The analyst's attitudes must also change. Although experimentation uses many commonplace techniques, these techniques must be used in a highly self-critical fashion rather than to prove a point or go on a statistical fishing expedition. One must always ask: Does the policy work, why, and how might it work better?

Careful experimentation is not a panacea for resolving the problems of regulated transportation; however, it does provide sensitive tools and a powerful discipline for developing policy. Experimentation can help better define the questions surrounding regulatory change, but even a successful experimental program does not eliminate our radical ignorance of the future. It may, however, increase the general confidence that what is true and workable today will persist into tomorrow (18).

ACKNOWLEDGMENT

The encouragement and comments by Edward Margolin and Douglas McKelvey are greatly appreciated, as is the assistance of Patricia Klaas. The findings and opinions are mine alone and not necessarily those of the National Transportation Policy Study Commission or of any individual commissioner.

REFERENCES

1. A.E. Pisarski. Institutional Impediments to Comprehensive Data Collection. TRB, Ad Hoc Task Force on Transportation Data, Draft Paper, Aug. 1978.
2. A. Kaplan. The Conduct of Inquiry: Methodology for Behavioral Science. Chandler Publishing Co., Scranton, PA, 1964.
3. C.H. Weiss. Evaluation Research: Methods of Assessing Program Effectiveness. Prentice-Hall, Englewood Cliffs, NJ, 1972.
4. H.J. Riecken and R.F. Boruch, eds. Social Experimentation: A Method for Planning and Evaluating Social Interventions. Academic Press, New York, 1974.
5. D.T. Campbell and J.C. Stanley. Experimental and Quasi-Experimental Designs for Research on Teaching. In Handbook of Research on Teaching (N. L. Gage, ed.), Rand McNally, Chicago, IL, 1963, pp. 171-246.
6. T.D. Cook and D.T. Campbell. The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings. In Handbook of Industrial and Organizational Psychology (M.D. Dunnette, ed.), Rand McNally, Chicago, IL, 1976, pp. 223-326.
7. G.V. Glass, V.L. Wilson, and J.M. Gottman. Design and Analysis of Time-Series Experiments. Colorado Associated Univ. Press, Boulder, 1975.
8. Charles River Associates. Measurement of the Effects of Transportation Changes. Urban Mass Transportation Administration, 1972.
9. R.R. Schmitt. Accessibility, Experimentation, and the Evaluation of Transportation Developments for Disadvantaged Groups. Johns Hopkins Univ., Baltimore, MD, Ph.D. dissertation, 1978.
10. D.T. Campbell. From Description to Experimentation: Interpreting Trends as Quasi-Experiments. In Problems in Measuring Change (C.W. Harris, ed.), Univ. of Wisconsin Press, Madison, 1963, pp. 212-254.
11. E.A. Suchman. Evaluative Research: Principles and Practice in Public Service and Social Action Programs. Russell Sage Foundation, New York, 1967.
12. T.C. Chamberlin. The Method of Multiple Working Hypotheses. Science, Vol. 15, Feb. 7, 1890, pp. 92-96.
13. R.R. Schmitt. Threats to Validity Involving Geographic Space. Journal of Socio-Economic Planning Sciences, Vol. 12, 1978, pp. 191-195.
14. D.B. Lee, Jr. Requiem for Large-Scale Models. Journal of the American Institute of Planners, Vol. 39, May 1973, pp. 163-178.
15. R.P. Carver. Special Problems in Measuring Change with Psychometric Devices. In Evaluation Research: Strategies and Methods, American Institute for Research, Pittsburgh, PA, 1970, pp. 48-66.
16. D.P. Warwick and C.A. Lininger. The Sample Survey: Theory and Practice. McGraw-Hill, New York, 1975.
17. D.T. Campbell. Reforms as Experiments. American Psychologist, Vol. 24, April 1969, pp. 409-429.
18. S.J. Mandelbaum. On Not Doing One's Best: The Uses and Problems of Experimentation in Planning. Journal of the American Institute of Planners, Vol. 41, May 1975, pp. 184-190.

Publication of this paper sponsored by Committee on Surface Freight Transport Regulation.

Evaluation of Trucking Entry Control: The Exempt Backhaul Case

Merrill J. Roberts, University of Maryland, College Park

This paper traces the potential effects of limiting trucking regulation by permitting independent operators to carry nonexempt general commodities on backhauls following the transportation of exempt agricultural products. These effects are determined by vehicle flows that establish the physical opportunity for traffic switches among regulated carriers and independent operators promoting efficiency, economically feasible modifications in traffic patterns, and the behavioral patterns of the regulated carriers, independent operators, and shippers. The logistical data are drawn from an Interstate Commerce Commission survey of more than 13 000 truck trips in 1975 that defines the competitive relations among the industry segments. These data indicate little prospect for productivity improvements due to saving empty trips; however, cost savings by independent operators are possible. They would gain in lease-bargaining strength, but orderly price competition between them and the regulated carriers would be quite limited except for Florida where it might be extreme. Service improvements are not indicated, but some deterioration may be hypothesized for Florida's outbound general commodities. The proposed regulatory change offers disadvantages with minimal gains. Adverse effects arise from the continuing regulation of outbound general commodities that can destabilize the Florida trucking market; the eligibility rule for inbound general commodities, which both limits competition and prevents intermarket capacity mobility; and the danger of disorderly market behavior. Even though regulation may be wasteful according to the dictates of the competitive model, partial deregulation may be counterproductive unless it is carefully designed so that the residual element of freedom and control mesh and do not clash as in this case.