

Excel Workbook Documentation

For Computations Made in the STB Rail Rate Study

January 2009

The Excel Workbook is designed to translate the data on revenue, tons, and miles retrieved from the Waybill sample using an Access database into charts showing trends in the rates that railroads charge their customers. This documentation consists of a list of the spreadsheets contained in the workbook, along with information about the data they contain and the manner in which they operate.

Before describing each of the sheets in the workbook, we will provide a description of the common characteristics found in sheets three (TAB [Tornqvist Index]) through ten (TAB [Revenue]). The data in these eight sheets are based on a summary of data from the Waybill sample or information based on simple arithmetic transformations of the Waybill sample data. Each of these eight sheets has three rows at the top. The top row contains the table title. The next row is blank, and the third row contains the following column headers:

- A. This column contains the name of the Standard Transportation Commodity Code (STCC) for which summary data follows to the right of it. There are several names that are unique to this study and do not follow strictly the categories established through the STCC system: (1) Intermodal—these are movements that are part highway and part rail; they are removed from the remaining STCC categories. For example, if some items that are STCC 20 (Food and Kindred Products) are loaded into a highway trailer that is then carried to a railroad ramp before being further transported by rail, that will be classified as intermodal and not as STCC 20. (2) Grain—these are all commodities in the STCC 0113 (Grain) and STCC 0114 (Oil Seeds) categories. These products are generally carried in bulk using hopper cars. They are very different from other Farm Products, such as fresh fruits and vegetables, which may require refrigerating equipment. (3) All Other—this is a category that is simply an amalgamation of all items that do not fit either the intermodal group or any of the other STCCs selected for separate examination.
- B. This column is hidden and contains either a “1” or a “0”. The value in this column controls the labeling information shown in columns A and C. The concept here is for columns A and C to show information about commodity groups only when that information changes. For example, the first four rows show data about “Intermodal” shipments, coded in column C with STCC “00”. The information in column B allows that the word “Intermodal” and the code “00” appear only in the first of those four rows. The number in column B is computed based on examining column D in TAB [Summary_Statistics] to identify the rows where changes occur in the column labeled “STCC Description”.
- C. STCC. This is the two-digit STCC for the commodity belonging to this row of data, with the following exceptions: (1) Intermodal is shown as “00”; (2) Grain is shown as “XX”; and (3) All Other is shown as “99”. The information is generated by referencing TAB [Summary Statistics] and TAB [STCC List],

both to be discussed later.

- D. **Dist Category.** This column provides a code for distance category, with the following definitions: (1) “S” signifies “Short”, which is less than 500 miles except for Metallic Ores and Non-metallic Minerals, which are short only for distances less than 100 miles. (2) “M” signifies “Medium”, which is between 500 and 1000 miles, except for Ores and Non-metallic Minerals, which are medium only for distances between 100 and 250 miles. (3) “L” signifies “Long”, which is generally any movement over 1000 miles (250 miles for Metallic Ores and Non-metallic Minerals), except that for Intermodal and Coal movements this category ends at 1500 miles. (4) “VL” signifies “very long”, which is a distance category of over 1500 miles, reserved for Coal and Intermodal shipments only. An “X” indicates that the row represents a combination that is not divided into any distance categories.
- E. **Car Ownership.** This column provides a code for the car ownership category. In most cases, the cells in this column contain an “X”, indicating that the combination in that row is not further divided into different car ownership categories. Only two commodity groups, Coal and Grain, show this division. For cells containing a “P” the row represents privately-owned cars. Rows containing “R” in this column have data for railroad-owned cars. There is an exception for 1985 and 1986. For those two years, the appropriate variable (Car_Ownership) has no entries; that is, the entry is null for each Waybill record. Therefore, the code in those two years shows “R” when those records have not been divided into car ownership categories at all. We elected not to make any adjustments to the data for this anomaly as the effect is very small.
- F. **Train Type.** This column contains the code for the train type category. As before “X” indicates that the row is not further divided into separate categories for train type. Unit trains are indicated by “U” (50 or more cars tendered at once); multi-car shipments are designated as “M” (6 to 49 cars); and single-car lots (5 and fewer cars) are designated “S”.
- G. **1985.** This column contains the data element that applies to that row for 1985. For example, in TAB [Revenue] it contains the total freight revenue for the combination of elements applying to that row. Looking at row 4 of TAB [Revenue], we see that we are talking about Intermodal shipments that are short in distance (< 500 miles) and that are not separated according to car ownership or train type. For shipments in this category, the total of freight revenue received for shipments terminating in the United States in 1985 was \$199 million. Each of the subsequent years’ revenue is provided in columns to the right. For example, we see that the revenue received for this category of shipment in 2000 was \$390 million. For some of the spreadsheets, there are four columns at the end to show change information: columns AD, AE, AF, and AG. These columns show the percent change in the data item from 2003 to 2004, 2004 to 2005, 2005 to 2006, and 2006 to 2007, respectively.

These four change columns are found only in the following sheets: TAB [Revenue], TAB [Real Revenue], TAB [Tons], TAB [Ton-Miles], and TAB [RealRPTM].

The following list describes each of the spreadsheets:

1. TAB [TI Graph]. The documentation of data in other sheets describes the computational process for creating the information shown in this chart. This chart shows the historic value of the Tornqvist Index for every year from 1985 through 2007.
2. TAB [Tornqvist Index]. This table shows the computation of the Tornqvist Index based on the real revenue per ton-mile. For each cell, the following formula generates the value: $\text{LN}(\text{RealRPTM}(\text{current year})/\text{RealRPTM}(\text{previous year}))$. The composite Tornqvist Index in row 71, uses the following formula to compute the change in the index between year y and year y-1:

$\text{EXP}(\text{SUMPRODUCT}(\text{K4:K70}, \text{RealRevCmdShare!K4:K70}))$.

That is, $e^{(\sum_i [\ln(R_{i,y}/R_{i,y-1}) * (R_{i,y} + R_{i,y-1}) / (T_{i,y} + T_{i,y-1})])}$

Where, \sum_i represents the sum over all output types i.
 $R_{i,y}$ = Real Revenue for output type i in year y.
 $T_{i,y}$ = Total of Real Revenue for all output types in year y.

A starting point is provided by setting the 1985 Tornqvist Index equal to 100, arbitrarily.

3. TAB [RealRevCmdShare]. This table shows the percentage share of real revenue received by each of the 67 groups used in the study. The real revenue share is shown as a sum of the current year's real revenue and the previous year's real revenue for each of the category rows (4-70) divided by the sum of the real revenues in the current and preceding year for the total real revenues (row 71)..
4. TAB [Change in RealRPTM]. This table shows, for each of the 67 output groups, the change in real revenue per ton-mile from the preceding year to the current year.
5. TAB [RealRPTM]. This table shows the amount of real revenue, in cents, received by the railroads for each ton-mile carried, separated by the 67 categories of output being considered here. The computation is based on the revenue and ton-mile information contained in TAB [Summary_Stats] with adjustment made for the value of the dollar and converted to cents. Summing the data in columns G-AC for rows 4-70 provides the data in row 71 for columns G-AC.

6. TAB [Ton-Miles]. The information in this sheet is computed not in this workbook, but instead in the database. The database results are pasted as values into TAB [Summary Stats]. The values are then passed to TAB [Summary_Statistics] and then forwarded on to this worksheet. Summing the data in columns G-AC for rows 4-70 provides the data in row 71 for columns G-AC.
7. TAB [Tons]. The information in this sheet is computed not in this workbook, but instead in the database. Each row in this table represents the total tons in thousands for each market segment combination indicated by that row in the sheet. The database results are pasted as values into TAB [Summary Stats]. The values are then passed to TAB [Summary_Statistics] and then forwarded on to this worksheet. Summing the data in columns G-AC for rows 4-70 provides the data in row 71 for columns G-AC.
8. TAB [Real Revenue]. This spreadsheet shows the real revenues, in millions of dollars, collected by the railroad industry for each category of rail shipment. Real revenues are calculated by taking the nominal revenue, multiplying by the factor from the appropriate element in the “yyyy Factor” column in TAB [IPD]. Summing the data in columns G-AC for rows 4-70 provides the data in row 71 for columns G-AC.
9. TAB [Revenue]. This spreadsheet shows the nominal revenues, in millions of dollars, contained in the Waybill Sample. These are revenues that are expanded based on the sampling rate. For example, revenues recorded for single-car shipments are multiplied by 40, as one in forty of single-car shipments are sampled in the process of collecting data for the Waybill Sample. This information is contained in a field of the Waybill Sample called “Total_Unmasked_Revenue”. This information is not available in the public version of the Waybill Sample data. The quantities in this spreadsheet are pasted as values based on computations performed in the database. However, the database computations aggregate revenue information into the categories used for analyzing the rate trends and does not contain information that is sufficiently detailed to reveal the rates charged for individual shipments. Numerical data in the sheet for rows 4-70 and for columns G-AC are obtained directly from the database computations with a conversion into millions of dollars. Summing the data in columns G-AC for rows 4-70 provides the data in row 71 for columns G-AC.
10. TAB [IPD]. This spreadsheet contains the annual measure of the Implicit Price Deflator (IPD) as an average for each year from 1985 through the last year from which Waybill Sample data are available. The IPD is a measure that will convert nominal dollars to constant dollars based on spending patterns associated with the Gross Domestic Product. The sheet also contains a column titled “yyyy factor” (where yyyy represents each of the years in the analysis). This factor is the IPD

for 2007 divided in turn by the IPD for each of the years in the list and then multiplied by 100.

11. TAB [Grain Graphs]. This sheet summarizes the real revenue and ton-miles for grain shipments based separately on the variables of car ownership, train type, and length of haul.
12. TAB [Coal Graphs]. This sheet summarizes the real revenue and ton-miles for coal shipments based separately on the variables of car ownership and length of haul.
13. TAB [Fuel \$]. This sheet sums the amount spent on fuel by each of the four major Class I railroads in the United States for the years 2004-2007. It then determines the amount of additional fuel expense for 2007 over 2004, both in nominal and real terms.
14. TAB [Cat Summaries]. This sheet provides a summary of how the data are divided into 67 categories for analysis.
15. TAB [STCC List]. The final sheet in the table shows the codes that are used for the different commodity groups analyzed in this study. The code that does not fit within existing STCC divisions is “Intermodal”. All shipments categorized as intermodal were removed from what they would otherwise be coded.
16. TAB [Summary_Statistics]. This spreadsheet shows the Revenue, Cars, Car-miles, Tons, Ton-miles, and Revenue per Ton-mile for each of the 67 categories of output analyzed in this study. The contents of this sheet are very similar to the contents TAB [Summary Stats]. The difference is that TAB [Summary Stats] contains data imported directly from the database and therefore does not contain information on all categories of output. Specifically, data for 1985 and 1986 are incomplete for the Grain and Coal commodities. The reason the data are incomplete for these combinations is that the Grain and Coal commodities are the only ones that divide outputs into categories based on car ownership information and that information is not available for 1985 and 1986. As a result, 13 rows of data are missing for each of those two years. TAB [Summary_Statistics] has been constructed to add these 26 missing rows, using quantities of zero for the output and revenue information in the “Private” car categories. We realize that there is a small distortion created by this process. However, the amount of distortion is very small and we have elected not to alter any of the results to account for this. The data on the number of Cars and Car-miles have been collected but not used in this analysis. Information on the computation of data in each of this sheet’s columns follows.
 - a. Sort. Data in this column are of the form yyyyNN, where yyyy represents the year and NN ranges from “00” to “66”, giving an index to the 67 output categories.
 - b. Year. This column contains the year number.

- c. STCC#. In most cases, this column shows the first two digits of the STCC group for the commodity group being represented. There are three exceptions: “00” is Intermodal; “02” is Grain; and “99” is All Other.
- d. STCC Name. This is the name for the STCC group number provided in the previous column.
- e. Dist Category. See the earlier description in paragraph “D”.
- f. Car Ownership. See the earlier description in paragraph “E”.
- g. Train Type. See the earlier description in paragraph “F”.
- h. Revenue (000s). The value in this column is transferred from TAB [Summary Stats].
- i. Cars. The value in this column is transferred from TAB [Summary Stats].
- j. Car-Miles. The value in this column is transferred from TAB [Summary Stats].
- k. Tons. The value in this column is transferred from TAB [Summary Stats].
- l. Ton-Miles. The value in this column is transferred from TAB [Summary Stats].
- m. RPTM. This value is computed by the database application by dividing the Revenue amount by the Ton-Miles amount. This value is posted to TAB [Summary Stats] from which it is retrieved by TAB [Summary_Statistics].
- n. Sort#. This value is in the form of yyyySSdCt, where...
 - i. yyyy is the year number.
 - ii. SS is the STCC#.
 - iii. d is the Dist Category.
 - iv. C is Car Ownership.
 - v. t is Train Type.

17. TAB [Summary Stats]. Each column in Summary Stats has an analogous column in TAB [Summary_Statistics] with the same name. The values in each TAB [Summary Stats] column and the analogous column in TAB [Summary_Statistics] are precisely the same, with the following exception: the STCC# for Grain is “XX” as opposed to “02”. As explained in the documentation for TAB [Summary Statistics], data in TAB [Summary Stats] are missing for Coal and Grain in 1985 and 1986 for private cars. This is because there is no car ownership information in the Waybill database for those two years.