Prepared for Submission to the 2005 Transportation Research Forum Annual Meeting

Ridership Accuracy and Transit Formula Grants

Xuehao Chu

Center for Urban Transportation Research
University of South Florida
4202 East Fowler Avenue, CUT 100
Tampa, Florida 33620

Phone: (813) 974-9831
Fax: (813) 974-5168
E-mail: xchu@cutr.usf.edu

This paper examines the accuracy of data in annual unlinked passenger trips reported to the National Transit Database (NTD) at the individual agency level. This examination takes a two-step approach. The first step compares the ridership reported by member agencies to the American Public Transportation Association (APTA) and the ridership reported to the NTD as recipients of transit formula grants. The NTD ridership can be as high as 50 percent more than the APTA ridership, and such significant positive deviation exists persistently over time across many agencies. The second step explores potential sources of these positive deviations by examining their components. Random errors, including both sampling errors and some of the non-sampling errors, do not help explain these one-sided deviations. Nor do occasional annual adjustments such as special events ridership to a direct count in the NTD ridership. Much of these positive deviations appear to be attributable to systematic non-sampling errors that result from undercounting in direct counts, from unintentional biases in procedures, or perhaps from intentional manipulation. Limited evidence in the literature, however, suggests that undercounting in direct counts is small at the systemwide level. The paper then quantitatively examines how these systematic non-sampling errors affect the allocation of two formula grants to Florida transit agencies: the Urbanized Area Formula Grant Program at the federal level and Florida's Transit Block Grant Program at the state level. The paper also discusses a strategy for reducing these systematic non-sampling errors.

Key words: National Transit Database, unlinked passenger trips, passenger miles, sampling, errors, formula grants

INTRODUCTION

Ridership is a factor in allocating transit formula grants at the federal level and in at least five states: Florida, Indiana, Iowa, New York, and Ohio. In FY 2003 the Federal program allocated about $3.5 billion to transit agencies in urbanized areas (FTA, 2003), while the funds allocated by the transit formula grant programs in the five states total around $200 million per year (FDOT, 2003; Jones, 2004a; IADOT, 2004; NYDOT, 2002; ODOT, 2004).

Both the formula and data inputs affect allocations. The transportation literature has focused on the use, design, and evaluation of the formula. Carstens et al. (1976), Hartman et al. (1994), and TCRP (2004) review state practices. Design studies include Dajani and Gilbert (1978), Levine (1978), GAO (1979), Alfa and Clayton (1986), Sinha et al. (1986), and Matherly (1997). Examples of evaluation are Miller (1979; 1980), Taylor (1991; 1995), Karlaftis and Sinha (1997), and Schmidt (2001). This paper contributes to the literature by addressing the role of data inputs. Specifically, it provides evidence of systematic errors in unlinked passenger trips reported to the National Transit Database (NTD), and examines their effects on the allocation of transit formula grants. By systematic errors we mean non-sampling errors that are positive in direction, significant in magnitude, widespread in scale, and persistent in time.

The general literature on allocation formulas for government grants-in-aid has a history of looking at how data inputs affect allocations (e.g., USDOC, 1978; Jabine et al., 2001; Louis et al., 2002). It is widely understood that errors in formula inputs lead to unfair allocations. More important, the effects of input errors can be amplified, attenuated, or influenced in other ways when the input errors interact with the properties of a formula.

The importance of addressing input errors is also questioned in the general literature. First, there is a tradeoff between increased input accuracy and the additional costs of producing these inputs. Second, accurate inputs may not be essential to achieve the goals of a formula grant program. Third, the losses are offset by the gains, so that the loss in overall social welfare resulting from these errors may be small. Fourth, correcting data errors may be less important than addressing more fundamental questions. It is suggested, for example, that what happens after the funds get to the recipient is an order of magnitude more important than marginal reduction in input errors.

However, the general literature puts priority on three inter-related areas: 1) Minimize the very large errors in data; 2) Identify and reduce persistent biases for identifiable geographic and demographic subgroups; and 3) Guard against intentional manipulation to get higher allocations. One important lesson from the general literature is that the data for all potential grant recipients come from the same source using the same general method. This paper contributes to this general literature by addressing the consequences of different agencies using different collection methods for formula inputs. Formula allocation procedures typically specify exactly where the input data should come from. For both the federal and Florida programs, for example, data on ridership come from the NTD. The problem is that transit agencies are allowed to use either a 100-percent count or a statistical estimate through sampling for data on ridership regardless whether a 100-percent is available (FTA, 2003).

Systematic errors can have undesirable consequences to grant allocations. First, they cloud the incentive role of ridership as an allocation factor. Transit agencies do consider the impact of changes in ridership from their actions on their share of transit formula grants. The Lee County MPO of Florida (2002), for example, considers the negative impact of raising transit fares by the local transit operator on the local share of the Urbanized Formula Grant Program because fare increases tend to decrease ridership. If these systematic errors are much greater than achievable ridership growth through improved services by transit agencies, much of the incentive role of ridership in a formula is lost. Second, systematic errors also lead to unfair allocations that deviate from what the original formula intended. The allocation of transit formula grants is a zero-sum game in most cases. One agency's gain is a loss to other agencies.

Such systematic errors also have undesirable consequences in the use of the NTD ridership at the individual agency level for other purposes. The NTD ridership, for example, is commonly used by transit agencies to compare themselves to individual peer agencies in terms of ridership per unit of services provided, ridership per unit of operating cost, or just ridership per capita. Such performance comparisons are important considerations in local decisions for transit funding. These local decisions would be seriously questioned if the NTD ridership for some of the agencies involved in such performance comparisons contained systematic errors.

The next section reviews the general transit formula programs that use ridership as an allocation factor and common methods for collecting ridership data. The paper then examines the accuracy of the NTD ridership through a two-step approach. The first step examines deviations of the NTD ridership from the ridership reported by member agencies to the American Public Transportation Association (APTA). The second step explores potential sources of positive deviations by examining their components, including random errors and systematic errors. After this comparison the paper draws the implications of unilateral errors in ridership by an agency to its allocation of transit formula grants. Following the presentation of these implications, a strategy for reducing systematic errors in ridership is discussed.

BACKGROUND

Techniques for Collecting Ridership Data

Ridership appears as unlinked passenger trips or passenger miles in transit formula grant programs. While the Urbanized Area Formula Grant Program uses ridership in terms of passenger miles, all the state programs mentioned earlier use unlinked passenger trips. The following discussion focuses exclusively on fixed-route bus services.

Counting Directly. Unlinked passenger trips may be counted directly by fareboxes semi-automatically or by drivers manually. Counting passengers through fareboxes is semi-automatic because drivers still need to manually record boarding for many non-cash fare categories. As more transit operators install electronic registering fareboxes on their entire fleet, counting passengers directly is becoming the transit industry's standard. counting passenger miles directly is rarely done now, but will be increasingly used. With automatic passenger counters installed on an entire fleet or smart cards used systemwide by all passengers, passenger miles could also be directly counted. The NTD reporting manual refers to the number of unlinked passenger trips counted directly as a 100-percent count (FTA, 2003). A 100-percent count may be smaller than the full 100-percent count because of possible undercounting and exclusion of special events ridership.

Estimating Statistically. When a sampling procedure is used, a transit agency has many options to achieve the target level of statistical confidence and precision levels in the estimate. One widely used option is the FTA-approved sampling plans (UMTA, 1988). These plans have two features that appeal to transit agencies (Chu, 2004). First, they follow a simple framework. For any of the plans, the same number of one-way vehicle trips is sampled on each sample day, and the interval between sample days is constant. Second, they are assumed to have universal applicability, i.e., any transit agency can use any of the sampling plans without customizing to its conditions. Other options are often called alternative techniques, which have two objectives in general. One is to improve sampling efficiency, i.e., reducing sample size for the same confidence and precision levels. The other one is to reduce the administrative costs of sampling. Two commonly used techniques to improve sampling efficiency are sample stratification (Smith, 1983) and ratio estimates (Furth and McCollom, 1987). One common technique to reduce administrative costs is cluster sampling (Furth et al., 1988).

Transit Formula Grant Programs

This section briefly reviews the Federal Urbanized Formula Grant Program and the Florida Public Transit Block Grant Program. Other state programs are also mentioned.

Urbanized Formula Grant Program. Part of this program provides capital, operating, and planning assistance to transit agencies in urbanized areas with populations of 200,000 or over based on a statutory formula. The allocation is based on population, population density, fixed-guideway route miles, bus and fixed-guideway vehicle revenue miles, and bus and fixed-guideway passenger miles traveled. All performance data are taken from the latest report year of the National Transit Database. Agencies may either use a 100-percent count or an estimate from a sampling procedure for annual ridership data reported to the NTD. When a 100-percent count results in undercounting, agencies may factor up the data to account for the missing percentage (FTA, 2003). When a sampling procedure is used, a transit agency may use any sampling procedure that meets the minimum confidence level of 95 percent and the minimum precision level of ±10 percent.

Florida Public Transit Block Grant Program. The Florida Department of Transportation administers the Public Transit Block Grant Program as created by Section 341.052 of the Florida Statutes. This program provides state funds for capital projects and bus transit operations to providers of transit services. No single agency is apportioned more than 39 percent of the funds available, but each agency is apportioned at least $20,000 annually. A statutory formula is used with one third of the funds available allocated on the basis of county population, revenue miles, and passengers carried, respectively. Data on revenue miles and passengers carried are for all modes combined, including purchased transportation. Data on passengers carried for this allocation come from the most recent NTD.

Other State Programs. The formulas of the other state programs differ in several ways (Hartman, et al., 1994; IADOT, 2002; NYDOT, 1999; ODOT, 2004). First, they differ in the amount of formula grants allocated using ridership. It is one-third in Florida versus 100 percent in Indiana, for example. Second, they differ in how ridership enters their allocation formulas. Third, they differ in the frequency of data submission and fund allocation. Almost all other programs allocate transit formula grants annually except New York, which allocates its formula grants on a quarterly basis with quarterly submission of ridership data from transit agencies. Fourth, they differ in the required statistical confidence and precision levels in ridership data. Florida requires the use of NTD data. New York requires the same statistical confidence and precision levels on an annual basis as those with the NTD but does not require the use of NTD data. Indiana has no statistical requirements on ridership data submitted from individual agencies (Jones, 2004b).

RIDERSHIP ACCURACY

Two-Step Approach

Ideally the accuracy in unlinked passenger trips reported to the NTD by individual agencies should be measured as the difference between what they report to the NTD and their full 100-percent count. This ideal approach, however, does not work because data are available on the NTD ridership but not on the full 100-percent count. Instead, this paper adopts a two-step approach that is not as rigorous as the ideal approach but still provides evidence on the accuracy of the NTD ridership for individual agencies. The first step would measure deviations of the NTD ridership from a 100-percent count, while the second step would explore the potential sources for any deviations observed.

Data on 100-percent counts could have been obtained through a survey of transit agencies specifically for this analysis. That was not done due to a lack of resources. Rather existing ridership data from the APTA are used. In addition to reporting annual ridership to the NTD, many member transit agencies also voluntarily report monthly ridership data to the APTA on a quarterly basis. These monthly ridership data from individual agencies form the basis for APTA's *Ridership Reports* (e.g., APTA, 2004), the basis for the annual ridership in APTA's *Public Transportation Fact Book* (e.g., APTA, 2003), and are directly used in the Transportation Services Index by the Bureau of Transportation Statistics (BTS, 2004). Both NTD and APTA ridership numbers are widely used in the transit industry for policy debates and decision making.

Quality Concern

There is an undocumented concern about the quality of the APTA ridership relative to that of the NTD ridership. This concern may have resulted from several factors. The APTA does not prescribe specific methods to its members for collecting the data (Bronson, 2003). The APTA does not prescribe specific statistical confidence and precision levels (Bronson, 2003; BTS, 2004). The APTA does not have an elaborate process for auditing, verifying, and validating the data from individual members (BTS, 2004). Finally, submission of ridership data is voluntary.

Our experience with the actual practice of many transit agencies indicates that this concern is largely unfounded for ridership data at the individual agency level. It is true that many member agencies do not report data for all modes and/or all months. This problem of missing data results largely from the voluntary nature of the process and does lead to a serious statistical problem of coverage errors in estimated total ridership at the national level. This coverage problem is irrelevant to the analysis in this paper, however.

What is relevant is the quality of ridership data at the agency level, where the quality concern is largely unfounded. If a large portion of member agencies did submit statistical estimates of their monthly ridership to the APTA, data quality would be a serious problem because of the absence of prescribed statistical confidence and precision levels and the absence of an elaborate process for auditing, verifying, and validating the data. The reality is that the vast majority of member agencies do not submit estimates, but rather they submit 100-percent counts to the APTA. Table 1 shows how data on unlinked passenger trips for reporting to the APTA and NTD were collected in 2003 by 23 Florida transit agencies, for which information is available. The absence of an elaborate process for catching errors has little negative effect. Furthermore, the absence of minimum confidence and precision levels has no negative effect on ridership data if they are 100-percent counts.

Table 1. Methods of Collecting Unlinked Passenger Trips for APTA and NTD Reporting, Florida, 2003

Collection Method	APTA	NTD
100%	21	15
Sampling	2	8
Total	23	23

Sources: Obtained directly from individual agencies through telephone conversations.

Deviations

The analysis in this section represents the first step of our two-step approach. Ridership deviations (NTD ridership minus APTA ridership) are assessed for three sets of transit agencies on an annual fiscal year basis. Since the NTD ridership is annual and on a fiscal year basis, the APTA monthly ridership data are converted to fiscal year data, based on the ending month of an agency's fiscal year. The first comparison is for directly-operated fixed-route bus services among 162 transit agencies in 2000 and among 135 transit agencies in 2002. These are all agencies reporting to the NTD, for which corresponding APTA ridership data are also available in the respective years. The results are summarized in Table 2 for three groups of these agencies. The NTD ridership is within one percent of the APTA ridership for the "No Deviation" group, less than the APTA ridership by more than one percent for the "Negative Deviation" group, and greater than the APTA ridership by more than one percent for the "Positive Deviation" group. A total of 57 transit agencies have a positive deviation in 2000, indicating that deviations of NTD ridership from APTA ridership are widespread. Furthermore, NTD ridership is greater than APTA ridership by 13.2 percent on average among the 57 agencies in 2000, indicating positive deviations of significant magnitudes.

Table 2. Differences in NTD and APTA Ridership for 162 Agencies in FY 2000

Groups	APTA (000s)	NTD (000)	NTD-APTA (000s)	100(NTD-APTA)/APTA (%)	Number of Agencies
No Deviation	1,468,774	1,468,105	-669	0.0	74
Negative Deviation	821,272	778,845	-42,427	-5.2	31
Positive Deviation	1,645,828	1,863,408	217,580	13.2	57
All	3,935,874	4,110,358	174,484	4.4	162

Sources: Excel files of member agencies reported ridership to APTA for calendar years 1999-2001 from APTA's Terry Bronson and an Excel file of ridership for individual agencies from the NTD website for fiscal year 2000. Ridership data are limited to directly-operated fixed-route bus services.

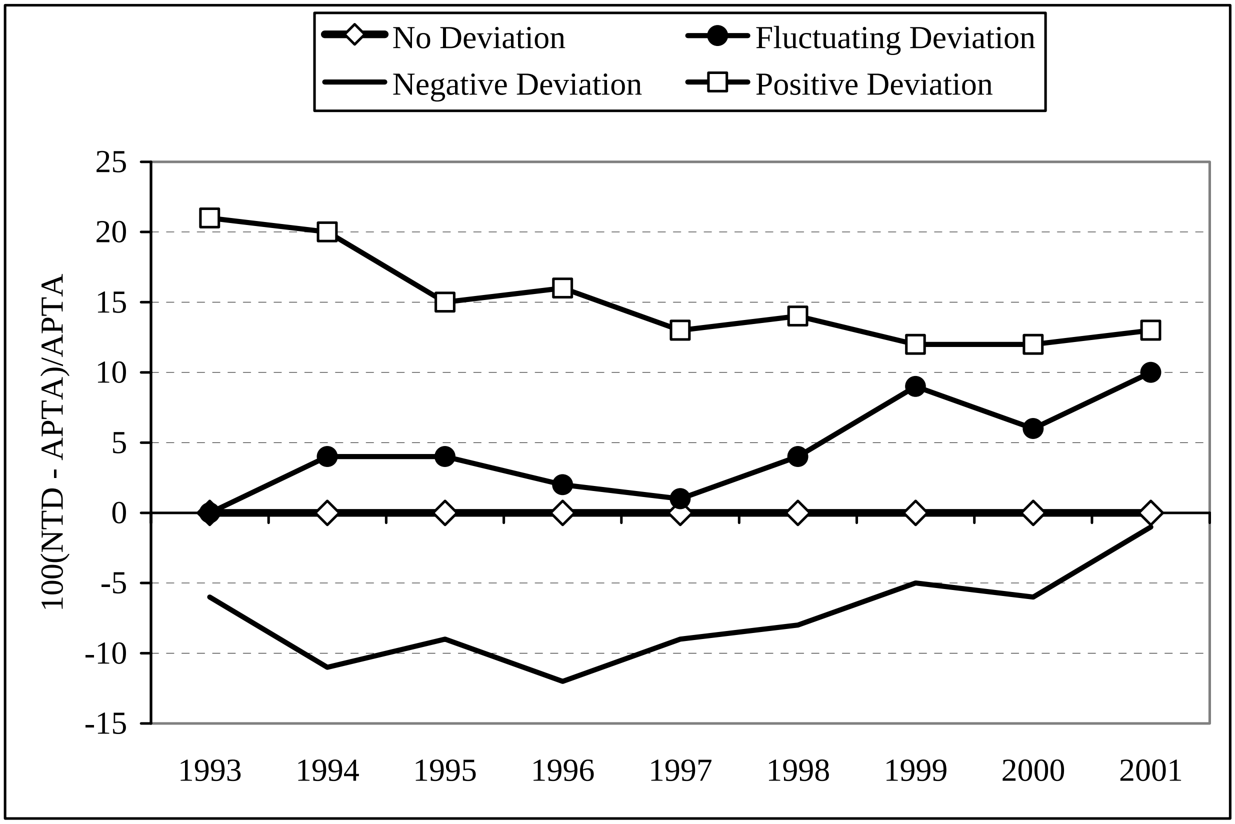
The second comparison is for directly-operated fixed-route bus services among the largest 27 bus agencies during the period from 1993 to 2001. These 27 bus agencies are categorized into four groups, based on their relative NTD and APTA ridership over the entire period. Among them, 5 agencies have no significant deviations in either direction; 11 with fluctuating deviations; 2 with negative deviations every year; and 9 with positive deviations every year. Figure 1 shows these differences in percentage terms. Again, the problem of positive deviations of NTD ridership from APTA is significant and widespread. For the group of agencies with positive deviations as a whole, the deviations average 15 percent over the entire period. The aggregate deviations of NTD ridership from APTA ridership are positive not only among the agencies with positive deviations during the period but also among the agencies with fluctuating deviations. Combined these agencies represent a total of 20 of the 27 largest bus systems. In addition, these positive deviations are consistently present over time. The aggregate deviations range annually between 12 percent and 21 percent for the group with positive deviations throughout the analysis period.

The third comparison is for all modes among Florida transit agencies from 1993 to 2001. The modes are combined for this comparison because the Florida State Block Grant Program uses the total ridership of all modes of service. Figure 2 shows the percent deviations for agencies that estimate NTD ridership with sampling. While not shown, deviations of NTD ridership from APTA ridership are negligible for all other Florida agencies. Agencies with large deviations, including VOTRAN, ECAT, HARTline, TalTran, and Space Coast, all use FTA-approved sampling plans. However, the deviations are far smaller for PalmTran, PSTA, and BCT even though they also use FTA-approved sampling plans. Again, the deviations are positive and significant in magnitudes in most cases.

Sources of Ridership Deviations

The qualitative analysis in this section represents the second step of our two-step approach. It is clear from the first step that positive deviations of NTD ridership from APTA ridership are significant, widespread, and persistently present over time. There are five types of potential sources that may have contributed to these deviations: APTA ridership being estimates, annual adjustments like special events ridership to a 100-percent count, 100-percent counts being undercounted, sampling errors, and non-sampling errors.

Figure 1. Percent Deviations in Directly-Operated Fixed-Route Bus Services for Largest 27 Bus Systems by Group

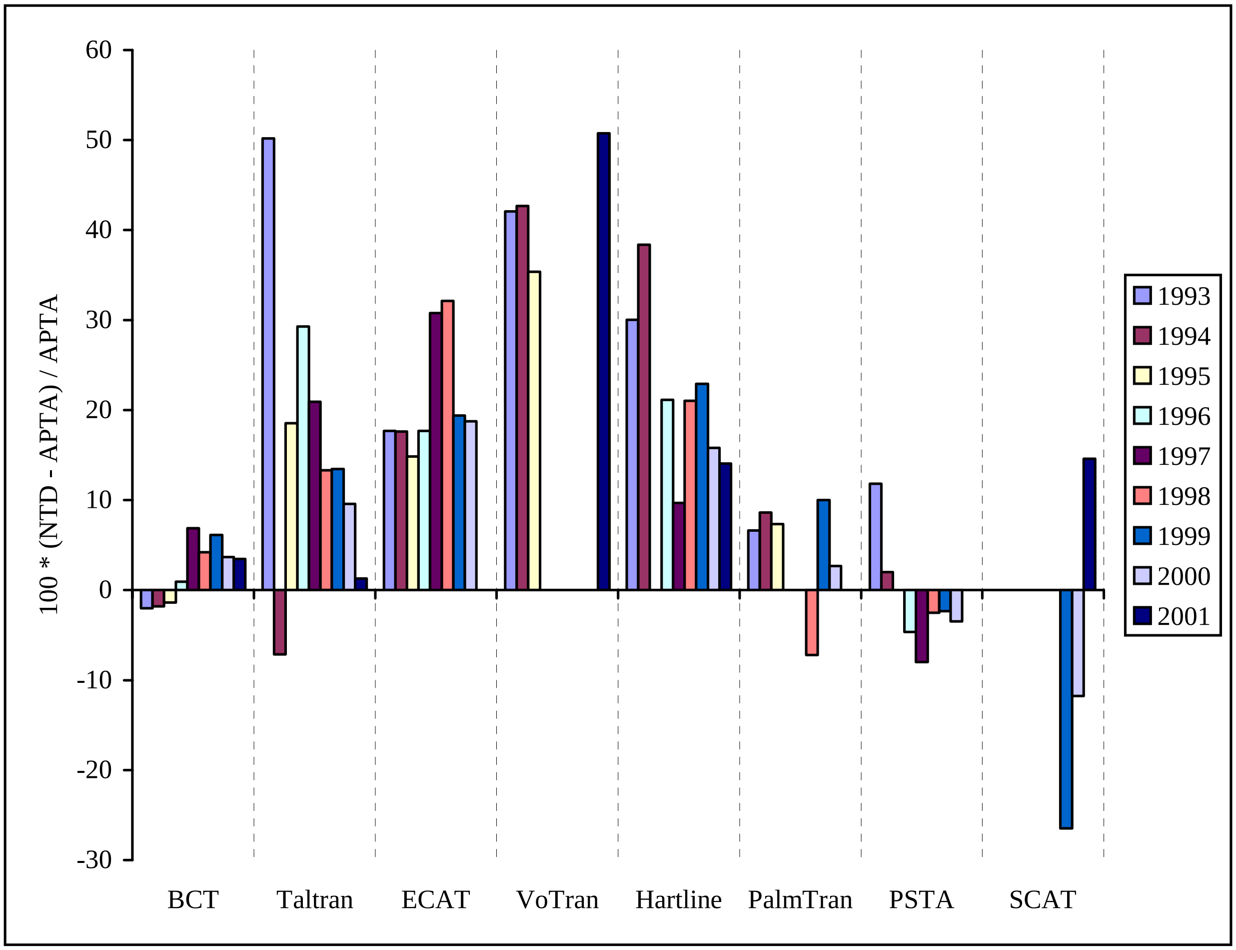


Notes: The number of agencies within each group: No Deviation = 5; Fluctuating Deviation = 11; Negative Deviation = 2; and Positive Deviation = 9.

Sources: APTA ridership data are from the quarterly APTA *Transit Ridership Reports* from the relevant years. NTD ridership data are from the annual *Data Tables*. The largest bus systems are based on those listed in APTA's *Transit Ridership Reports for Largest Bus Systems*.

One potential complication of using APTA data is that not all agencies report an estimate to the NTD and a 100-percent count to the APTA. Based on conversations with many of them, the methods that transit agencies use to collect APTA ridership fall into five categories in relation to their NTD method: 1) Report a 100-percent count to both entities; 2) Report an estimate to both entities using the same sample; 3) Report a 100-percent count to the APTA but an estimate to the NTD; or 4) others. In the first two cases, the two annual ridership numbers should be similar. In the third case, these two numbers can deviate if the 100-percent count to the NTD includes some annual adjustments like special event ridership. Any deviation in this case is likely to be small. An example of the fourth case would be reporting a revenue-based estimate to the APTA but an estimate based on an FTA approved sampling plan to the NTD. In this case, the two ridership numbers can deviate in either direction. Our conversations with many agencies indicate that few agencies may fall into the fourth case. In the third case, however, these two numbers can deviate. When the deviation is negligible, any one of these four cases is a possibility. Without knowing the specific methods that individual agencies use in obtaining either NTD or APTA ridership, these negligible deviations do not help assess the accuracy in the NTD ridership. A significant deviation, on the other hand, would most likely to indicate a 100-percent count being reported to the APTA but an estimate to the NTD. That is, a significant deviation would indicate either significant undercounting in 100-percent counts reported to the APTA or upward sampling and non-sampling errors in the NTD estimate.

Figure 2. Percent Deviations in Total Ridership for Florida Transit Systems



Notes: The bars from the left to the right for a given agency follow the temporal order of the fiscal years 1993 through 2001. An absence of bars indicates missing data. These operators are:

BCT = Broward County Mass Transit Division
TalTran = City of Tallahassee-TALTRAN
ECAT = Escambia County Area Transit
VoTran = County of Volusia
Harline = borough Area Regional Transit Authority
PalmTran = Palm Tran, Inc.
PSTA = Pinellas Suncoast Transit Authority
SCAT = Space Coast Area Transit

Sources: APTA ridership data are from the quarterly APTA *Transit Ridership Reports* from the relevant years. NTD ridership data are from the annual *Data Tables*.

The NTD requires agencies to report all ridership on its fixed-route services, including special events ridership. Since the APTA ridership is monthly data, some agencies do not include special events ridership to the APTA ridership. However, they do add such ridership during their entire fiscal year to the directly counted ridership before reporting to the NTD. In almost all cases, such annual adjustments are small and are unlike to account for much of the positive deviations.

A 100-percent count is most commonly made using electronic registering fareboxes (Furth, 2000). The majority of the passengers are registered automatically as they pay their fare, and the remainder are registered manually by the operator. The synthesis by Furth (2000) indicates that operator errors in registering passengers usually have a systematic downward but small error at the system level. If this is true, undercounting in 100-percent counts is unlikely to account for most of the observed positive deviations of NTD ridership from APTA ridership.

Statistical sampling does lead to sampling errors. When sampling is done correctly, however, these sampling errors go either direction, and over time the positive and negative errors should balance each other out. As a result, sampling errors are unlikely to account for much of the one-sided deviations.

Non-sampling errors can result from many sources (FCSM, 1978). The list of bus trips may be incomplete. Sampling may be non-random. Missed random trips may be replaced with conveniently selected trips. Errors could also exist in data recording. In addition to these sampling related sources, non-sampling errors can also exist with the estimation procedure. Sample averages may be computed incorrectly. Data on the number of actually run vehicle trips may be incorrect. Institutions may also play a role. Agencies may not emphasize technical competency of their staff. The NTD process is considered by most transit agencies as a technical and financial burden (Perk and Kamp, 2003). The extensive NTD process of auditing, verifying, and validating is designed to capture anomalies, such as sudden changes over time. Persistently over-estimated ridership over time would appear normal to this process. What would be more likely to be caught as anomalies would be the random fluctuations in ridership resulting from sampling errors.

In general non-sampling errors from these potential sources should also fluctuate in both directions just like sampling errors. They may not be random, however, if they result from unintentional biases in procedures or perhaps international manipulation. Agencies have a strong incentive to not underestimate their NTD ridership. When their execution of current procedures leads to persistently higher NTD estimates of ridership, on the other hand, agencies have a strong financial incentive to keep these procedures and to execute them as they used to.

Summary

The NTD ridership is significantly greater than the APTA ridership at the individual agency level, and these positive deviations are widespread and persistently present over time. Occasional annual adjustments like special events ridership may have contributed some to these positive deviations. Undercounting in 100-percent counts is like to have also contributed.

However, most of these positive deviations appear to be attributable to systematic non-sampling errors that have resulted from unintentional biases in procedures or perhaps from intentional manipulation.

IMPLICATIONS

This section quantifies the effects of hypothetical errors in ridership on the allocations of transit formula grants. The analysis focuses on Florida, and considers both the Urbanized Formula Grant Program at the federal level and the Florida Formula Block Grant Program at the state level. The allocation formulas are first analyzed to determine the elasticity of funding allocation to individual agencies with respect to unilateral ridership errors. Dollar changes in allocation to any agency as a result of a unilateral ridership error are then determined, using the elasticity result.

Elasticity Analysis

Florida Block Grant Program. To derive the allocation elasticity, define the following parameters:

- A_i = funding allocated to agency i,
- A = total annual funding available for the Public Transit Block Grant Program,
- P_i = annual passengers carried by agency i,
- P = annual passengers carried by all agencies eligible with $\sum P_i = P$, and
- $S_i = P_i / P$ with $\sum S_i = 1$.

The statuary formula requires that funding be allocated according to the following relationship:

$$A_i = \left(\frac{A}{3}\right)\left(\frac{P_i}{P}\right) \tag{1}$$

This allocation formula implies an allocation elasticity as follows:

$$E_i \equiv \frac{\partial A_i}{\partial P_i}\frac{P_i}{A_i} = 1 - S_i \tag{2}$$

which is bounded within (0,1). The larger an agency is, the smaller its elasticity. The elasticity for most agencies is close to 1. For an agency carrying $100S_i$ percent of Florida’s passengers, a one-percent unilateral upward bias in passengers reported would result in a $100(1-S_i)$ percent increase in its state funding from the Public Transit Block Grant Program.

Other State Programs. While the four other state programs all use ridership in allocating some of their transit formula grants, they differ from the Florida program in their allocation elasticity with respect to differences in ridership because ridership enters their formulas differently.

Indiana: The Public Mass Transportation Fund divides annual total funds among five groups of transit agencies with fixed proportions, and within each group allocates one-third to individual agencies based on the ratio of each agency's per capita ridership multiplied by its locally derived income over the same quantity for the entire group, and another one-third based on the ratio of each agency's ridership per vehicle revenue mile multiplied by its locally derived income over the same quantity for the entire group (Hartman et al., 1994). The allocation elasticity under this formula also takes the form in equation (2), with S_i measured within the group to which agency *i* belongs.

Iowa: The State Transit Assistance Program allocates one-quarter of total formula funds to urban transit agencies based on each agency's ridership per unit of operating cost relative to the sum of this ratio across all urban agencies (IADOT, 2002). Let C_i be the ratio of ridership per unit of operating cost for agency *i* to the sum of ridership per unit of operating cost across all urban agencies. The allocation elasticity for its urban program also takes the form in equation (2) with S_i replaced by C_i. The more cost-effective an agency is, the smaller its elasticity. This elasticity is bounded within [1-$C_{\max}$, 1-$C_{\min}$], where $C_{\max}$ and $C_{\min}$ are the maximum and minimum of all C_i ratios.

New York: The State Mass Transportation Operating Assistance Program allocates formula funds to small systems based on fixed amounts per vehicle revenue mile and per passenger mile on a quarterly basis (NYDOT, 2004). The recent fixed rate per passenger, for example, has been $0.405. With this formula, the allocation elasticity is constant at 1.

Ohio: The Ohio Public Transportation Grant Program allocates one-fifth of the total funds to transit agencies based on each agency's share of ridership and another one-fifth based on a ratio V_i, defined as ridership per vehicle revenue mile for agency *i* relative to the sum of ridership per vehicle revenue mile for all transit systems (ODOT, 2004). Its allocation elasticity is given by the average of (1-S_i) and (1-V_i) with S_i and V_i as the weights, or

$$E_i = \frac{S_i(1-S_i)+V_i(1-V_i)}{S_i+V_i} \tag{3}$$

How the size of an agency (S_i) or its service effectiveness (V_i) affects its elasticity is more complicated than the cases described above. However, this elasticity still is bounded within (0,1) as long as all V_i ratios are no greater than 1.

Urbanized Formula Grant Program. This analysis focuses on the bus incentive portion, which is apportioned on the basis of passenger miles weighted by passenger miles per unit of operating cost (FTA, 2003). Also, this bus incentive portion represents 5.57 percent of the Urbanized Area Formula Program, which is $2,997,316,081 for FY 2001. Similar to the Florida program, the following are defined:

- F_i = funding allocated to agency *i*,
- $F = \sum F_i$,
- M_i = annual passenger miles on directly-operated fixed-route bus services for agency *i*,
- O_i = operating costs of directly-operated fixed-route bus services for agency *i*, and

- $Z_i = M_i / \sum M_j$.

The statuary formula requires that the funds be allocated according to either one of the following two equivalent relationships. One formula is:

$$F_i = F\left[\frac{M_i\left(\frac{M_i}{O_i}\right)}{\sum_j M_j\left(\frac{\sum_j M_j}{\sum_j O_j}\right)}\right] \qquad (4)$$

This is based on the national distribution of passenger miles weighted by passenger miles per unit of operating cost. That is, how much funding a single agency gets is proportional to its national share of passenger miles weighted by passenger miles per unit of operating cost. This proportion is given by the bracket on the right hand side of equation (4). The other formula is:

$$F_i = \left[M_i\left(\frac{M_i}{O_i}\right)\right]\left[\frac{F}{\sum_j M_j\left(\frac{\sum_j M_j}{\sum_j O_j}\right)}\right] \qquad (5)$$

This is based on the dollar unit value given by the second bracket on the right hand side of equation (5). This dollar unit value is the amount of total funding allocated to each unit of passenger mile weighted by passenger miles per unit of operating cost. The funding to a single agency is determined by the product of this dollar unit value by its passenger miles weighted by passenger miles per unit of operating cost. This unit value is $0.00490633 for FY2001.

Either allocation formula implies an allocation elasticity as follows:

$$E_i = \frac{\partial F_i}{\partial M_i}\frac{M_i}{F_i} = 2\left(1 - Z_i\right) \qquad (6)$$

which is bounded within (0,2). The larger an agency is, the smaller its elasticity. It is close to 2 for most agencies. For an agency carrying $100Z_i$ percent of passenger miles among all eligible agencies, a one-percent upward bias in its passenger miles carried would result in a $200(1\text{-}Z_i)$ percent increase in federal funding from the bus incentive portion of the Urbanized Area Formula Program. The elasticity is 1.98, 1.90, 1.80, and 1.60 for agencies representing 1 percent, 5 percent, 10 percent, and 20 percent of nationwide passenger miles, respectively.

Allocation Analysis

For Florida transit agencies for which FY 2001 data are available, this section illustrates increases in an agency's ridership-based allocations from both the federal and state programs that result from a unilateral positive error in reported ridership. For the Florida Public Transit Formula Grant Program, the illustrative unilateral error is 10 percent in the number of unlinked passenger trips for all modes. The allocation does not consider the 39 percent upper limit or the $20,000 lower limit mentioned earlier. For the Urbanized Formula Grant Program, the illustrative unilateral error is also 10 percent but in the amount of passenger miles for directly operated fixed-route bus services. Table 3 summarizes the results for individual agencies. A unilateral 10 percent error in ridership would mean an increase ranging from 9.5 percent for the Miami-Dade Transit Authority to 14.2 percent for the Sarasota County Transportation Authority in combined state and Federal incentive formula grants. These increases in grant allocations can represent a strong incentive for transit agencies to report ridership with positive errors.

Table 3. Total Allocated State and Federal Incentive Grants and Added Funds from Hypothetical Unilateral 10% Upward Differences in Ridership, Florida Agencies, FY 2004

Agencies	Base Allocation of State & Federal Incentive Grants	Added Funds from 10% Bias	
		Amount	% of Base Allocation
Manatee County Area Transit	$93,583	$10,641	11.4
Pinellas Suncoast Transit Authority	$1,193,306	$145,581	12.2
Lee County Transit	$337,367	$44,833	13.3
Broward County Mass Transit Division	$4,714,308	$593,086	12.6
Gainesville Regional Transit System	$728,852	$85,401	11.7
Lakeland Area Mass Transit District	$183,088	$22,549	12.3
County of Volusia dba: VOTRAN	$551,743	$71,072	12.9
Miami-Dade Transit	$10,525,649	$1,000,632	9.5
Central Florida Regional Transportation Authority	$3,481,763	$466,118	13.4
City of Tallahassee-TALTRAN	$432,736	$49,057	11.3
Palm Tran, Inc.	$824,588	$104,902	12.7
Escambia County Area Transit	$193,394	$24,225	12.5
Jacksonville Transportation Authority	$1,320,319	$174,864	13.2
Hillsborough Area Regional Transit Authority	$1,226,014	$150,087	12.2
Sarasota County Transportation Authority	$295,832	$42,131	14.2
Space Coast Area Transit	$106,223	$11,788	11.1
Pasco County Public Transportation	$39,588	$4,700	11.9
Bay County Council On Aging	$24,174	$2,491	10.3
Indian River County Council on Aging, Inc.	$37,852	$3,821	10.1
Polk County Transit Services Division	$17,234	$1,744	10.1

Notes: The 10 percent upward differences are hypothetical for illustration purposes. For any given agency, the added funds only reflect the impacts of a 10-percent upward bias in its reported ridership on its only grant allocations. The impacts of a 10-percent upward bias in any one agency's reported ridership on other agencies' grant allocations are not included.

STRATEGY

An effective long-term strategy for reducing the systematic non-sampling errors in ridership would be to require the use of the same 100-percent count of unlinked passenger trips for all reporting purposes. Many agencies use a 100-percent count of unlinked passenger trips for all internal purposes, but use sampling techniques for estimating both unlinked passenger trips and passenger miles for transit formula grant programs. In addition to potential systematic errors in ridership, another problem with this practice is the inconsistent numbers in unlinked passenger trips between annual reporting to formula grant programs and other reporting purposes.

Requiring the use of a 100-percent count also gives transit agencies an incentive to improve the accuracy of their counts. Improving the estimation of an agency's ridership that has systematic positive errors leads to lower allocations of transit formula grants to that agency. Improving the accuracy of a 100-percent count, in contrast, leads to higher reported ridership and higher allocations as a result.

Requiring the use of a 100-percent count may also encourage transit agencies to estimate passenger miles as the product of the 100-percent count of unlinked passenger trips with an estimated average passenger trip length through alternative sampling techniques. Many agencies report a 100-percent count of unlinked passenger trips to all reporting entities but still use FTA-approved sampling plans for estimating passenger miles. The problem with this practice is the inconsistency between annual unlinked passenger trips and passenger miles and the implied average passenger trip length. Some agencies are already using this average trip length approach to estimating passenger miles but are still using FTA-approved sampling plans. Alternative techniques are available that should require much smaller sample sizes in general and still meet FTA's minimum confidence and precision levels (Chu, 2004).

One concern for requiring transit agencies to report a 100-percent count is how allocations may change within a state. In terms of Florida's Program, the five agencies using FTA-approved sampling plans for estimating unlinked passenger trips and reporting ridership with positive systematic errors would lose $357,000 for FY2004. This would represent a loss of about 7.9 percent of what they would receive without this change. In contrast, all other Florida agencies would collectively gain the same amount, representing an average gain of 2.5 percent.

A far more significant concern of adopting this strategy is its potential negative impact on a state. In terms of the Federal program, for example, Florida would lose $307,352 for FY2004 if it adopts the strategy alone. This loss of federal funding would represent about 3.8 percent of what Florida would receive without adopting the strategy. Over time, this loss is likely to be smaller as more and more agencies across the country are switching or considering switching to using a 100-percent count to report as NTD ridership. Furthermore, any such loss is likely to be offset to some degree by savings in sampling costs from using more efficient sampling techniques based on passenger trip length. This is a real concern, however. For this long term strategy to be truly effective, the Federal Transit Administration must adopt it nationwide, and at the same time reduces the reporting burden on transit agencies.

CONCLUSIONS

The paper has examined the accuracy of data in unlinked passenger trips reported to the National Transit Database. Rather than using the ideal approach in which the NTD ridership would be compared with the full 100-percent count, the paper takes a less rigorous two-step approach due to a lack of data on full 100-percent counts. The first step is quantitative and compares the ridership reported by member agencies to the American Public Transportation Association and the NTD ridership. The second step is qualitative and explores potential sources of observed deviations between NTD and APTA ridership numbers from the first step by examining their components.

It is clear that the NTD ridership is significantly greater than the APTA ridership for many transit agencies over a number of years. Random errors, including both sampling errors and some of the non-sampling errors, do not help explain these one-sided deviations. Nor do occasional annual adjustments such as special events ridership to 100-percent counts in the NTD ridership. Evidence in the literature, though limited, suggests that undercounting in the APTA ridership is small at the system level and hence does not explain the significant deviations in the NTD ridership from the APTA ridership. Most of these positive deviations appear to be attributable to systematic non-sampling errors in the NTD ridership that result from unintentional biases in procedures, or perhaps from intentional manipulation.

Through the grant allocation formulas, especially the formula in the Urbanized Formula Grant Program, these systematic non-sampling errors represent strong financial incentives for them to be maintained. These systematic errors can significantly increase an agency's share of the portion of a transit formula grant program that involves ridership. The error elasticity of grant allocation is close to 2 for most agencies through the Urbanized Formula Grant Program and close to 1 for most agencies through the state programs. Over-reporting ridership by 10 percent by an agency, for example, would result in an almost 20 percent increase in allocation to that agency from the incentive tier of the Urbanized Formula Grant Program. The allocation of transit formula grants to individual agencies is a zero-sum game in most cases. One agency's unfair gain is the unfair loss to other agencies.

It is important to point out the shortcomings of this approach in order for the reader to assess the paper's contribution appropriately. One shortcoming is its reliance on limited evidence in the literature on the degree of undercounting in 100-percent counts. While this limited evidence indicates that undercounting is a small problem at the system level on average, occasional stories suggest a more serious problem for some agencies, especially during the first a few years of using electronic registering fareboxes. It is hard to imagine, however, that a significant portion of the observed positive deviations between NTD and APTA ridership numbers could have resulted from undercounting in 100-percent counts of ridership. If it were the case, it would be a serious problem in itself for the NTD. If Florida is any indication, the major of transit agencies report a 100-percent count of unlinked passenger trips as well to the NTD. In any case, there is a need to better understand the extent of undercounting in 100-percent counts.

Another shortcoming of this paper is its failure to pinpoint specific non-sampling errors that have contributed to the systematic errors in ridership reported to the NTD. It is difficult for a researcher to find out the exact sources of non-sampling errors for any particular agency. If done on purpose, agencies are likely to avoid a direct answer when asked. Otherwise, poor record-keeping and staff turnover also prevent the researcher from conclusively identifying the sources. Rather than using resources to identify direct evidence in unintentional biases in procedures or intentional manipulation if any, a far more fruitful approach would be to support a change of policy related to reporting ridership to the National Transit Database.

The current policy gives transit agencies the option of either reporting an estimate from sampling or a 100-percent count, regardless whether a 100-percent count is available. The policy should be changed to require all transit agencies to report a 100-percent count both at the Federal level and at the state level when there is a state transit formula grant program. Exclusive of any impact on grant allocations, there is little cost to this policy change. Most agencies already report a 100-percent count to other entities, including their governing boards and the American Public Transportation Association. If they have not already, more transit agencies are switching and considering switching to electronic registering fareboxes.

This policy change has several benefits. It will result in consistent numbers in unlinked passenger trips being reported to formula grant programs and to other entities. It will give an incentive to transit agencies to improve the accuracy of their counts. It will also enable transit agencies to sample more efficiently for estimating passenger miles by multiplying their 100-percent count of unlinked passenger trips and average passenger trip length.

ACKNOWLEDGMENT

The Florida Department of Transportation (FDOT) provided funding through the National Center for Transit Research. Tara Bartee is the FDOT Project Manager. Christian Senn compiled the APTA ridership data. Terry Bronson provided electronic data on APA ridership for calendar years 1999-2001. The paper has benefited from comments provided by and discussions with Victoria Perk and Steve Polzin. The author wants to particularly thank Ed Fleischman and his staff in the Federal Transit Administration's Office of Oversight and members of the NTD contractor for their insightful comments during a presentation to them.

REFERENCES

Alfa, AS, and D Clayton (1986), An Improved Procedure for Allocating Public Transit Operating Subsidies in Canada, Journal of Advanced Transportation, 20(3), pp. 223-238.

APTA (2003), 2003 Public Transportation Fact Book, Washington, D.C.

APTA (2004), APTA Ridership Report: Third Quarter 2003, http://www.apta.com/research/stats/ridershp/index.cfm, accessed February 2004.

Boyle, Daniel K (1998), Passenger Counting Technologies and Procedures, TCRP Synthesis 29, Washington, D.C., Transportation Research Board.

Bronson, Terry (2003), e-mail communication.

Bureau of Transportation Statistics (BTS) (2004), Transportation Services Index, http://www.bts.gov/programs/economics_and_finance/transportation_services_index/html/public_transit_ridership.html.

Carstens, RL, CR, Mercier, and EJ Kannel (1976), Current Status of State-Level Support for Transit, Transportation Research Record, 589, pp. 14-19.

Chu, Xuehao (2004), Customized Sampling Plans: A Guide to Alternative Sampling Techniques for National Transit Database Reporting, Center for Urban Transportation Research, Tampa, Florida.

Dajani, Jarir S, and Gorman Gilbert (1978), Measuring the Performance of Transit Systems, Transportation Planning and Technology, 4, pp. 97-103.

Federal Committee on Statistical Methodology (FCSM) (1978), Statistical Policy Working Paper 4 – Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics, http://ntl.bts.gov/DOCS/sw4.html.

Federal Transit Administration (FTA) (2003), National Transit Database 2003 Reporting Manual, U.S. Department of Transportation, http://www.ntdprogram.com/NTD/ntdhome.nsf/?Open.

Florida Department of Transportation (FDOT) (2003), 2002 Program Objectives and Accomplishments, Office of Management and Budget.

Furth, Peter G (2000), Data Analysis for Bus Planning and Monitoring, TCRP Synthesis 34, Washington, D.C., Transportation Research Board.

Furth, Peter G., and Brian McCollom (1987), Using Conversion Factors to Lower Transit Data Collection Costs, Transportation Research Record, 1144: 1-6.

Furth, Peter G., Keith L. Killough, and Gary F. Ruprecht (1988), Cluster Sampling Techniques for Estimating Transit Patronage, Transportation Research Record, 1165: 105-114.

General Accounting Office (GAO) (1979), Analysis of the Allocation Formula for Federal Mass Transit Subsidies, Washington, D.C.

Hartman, RJ, EM Kurtz, and AB Winn (1994), The Role of Performance-Based Measures in Allocating Funding for Transit Operations, TCRP Synthesis of Transit Practice, 6, Transportation Research Board.

Iowa Department of Transportation (IADOT) (2002), Iowa State Transit Assistance Formula Allocation, http://www.iatransit.com/regulations/sta_committee/current_sta.pdf.

IADOT (2004), Revised STA Projections for FY 2004, accessed on February 25, 2004. http://www.iatransit.com/history/sta_information/revised_fy04_STA_1-7-04.xls.

Jabine, Thomas B., Thomas A. Louis, and Allen L. Schirm (2001), Choosing the Right Formula, Washington, D.C., National Research Council.

Jones, Brian (2004a), Excel file via e-mail on total formula funds allocated through the Indiana Public Mass Transportation Fund since 1986.

Jones, Brian (2004b), e-mail communication.

Karlaftis, MG; Sinha, KC (1997), Performance-Based Transit Operating Subsidy Allocation: A Before and After Study, Journal of Public Transportation, 1(2), pp. 1-23.

Lee County MPO (2002), Long-Term Options for Funding Transit and Transportation Disadvantaged Services in Lee County-FINAL REPORT, North Fort Myers, Florida.

Levine, HA (1978), Using Efficiency Incentives to Allocate Transit-Operating Deficits, Traffic Quarterly, 32(1), pp. 87-104.

Louis, TA, TB Jabine, and MA Gerstein (2002), Statistical Issues in Allocating Funds by Formula, Washington, D.C., National Research Council.

Matherly, D (1997), Developing a Performance-Based Transit Allocation Formula: Case Study for a Participatory Process, Transportation Research Record, 1604, pp. 83-91.

Miller, James H (1979), An Evaluation of Allocation Methodologies for Public Transportation Operating Assistance, Transportation Journal, 19 (1), pp. 40-49.

Miller, JH (1980), Use of Performance-Based Methodologies for The Allocation of Transit Operating Funds, Traffic Quarterly, 34(4), pp. 555-574.

New York Department of Transportation (NYDOT) (1999), Rules and Regulations Part 975, Statewide Mass Transportation Operating Assistance Program.

NYDOT (2002), 2001 Annual Report on Public Transportation Assistance Programs in New York State, Passenger Transportation Division, http://www.dot.state.ny.us/pubtrans/annual/2001annual2.html.

Ohio Department of Transportation (ODOT) (2004), 2004 Ohio Public Transportation Grant Program (OPTGP) - Formula Criteria**,** accessed February 2004. http://www.dot.state.oh.us/ptrans/downloads/2004%20Formula%20Criteria%20final.doc.

Perk, Victoria, and Nign Kampa (2003), Handbook of Automated Data Collection Methods for the National Transit Database, Center for Urban Transportation Research, Tampa, Florida.

Schmidt, S (2001), Incentive Effects of Expanding Federal Mass Transit Formula Grants, Industrial and Labor Relations Review, 20(2), pp. 239-261.

Sinha, KC, JN Parsons, DT Barnum, and EA Sharaf (1986), Development and Implementation of a Performance-Based State Transit Subsidy Allocation Formula for Indiana, Transportation Research Record, 1076, pp. 30-38.

Smith, Robert L., Jr. (1983), Innovation and Incremental Improvement in Bus Transit Passenger Data Collection, Washington, D.C., U.S. Department of Transportation.

Taylor, BD (1991), Unjust Equity: An Examination of California's Transportation Development Act, Transportation Research Record, 1297: pp 85-92.

Taylor, BD (1995), Program Performance versus Transit Performance: Explanation for Ineffectiveness of Performance-Based Transit Subsidy Programs, Transportation Research Record, 1496, pp. 43-51.

TCRP (2004), A Status Report on the Use of Performance-Based Measures in Allocating Transit Funding, Transit Cooperative Research Program, Transportation Research Board.

Urban Mass Transportation Administration (UMTA) (1988), Sampling Procedures for Obtaining Fixed Route Bus Operating Data Under the Section 15 Reporting System, Circular UMTA-C-2710.1A, U.S. Department of Transportation.

US Department of Commerce (USDOC) (1978), Report on Statistics for Allocation of Funds, Statistical Policy Working Paper 1, Washington, D.C.