Assessment of the Geographical Accuracy of the Carload Waybill Sample for State Rail Planning

WILLIAM R. BLACK and JOHN C. ROBBINS

ABSTRACT

Carload waybill statistics as collected by the Interstate Commerce Commission and the Federal Railroad Administration have been the primary data source on railroad commodity flows for more than three decades, and a considerable amount of geographic research during that time period has made use of these data. Using recently released carload waybill data at the freight station level and actual flow data for the Penn Central, Norfolk and Western, and Southern railroads, a statistical comparison of the two data sets is undertaken. On the basis of correlation coefficients the waybill sample does a good job of estimating the universe. However, an examination of the standard errors indicates values large enough to cause some concern.

It has been 30 years since Ullman (1) introduced geographers and transport planners and analysts to a set of state-to-state railroad flow data referred to as the 1 percent Carload Waybill Sample. He used the sample as his major data source in the classic American Commodity Flow (2), and in later years he was involved in efforts to improve the quality of these data (3).

Over the years these data have been used to analyze commodity flows in the United States (4), to study the demand for transportation (5), and to examine the economics of light density rail lines (6). Public sector state and national railroad planning studies have made extensive use of these data (7,8). Regional rail traffic statistics (9) and projections (10) are based primarily on data of the Waybill Sample. According to Rhodes and Briggs (11) the sample is also used by the rail industry to evaluate the effects of rate changes, to identify possible merger impacts, and to develop marketing strategies. More recently researchers have used the sample to analyze hazardous materials routing procedures (12,p.39) and to evaluate the temporal stability of certain spatial interaction flows (13,p.190; 14) and models (15,p.191). Some of these studies have used the state-to-state flows and others have worked with the individual waybill sample elements before aggregation, or aggregated in an atypical fashion (e.g., on a line basis).

In spite of this high level of usage there is little in the literature regarding these data and their level of accuracy in geographical and regional research. It is the purpose of this paper to remedy this shortcoming. To this end, in the paper that follows the historical background of the 1 percent Carload Waybill Sample and the attributes of the data obtained are discussed. This is followed by some statistical checks of the accuracy of these data. Before proceeding, some comments on the nature of the waybills are merited.

A waybill is a document that accompanies a rail freight shipment from origin to destination and specifies: originating railroad, originating station (FSAC and SPLC), terminating railroad, terminating station (FSAC and SPLC), waybill number and date, commodity (STCC), freight charges, billing weight, short lines miles, and number of carloads. The waybill may also specify up to eight interchanges and bridge lines used in shipping the commodity. Although additional information is provided (16), these elements are the ones of interest to most researchers. As the list implies, most of the data are coded with freight stations identified by their freight station accounting code (FSAC) and standard point location code (SPLC) and commodities represented by their standard transportation commodity classification (STCC) code.

BACKGROUND OF THE SAMPLE

It should be evident that such a waybill would be useful for many types of transportation research. This has been recognized since the early years of this century when waybills were compiled on an irregular basis to identify the volume of traffic moving between different points or the total volume of a given commodity type that was being carried. Such compilations involved the universe of information and there were no problems of sampling.

According to Smith (17) the first national sample of these data was collected in 1932 by the Federal Coordinator of Transportation. It covered all terminations on a single day in 1932 and it was recognized that such a sample had considerable bias and, as a result, it was used little. Smith also noted that the Board of Investigation and Research undertook the collection of a national sample of waybills in 1939. That sample involved all traffic terminating on one day of each month for all Class I railroads. Because of the staggering of days in the sample, it represented a significant improvement over the previous sampling effort.

It was in 1946 that the Interstate Commerce Commission (ICC) began work on the sampling design of a national waybill sample. After considerable debate and discussion the ICC decided that the sample would consist of all waybills ending in "01." It was believed that this would yield an unbiased 1 percent sample of total carload rail shipments in the United States.

The sample and sampling design have been altered over the years since it was first collected [ICC order issued September 6, 1946 (11,p.235)]. Through

1966 the sample was collected and published annually by the ICC. Due to budgetary problems there was no sample collected in 1969. In 1970 the Federal Railroad Administration of the U.S. Department of Transportation took over the processing of the sample although the data were still collected by the ICC. Due to these changes the data are referred to as the ICC Waybill Sample, the FRA Waybill Sample, or the DOT Waybill Sample. Regardless of the titles, all references are to the same sample data.

For several years the data collected were published by the ICC in their State to State Distribution series. This series gave the carloads and tonnages of nearly three dozen commodity groups moving between the states of the United States. Beginning in September 1971 the U.S. Department of Transportation took over the publishing function with the appearance of waybill statistics for the year 1969 (18). They continued this practice until 1972. Due to a low level of demand the U.S. DOT ceased publishing the data, but they remained available in computer printout form and this continues to be the case. It should be noted that the bulk of the research undertaken to date has involved the use of these state-to-state flows.

Although the individual elements of the 1 percent sample, that is the sample waybills, were not available to planners and researchers in that disaggregated form in the past, this policy was altered in 1982. At that time the ICC ordered that individual waybills (as stored on computer) for current and previous years could be made available to states and railroads for planning and analysis purpose (19).

ACCURACY OF THE SAMPLE

Despite the broad use of waybill data, there has not been a large volume of research done on their accuracy. As the agency charged with ensuring the quality of the data, the FRA requires each rail carrier to submit a sample tape for one quarter of the year. Information that is given on the tape is compared with the carrier's Quarterly Commodity Statistics (QCS) report filed for that same quarter. This latter report gives the total carloads terminated on the railroad's lines for each carload commodity class. If the sample information, expanded to represent the QCS universe, falls within an acceptable confidence interval as identified by FRA, the sample is deemed complete.

Whether or not the data submitted on the waybills is correct is checked by the ICC. That agency may require hard copies of the waybills to check against data submitted on computer tape. Neither this check nor the QCS check necessarily ensures the geographical accuracy of the waybill sample.

This shortcoming was recognized by Harris (20) in research that he undertook before examining the cost savings of rail line abandonment in the United States (6). In his preliminary research Harris sought to examine statistical aspects of the waybill sample because it was to be the data source for his subsequent research. He wanted to identify whether the waybill data were biased based on the size of station, or biased based on line traffic density. In addition he wanted to know whether the waybill sample could reliably predict actual carloadings originating and terminating on a line.

Using FRA-supplied waybill data from the sample for 1973, 1974, and 1975, as well as actual data for those years supplied by the Southern Pacific Transportation Company for their rail operations, Harris examined his research questions using a series of statistical tests. On the basis of this research he concluded that there is no significant bias in the waybill sampling distribution by size of station; there was a slight bias in the waybill sampling distribution by size of line segment (i.e., there was a significant undersampling of traffic originating on high-density lines and an oversampling of low-density lines); errors of prediction were large for individual lines having few observations in the waybill sample; aggregation to increase total traffic on lines or to increase the size of the line resulted in error levels of about 20 percent; and pooling of traffic for several years increases the accuracy of the line estimates (20,pp.42-43). Aside from Harris's work on the accuracy of the waybill when aggregated to rail lines, there does not appear to be any research in this area.

From a geographical perspective it is possible to work with waybill data on a station basis. This is a type of minimal aggregation level that usually represents the traffic of several rail users (shippers and receivers). It is the finest level of detail for which this type of data is available, and it is at this level that the data were evaluated for accuracy.

It would also be of interest to evaluate the accuracy of the waybill data when they are aggregated to the state level because a significant amount of research has been undertaken using the state-to-state rail flows. However, this would be extremely difficult because railroads do not usually keep data on a state basis. In addition, most states are usually served by several railroads. As a result one would have to set up a "universe" of data aggregated by states in order to have something to compare with the state-level waybill data. This is hardly practical.

At the highest level of aggregation the national rail data compiled by the Association of American Railroads (21) may be compared with the waybill data from the sample that has been aggregated to the national level. Comparing the carloads and tonnages of these two data sets results in the data given in Table 1. During the period from 1972 to 1981 the expanded sample data for carloads were consistently less than the universe. The same was true of tonnage information. One possible reason for the observed differences is that the waybill data exclude originated traffic from Mexico or Canada. It is unlikely that such international rail trade would account for from 1.6 to 4.9 million carloads or from 90 to 276 million tons of traffic per year during the 1972 to 1981 time period. Although there are other reasons for the difference according to the U.S. Department of Transportation (22), the fact that the difference is so small in 1982 (see Table 1) suggests that the previous differences were due to error. That the level of difference between the data sets is decreasing for recent years, and particularly for 1982, does not necessarily mean that the error is decreasing. It may simply mean that the traffic of certain railroads may have been oversampled. In effect, the observed differences in the data imply the existence of error, but perfect agreement would not necessarily mean an absence of such error.

DATA USED

One of the major barriers to assessing the accuracy of the waybill sample data in the past has been the unavailability of data on actual rail operations that were comparable to some aggregation level of the waybill sample. Although such data are available to the rail industry and federal agencies and such analyses may have been undertaken by corporate and governmental planners, the results of such analyses, if they were undertaken, have not been made public.

Release of the sample data at the individual way-

TABLE 1 Comparison of National "Universe" and Expanded Waybill Sample[a]

Year	Carloads	Sample Carloads	Tonnage	Sample Tonnage
1972	26.1	22.8	1,448	1,296
1973	27.3	23.3	1,532	1,340
1974	26.2	21.3	1,531	1,255
1975	23.2	19.2	1,395	1,177
1976	23.5	20.3	1,407	1,255
1977	23.2	20.1	1,395	1,248
1978	23.4	21.0	1,390	1,320
1979	23.9	22.0	1,502	1,412
1980	22.6	20.7	1,492	1,402
1981	21.6	20.0	1,453	1,363
1982	18.5	18.4	1,269	1,271

Note: Actual carload data are from the Association of American Railroads; sample carloads were compiled by the authors from printouts of the state-to-state flows. Data for 1982 were supplied by the ICC.

[a]Carloads and tonnages are in millions.

bill level by the Interstate Commerce Commission (19) provided waybill data from the sample. "Universe" data on rail operations were drawn from state-level rail planning studies (23,7) in Indiana. The first planning study provided data on all Penn Central Railroad operations in Indiana for 1973. Those data consisted of information on tonnages, commodities, and revenues for every carload of traffic originating or terminating at stations within the state. The 1982 study provided selected data for stations of the Norfolk and Western Railway and the Southern Railway in Indiana for 1980 and 1981, respectively.

RESEARCH DESIGN

Using the 1973 Penn Central Railroad data and the other data sources noted, three series of statistical tests were run. The first series of tests was as follows:

1. All carloads for all Penn Central stations in Indiana were estimated using the 1 percent waybill sample carloads for those stations. Stations not in the waybill sample were read as zero carloads. There were 493 stations in this test.
2. All Penn Central stations that had equal to or greater than 100 carloads were estimated using their corresponding waybill estimate. In this test there were 195 stations.
3. All stations in the 1 percent waybill sample, which would be all stations with more than one carload (or as expanded, 100 carloads), were used to estimate their corresponding station totals in the Penn Central "universe." There were 180 stations in this test.
4. The fourth test involved all Penn Central stations with fewer than 40,000 annual carloads being estimated by their corresponding carloads from the waybill sample data.

The purpose of this fourth test was to ensure that the results of the analysis were not being biased by large stations in Indiana.

A second series of tests sought to assess the accuracy of estimating rail branch line carloads by station. For these tests four samples of 15, 30, and 45 stations were created; these were treated as branch lines. Each of the 12 samples of stations was then estimated using the corresponding station traffic levels in the waybill sample.

The third series of tests involved data from the Norfolk and Western Railway (N&W) in 1980 and the Southern Railway (SR) in 1981. As in the case of the Penn Central tests these tests involved

1. All carloads for the N&W and SR stations were estimated by the corresponding waybill sample carloads for those stations. As before, stations not in the waybill data were read as zero carloads. These tests involved 146 N&W and 39 SR stations.
2. All stations of the N&W and SR in the waybill sample, which would be all stations with more than one carload, were compared with their corresponding values in the railroad-supplied data. These tests involved 64 N&W and 23 SR stations.

One point that should be noted before proceeding is that before the collection of the 1981 data there were some major changes in the sampling design of the waybill sample. These changes were brought about by the failure of the 1 percent sample to do a reasonable job of picking up unit trains and resulted in a variable percentage sampling rate in the case of multiple-car shipments.

There is also some question about exactly how comparisons of this type should be carried out. It is obvious that the waybill sample is not statistically independent of the universe from which it was drawn. Because of this there are problems of a philosophical nature regarding statistical inference. But the present study seeks only to analyze the accuracy of the sample data and as a result it appears to be appropriate to use correlation and regression analysis to indicate the accuracy of the sample. At the same time considerable attention will be given to the regression parameters and the standard error of estimate in the comparisons that follow.

RESULTS OF THE TESTS

Table 2 gives the results of the first series of tests undertaken. Certain facts should be noted from the table. Each of the four Penn Central tests displays an amazingly accurate and consistent performance. Explained variation, as measured by the coefficient of determination (r^2), ranges from 98.35 to 99.19 percent. The intercepts (a) are near zero, and the regression coefficients (b) range from 111.987 to 115.451. In theory these latter values should be 100 because the sample is 1 percent of total carloads. As was true for Table 1 the differences there and the departure of the regression coefficients from 100 is believed to be due to the absence of certain data from the waybill sample. Finally, the standard errors (SE) are reasonably small but still large enough that they should not be ignored. As expected, this error tends to be less for the largest samples. In the best case (Test 4) the estimates derived from the regression expanded sample would tend to miss actual values by 457 carloads or less in nearly two-thirds of the estimates.

TABLE 2 Accuracy of the Waybill Sample for Estimating 1973 Penn Central Rail Traffic by Station—Summary Statistics[a]

Test	N	$\bar{Y}$	r^2	a	b	SE
1	493	1,252	0.9919	21	112.123	481
2	195	3,127	0.9914	54	111.987	759
3	180	3,369	0.9913	38	112.062	788
4	491	1,001	0.9835	−2	115.451	457

[a]Letters in the table are defined as follows: N = number of observations (stations), $\bar{Y}$ = average number of carloads per station based on railroad-supplied data, r^2 = coefficient of determination, a = intercept of the regression equation, b = regression coefficient of slope value, and SE = standard error of estimate.

The second series of experiments involved the analysis of small sample accuracy and accordingly of whether the waybill sample will yield reasonable estimates for stations along individual branch lines. Results of the 12 small sample tests are given in Table 3. As the data in the table indicate, the average traffic per station on the 12 generated lines, Y, ranged from 635 to 3,249 carloads. In all cases the results were good with the majority of r^2 values exceeding 0.990. None of the intercept (a) values differed significantly from the expected 0.0 and the regression coefficients (b) were similar to those obtained earlier and averaged about 116. The standard errors (SE) were also reasonably good and ranged from 158 to 742 carloads with an average of 351.7 carloads.

TABLE 3 Tests of Small Sample Accuracy—Summary Statistics[a]

Sample	N	$\bar{Y}$	r^2	a	b	SE
1	15	1,736	0.997	-17.933	101.602	251
2	15	1,031	0.991	8.636	127.804	186
3	15	1,157	0.978	-70.219	105.196	442
4	15	3,249	0.999	32.151	119.167	259
5	30	1,521	0.987	4.721	107.046	426
6	30	1,311	0.969	-82.060	115.447	571
7	30	635	0.988	-6.947	127.605	158
8	30	2,155	0.999	54.079	119.389	230
9	45	2,126	0.928	-134.219	123.051	742
10	45	3,019	0.998	-116.219	111.528	558
11	45	861	0.994	38.735	119.435	178
12	45	1,893	0.999	35.886	119.221	219

[a]For column heading identification see Table 2.

The third series of tests was not as promising (Table 4). These were the tests involving the 1980 and 1981 station data of the N&W and Southern Railways, respectively. In the case of the N&W these tests were not as good with coefficients of determination in the neighborhood of 92 percent and standard errors that were too large for the flows involved. Nevertheless, the (a) intercept values were reasonably close to zero and the regression coefficients in the vicinity of 106 were reasonable. The Southern tests were less conclusive with explained variation in the vicinity of 95 percent, but with standard errors in excess of 1,100 carloads in the best case and intercept and slope values that can only be described as unacceptable.

The generally poorer results of these last four tests may be attributable to a number of factors. First, the "universe" data used in these tests were not audited and as a result are not nearly of the quality of the Penn Central data. Second, there has been a considerable increase in the amount of unit train traffic. This might account for greater error levels in 1980 and 1981 in comparison to the 1973 Penn Central case. Third, as previously noted, the sampling procedure was changed before the 1981 sample and it can only be assumed that the 1980 sample was considered too poor to continue with the existing sampling design. Fourth, errors in the 1981 results may simply indicate that more work needs to be done on the current sampling methods because parameter values indicate a tendency toward undersampling. Continuing this final point it should be noted that the newer sampling procedures will not necessarily increase the quality of the data at the station level although it may do this at the state level.

TABLE 4 Accuracy of the Waybill Sample for Estimating 1980 Norfolk and Western and 1981 Southern Rail Traffic by Station—Summary Statistics[a]

Test	N	$\bar{Y}$[b]	r^2	a	b	SE
N&W (1)	146		0.917	47.324	105.897	661
N&W (2)	64		0.919	1.591	106.509	950
Southern (1)	39		0.952	-254.598	207.049[c]	1132
Southern (2)	23		0.949	-547.493	211.752	1456

[a]For column heading identification see Table 2.
[b]The mean flow per station has been deleted to avoid disclosure of confidential shipper information.
[c]Due to changes in sampling and weighting, the 1981 Southern data are no longer a 1 percent sample but rather a weighted estimate. The actual values obtained for b were 2.07049 and 2.11752. They are presented in the table as though they were for a 1 percent sample for comparability.

CONCLUSIONS

The purpose of this paper has been to evaluate how accurate the carload waybill sample is in relation to actual carload data by station. Three series of tests were undertaken between the sample and actual flow data; a total of 20 comparisons were performed. On the basis of these comparisons and tests it may be stated that

1. The 1973 sample was good and the 1980 and 1981 samples were good (based on correlation coefficients) at estimating the actual number of carloads by station for large numbers of stations.
2. Even in the best of the large sample cases regression estimates of actual carloads yield standard errors that range from 450 to nearly 800 carloads per estimate. Although this does not completely undermine the utility of the waybill sample, it is something of which researchers and planners should be aware.
3. For small numbers of stations the results also appear to be good in terms of standard errors. This would suggest that the estimates may be good for branch line economic analysis provided the branch lines have a large volume of carloads.
4. It is not at all clear that the current sampling procedures are doing as good a job of estimating station traffic levels as former methods did. Differences observed here may be due to the sampling procedures, or to a substantive change in the character or nature of rail traffic (e.g., more multiple car moves per waybill), or to the quality of the more recent data used for the tests.

The first three conclusions suggest that the waybill sample may be used with some confidence for estimating traffic at stations within a state, estimating traffic for stations on branch lines and for the economic analysis of these, representing rail traffic at places in correlation studies, developing maps of potential rail traffic for state-size areas, and analyzing interstate rail flows for different commodities. Although this last point does not stem directly from this research, it is a reasonable inference because the aggregation of the station sample data used here can only serve to smooth out errors that may exist due to minor over- or undersampling.

Although the results of the tests described here appear to elevate the waybill sample to a much higher level of credibility, there is still the fourth conclusion that should make researchers hesitate to rest easy about the current sampling methods. Whenever possible researchers and planners should attempt to assess the accuracy of the data from the waybill sample before using it for major policy decisions of a public or corporate nature. It is apparent that a major evaluation of the current sampling procedure needs to be undertaken before analysts use it as the basis for substantial empirical work. In the interim a considerable amount of insight may be provided by the analysis of earlier waybill sample data.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to the Penn Central Corporation, the Consolidated Rail Corporation, the Norfolk Southern Corporation, and the Interstate Commerce Commission for permission to use the data that appear in this study. They are also grateful for the comments of James R. Blaze of Conrail, Steven D. Eisenach of Norfolk Southern, and Sidney Fine and William R. Southard of the ICC on an earlier draft of this paper.

REFERENCES

1. E.L. Ullman. Transportation Geography. In American Geography: Inventory and Prospect (P.E. James and C.F. Jones, eds.), Syracuse University Press, Syracuse, N.Y., 1954.
2. E.L. Ullman. American Commodity Flow. University of Washington Press, Seattle, 1957.
3. E.L. Ullman. Inadequacy of U.S. Flow Statistics. In Transport Flow Data (H.O. Whitten, ed.), The American University, Washington, D.C., 1968, pp. 153-155.
4. S. Spiegelglas. Some Aspects of State to State Commodity Flows in the United States. Journal of Regional Science, Vol. 2, 1960, pp. 71-80.
5. E. Perle. The Demand for Transportation. Research Paper 95. Department of Geography, The University of Chicago, Chicago, Ill., 1964.
6. R.G. Harris. Economics of Traffic Density in the Rail Freight Industry. Bell Journal of Economics, Vol. 8, 1977, pp. 556-564.
7. Division of Regional Transport Research, Indiana University. Indiana State Rail Plan Update 1983. Indiana Department of Transportation, Indianapolis, 1982.
8. Secretary of Transportation. Rail Service in the Midwest and Northeast Region. U.S. Department of Transportation, 1974.
9. 1972 Carload Waybill Statistics. Statement TD-1. U.S. Department of Transportation, 1974.
10. J.R. Pucher. Projections of 1980 Freight Demands for Selected Midwestern Railroads. DOT-TST-76T-40. Office of University Research, U.S. Department of Transportation, 1976.
11. R.G. Rhodes and R.E. Briggs. The ICC's Carload Waybill Sample. ICC Practitioner's Journal, Vol. 35, 1968, pp. 235-251.
12. P.E. Johnson. Predicting Routes for Highway or Railroad Shipments of Certain Hazardous Materials. AAG Program Abstracts 1983. The Association of American Geographers, Washington, D.C., 1983.
13. D.C. Knudsen. An Analysis of Temporal Change in an Interaction System. AAG Program Abstracts 1983. The Association of American Geographers, Washington, D.C., 1983.
14. D.C. Knudsen. Modelling Change in a Commodity Flow System: An Examination of U.S. Rail Freight Flows, 1972-1981. Ph.D. dissertation, Indiana University, Bloomington, 1984.
15. P.A. Williams. Forecasting Commodity Flows from Historical Sample Data. AAG Program Abstracts 1983. The Association of American Geographers, Washington, D.C., 1983.
16. S. Fine and R. Owen. Documentation of the ICC Waybill Sample. Office of Policy and Analysis, Interstate Commerce Commission, 1981.
17. R.T. Smith III. Technical Aspects of Transportation Flow Data. Journal of the American Statistical Association, Vol. 49, 1954, pp. 227-239.
18. 1969 Carload Waybill Statistics. Statement TD-1. U.S. Department of Transportation, 1971.
19. Interstate Commerce Commission. Release of Waybill Data to States and Railroads. 47 Fed. Reg. 7,778, 1982.
20. R.G. Harris. A Statistical Analysis of the FRA Waybill Sample. Office of Rail Economics and Policy Development, Federal Railroad Administration, U.S. Department of Transportation, 1977.
21. Railroad Facts. Association of American Railroads, Washington, D.C., 1983.
22. 1979 Carload Waybill Statistics. Statement TD-1. U.S. Department of Transportation, 1980.
23. Center for Urban and Regional Analysis, Indiana University. Indiana State Rail Plan. Public Service Commission of Indiana, Indianapolis, 1975.

Publication of this paper sponsored by Committee on State Role in Rail Transport.