



Transportation Research Forum

Leading Transportation Indicators: Forecasting Waterborne Commerce Statistics Using Lock Performance Data

Author(s): Mark A. Thoma and Wesley W. Wilson

Source: *Journal of the Transportation Research Forum*, Vol. 44, No. 2 (Summer 2005), pp. 33-44

Published by: Transportation Research Forum

Stable URL: <http://www.trforum.org/journal>

The Transportation Research Forum, founded in 1958, is an independent, nonprofit organization of transportation professionals who conduct, use, and benefit from research. Its purpose is to provide an impartial meeting ground for carriers, shippers, government officials, consultants, university researchers, suppliers, and others seeking exchange of information and ideas related to both passenger and freight transportation. More information on the Transportation Research Forum can be found on the Web at www.trforum.org.

LEADING TRANSPORTATION INDICATORS: FORECASTING WATERBORNE COMMERCE STATISTICS USING LOCK PERFORMANCE DATA

This paper develops and applies a forecasting model for transportation data based on the leading economic indicators literature. The specific application is to forecast river tonnages. Waterborne commerce data reflect tonnages of commodities moved on various rivers and in various directions. They are released after the Lock Performance Monitoring System data which reflect tonnages moving through specific locks. The model presented here forecasts waterborne commerce data from lock performance data. The results suggest that even a very simple model can provide precise forecasts.

by **Mark A. Thoma and Wesley W. Wilson**

INTRODUCTION

Market participants often need information about future market conditions at the earliest possible date. Such information is used for a variety of purposes. Transportation firms use forecasts to assess equipment needs both in total and in location, to negotiate rates, to assess cost recovery, and to make employment decisions. Government agencies use such data to gauge the quality of data collection efforts, validate trends in data, and provide information. Transportation statistics, however, are often released with a lag, and sometimes the lag is quite significant.¹

This paper develops a procedure for predicting future values of a variable of interest based upon its past history and upon leading indicators that appear either at a higher frequency or are released between the release dates of the variable of interest. This provides useful information to regulators and other market participants earlier than if they wait for the release of the actual data.

The example used to demonstrate the procedure, predicting waterborne commerce (WBC) annual tonnages from the Lock Performance Monitoring System (LPMS) data released between WBC data releases, illustrates the usefulness of the technique to market participants. Also, planning by lockmasters, the Army Corps of Engineers, and others is made simpler by having accurate forecasts of the WBC data as early as possible. The LPMS

data collected by the National Data Center are available much sooner than the WBC data. The Waterborne Commerce Statistics Center schedules the release of waterway tonnage statistics annually in September. The reports represent annual flows in total, upbound, and downbound for different rivers. The corresponding LPMS data, which measures tonnage passing through particular locks, are generally released in March.² When the LPMS data is used to predict the September WBC value, planners are given needed information much sooner.

The prediction problem faced by forecasters using LPMS data to predict WBC data is analogous to the problem faced by macroeconomic forecasters and can be solved in a similar manner. In particular, macroeconomic forecasters are often confronted with the problem of forecasting future values of output, inflation, unemployment, and other macroeconomic variables from data on variables available prior to the release of the variable of interest. In doing so, macroeconomic forecasters have developed a list of leading economic indicators (LEI) released monthly by the Conference Board.³ Upon release, these variables are used to predict future values of key macroeconomic variables using time-series techniques.

The approach used in this paper is analogous to that of the macroeconomic forecasters. In the prediction of WBC data, the LPMS data play the same role as the LEI's play in the prediction

of macroeconomic variables. Indeed, the LPMS data contain information on tonnages that pass through the locks of waterways, while the WBC data reflect total tonnages moved on the waterway. In fact, for a stylized simple network, it can be shown readily that the data can be perfectly correlated. In more complex settings, the data are correlated, but the correlation is not perfect. Thus, the LPMS data gives a nearly ideal LEI for WBC. This paper shows that the time-series techniques employed by macroeconomic forecasters can be used to generate forecasts of WBC data based upon LPMS data.⁴

The model described and used below is developed with an eye toward applicability to other river systems and bi-directional movements as well as to other transportation data series. Of course, the number of variables and the length of the data set used to predict a specific variable is limited only by the amount of data at the econometrician's disposal. In the present case, because of significant changes in data collection that resulted in a lack of consistent data over time, the analysis is limited by a short time period data set. As a result, the procedure developed here is as simple as possible without sacrificing predictive power. The aim is to provide a procedure that is relatively easy to duplicate for many river systems, and can be easily amended to integrate other features e.g., directional movements, commodity movements, etc., into the model. To help in this regard, the procedure is described with specific steps at the end of the paper.

There are a number of studies that forecast transportation data.⁵ Babcock and Lu (2001) address what they describe as a neglected area of water transportation forecasting, the short-term forecasting of inland waterway traffic. Their paper builds an ARIMAX forecasting model for grain tonnage on the Mississippi River and finds that the model provides accurate forecasts. Tang (2001) develops a time-series forecasting model for grain tonnage on the McClellan-Kerr Arkansas River. As she notes, data problems preclude the development of a structural economic model for forecasting, but given the focus on short-term forecasting, a time-series model is an attractive alternative. She finds that the models perform well so long as care is taken to identify and model structural breaks in the data. Thoma and

Wilson (2004a) use vector autogressions and variance decompositions to examine weekly transportation data with a specific purpose of examining shocks to barge quantities and rates from changes in ocean freight rates, and rail rates and deliveries. Thoma and Wilson (2004b) model co-integrating relationships between river traffic, lock capacities, and a demand measure over an extended time period (1953-2001). The co-integrating relationship is then used to develop forecasts of river traffic.

This paper also uses time-series models to overcome data and other problems that preclude estimation of a full structural model, but goes beyond the various forecasting models by adding the ability to update forecasts based upon the release of a leading indicator that appears prior to the release of data on the variable to be forecast. Specifically, the difference of this model from the forecasting models above is that the variable to be forecasted (WBC_t), is forecasted with a contemporaneous variable (LPMS_t). Its primary purpose is to use a highly correlated variable that is released earlier to forecast another variable's level for the same time period. The forecasting models that have appeared previous to the one used in this paper estimate a model based on theoretical and/or time-series properties and then forecast future values based on data up to the current time period.

RELATING LPMS AND WBC DATA USING KEY LOCKS

LPMS data is used as an economic indicator for WBC data. As noted earlier, the former measures tons locked at a specific location (or set of locations), while the latter measures total tons that originate, terminate or travel on a river.⁶ In this paper, the WBC series used is tons traveling on the Mississippi River in total (all commodities), food and farm product tons, and coal tons. In this section, a stylized example is provided to illustrate the data collection efforts for each variable and how the LPMS and WBC data are connected. However, it is important to note that the LPMS data and the WBC data measure different quantities.

In the example, there is a river that has five locks labeled L1, L2, L3, L4, and L5. There are 10 barge loads that move on the river during

the time period that tonnages are measured. The tonnages and movements between locks are:

- Load 1 20 tons through locks L1, L2, and L3
- Load 2 50 tons through locks L1 and L2
- Load 3 30 tons through locks L1, L2, L3, L4, and L5
- Load 4 10 tons through locks L3, L4, and L5
- Load 5 15 tons through locks L2, L3, and L4
- Load 6 80 tons through locks L4 and L5
- Load 7 25 tons through lock L3
- Load 8 35 tons through locks L2 and L3
- Load 9 45 tons through locks L4 and L5
- Load 10 60 tons through locks L1, L2, L3, L4, and L5

The WBC data reflect the sum of all loads (in tons) moved on the river. Thus, the WBC measurement is $20+50+30+10+15+80+25+35+45+60 = 370$. The LPMS data are tons passing through each lock. To see how the LPMS data are calculated in this example, the data are organized as in Table 1.

The totals for each individual lock are the LPMS measurements. Notice that there are two local peaks in the LPMS measurements, L2 at 210 and L4 at 240. Notice also that lock 5 at 225 has a larger volume than lock 2 at 210, but because these two locks are geographically

distinct, L2 will be the best indicator in its local area and L4 the best in its region as they capture the most local volume. These locks are defined as key locks for this river, i.e. those locks that handle the most traffic in particular regions on the river.

There are five potential leading indicators, L1 through L5. Because of data limitations and for other reasons noted below, parsimonious representations of the data are desirable when constructing forecasting models. Thus, it is desirable to focus on the most informative leading indicators. Importantly, changes in overall tonnage moving on the river as measured by the WBC data are most likely to be reflected at key locks. In the example above, there is only one load, Load 7, which does not pass through either key Lock 2 or key Lock 4. Thus, any change in tonnages on any load except Load 7 will be captured as a change in tonnages at one of the two key locks, or in the case of Loads 3, 5, and 10, a change in tonnages at both of the key locks.

This example is contrived, but the point that changes in tonnages through key locks are useful for capturing changes in overall tonnages moving on the river is true generally. This supports the use of data from key locks in the LPMS to predict the value that will be released in the WBC data in September.

Table 1: LPMS Data Example (tons)

Lock	L1	L2	L3	L4	L5
Load 1	20	20	20		
Load 2	50	50			
Load 3	30	30	30	30	30
Load 4			10	10	10
Load 5		15	15	15	
Load 6				80	80
Load 7			25		
Load 8		35	35		
Load 9				45	45
Load 10	60	60	60	60	60
Totals	160	210	195	240	225

The identification of key locks can be achieved through a combination of methods. First, publications such as the *Grain Transportation Report* from the Transportation Services Branch of the U.S. Department of Agriculture lists key locks to track grain movements. The U.S. Army Corps of Engineers Navigation Data Center has a *Key Lock Report* and thus provides another list of key locks. For example, these two publications taken together identify locks 15, 25, 26 and 27 on the Mississippi River and lock 8 on the Illinois Waterway as key locks for Mississippi river tonnages (the *Grain Transportation Report* lists locks 15, 25, 26, and 27 while the Navigation Data Center lists locks 25 and 27, and both list Illinois Waterway lock 8). These sources can then be buttressed with statistical methods if desired.⁷ For example, examining the correlation between the LPMS data for each lock on the Mississippi and Illinois Rivers with the WBC tonnages for the Mississippi River to identify which locks have the highest correlation supports the choice of locks 15 (.8705), 25 (.7899), 26 (.8800), and 27 (.8533) on the Mississippi River and lock 8 (.8600) on the Illinois Waterway as key locks (the correlation coefficients are in parentheses).

Finally, note that the tonnage data are very similar for locks 25, 26, and 27 (i.e. the correlation coefficients are .8417 for locks 25 and 26, .7408 for locks 25 and 27, and .9556 between 26 and 27). Because the data set is small, only 12 annual observations (1990-2001), degrees of freedom are an important consideration. Finding the minimum set of variables needed to predict the WBC data can enhance the precision of the estimates and provide better forecasts. To decide among these three locks, correlations were ex-

amined, and regressions of the WBC data on the locks data independently and as a set were also examined. The conclusion was that locks 25 and 27 are representative of all three locks, though the particular choice has very little effect on the predictive power of the model developed below. Another consideration in this choice comes from the correlations between locks 25, 26, and 27 which showed that the lowest correlation, .7408, was between locks furthest apart, locks 25 and 27. Thus, including these two locks adds the variables with the most independent information. Note that this is the set used in the Navigation Data Center report which also influenced our choice to focus on this subset of locks. Thus, the set of key locks to be examined is lock 8 on the Illinois Waterway, and locks 15, 25 and 27 on the Mississippi River.

THE FORECASTING MODEL

The goal of the forecasting model is to predict WBC data, as reported in *Waterborne Commerce of the United States*, using LPMS statistics. To this end, the WBC data used here are Mississippi River total tons from 1990-2001 measured in millions of tons, as well as tons of coal and tons of food and farm products.⁸ These variables, which are the variables to be forecasted using key locks identified above, reflects total tonnages and two commodity tonnages on the Mississippi River System. The key lock data are taken from the Summary of Lock Statistics provided by the U.S. Army Corps of Engineers Navigation Data Center's *Key Lock Report*.⁹

Given these data, the basic form of the model used to predict the WBC data release is:

$$(1) \quad WBC_t = \beta_0 + \beta_1 WBC_{t-1} + \beta_2 L8_t + \beta_3 L15_t + \beta_4 L25_t + \beta_5 L27_t + e_t \cdot$$

WBC = Mississippi River total tons, food and farm tons, or coal tons

$L8$ = total tons passing through lock 8 on the Mississippi

$L15$ = total tons passing through lock 15 on the Mississippi

$L25$ = total tons passing through lock 25 on the Mississippi

$L27$ = total tons passing through lock 27 on the Mississippi

In time-series forecasting models, the lagged value of the dependent variable, i.e. the lagged value of WBC on the right-hand side in this model, summarizes all information up to time $t-1$. The lock variables capture any new information in the LPMS data release.¹⁰ Note that the left-hand side variable WBC and the right-hand side variables on lockages are both dated time t , indicating that the data are for the same year. However, as already noted, the data are not released at the same time. Since the LPMS data are released earlier than the corresponding WBC data, the LPMS lockage variables can be used to predict the WBC variables.

More generally, other explanatory variables can be included in the model as well, but, given the small number of observations that are available, this alternative is not explored here.¹¹ With a larger annual data set or a large number of monthly observations, it would be possible to include additional variables to enhance the predictive power of the model. For example, with a large monthly data set a model such as 2 can be estimated.

$$(2) \quad WBC_t = \beta_0 + \sum_{i=1}^P \beta_{1i} WBC_{t-i} + \sum_{i=0}^Q \beta_{2i} Locks_{t-i} + \sum_{i=0}^R \beta_{3i} OtherVars_{t-i} + e_t$$

In this model, P lags of the dependent variable are included to summarize past information, a vector of current and Q lags of lockages are included to capture the information contained in the LPMS release, and a vector of the current and R lags of other variables that are helpful in producing forecasts is also included.

This is a fairly general specification as it allows for the inclusion of current and lagged tonnages of all key locks, a p^{th} order lag structure, and current and lagged values of any other variables helpful in forecasting the WBC data.¹² However, this is not necessarily the model that predicts best. To find the model that has the most predictive power, it is important to find the most parsimonious representation of the data.

When the goal of econometric estimation is to test theoretical propositions and to discover the values of important theoretical quantities, it is important to have unbiased estimators. Thus, including too many variables, which in general only reduces the power of test statistics but does not bias parameter estimates, is preferred to in-

cluding too few variables in the model, which can bias parameter estimates. Thus, the tendency is to include too many rather than too few variables when there is uncertainty as to the best empirical specification of a model.

The situation is different when a model is used for forecasting. Here, the goal is to produce forecasts as close as possible to the actual values and, if large gains in forecasting efficiency can be obtained by accepting small bias in estimation, such tradeoffs are generally adopted. Including too many variables in a model, i.e. simply “throwing in the kitchen sink” can produce poor forecasts even if no biases are introduced in the process. This is because the extra variables will, to some extent, explain the in-sample variation in the dependent variable even if, in fact, the variables are unrelated simply due to statistical chance causing the two variables to co-vary. This means that out-of-sample forecasts will be inappropriately sensitive to these variables and the forecasts will deteriorate. When the goal is precise forecasts, it is desirable to produce as parsimonious a model as possible to avoid this

pitfall. This is particularly true when, as in the case investigated here, where the degrees of freedom in estimation are small. Thus, effort is devoted to finding the model that predicts the best with the smallest number of right-hand side variables.¹³

To accomplish this for the model used here, various permutations of the locks are included in the model and the out-of-sample predictive power of each model permutation is examined for 1997-2002. That is, the model is estimated using data from 1990 through 1996. Then the model is used to predict the 1997 WBC value from the 1997 LPMS data release. Next, estimate the model using data through 1998 and predict the 1999 WBC value, and so on. Then, after all five values have been predicted, calculate the percentage forecast error at each point in time and the mean-squared error.

As an example, when total tons are examined, a comparison of a model including all four lock variables with one that drops the data for Illinois Waterway lock 8 reveals that

the model that drops the data for lock 8 predicts better out-of-sample than a model including lock 8 according to the mean-squared out-of-sample predictive error criterion. Interestingly, examination of mean-squared prediction errors for other models used to predict total tons reveals that models with locks 15, 25, and 27 individually outperform models with various combinations of locks, e.g. a model with any of these variables alone on the right-hand side outperforms a model with both locks 25 and 27 on the right-hand side. Among the models with locks 15, 25, and 27 entering alone, the mean-squared prediction errors are very similar and the choice is not critical, but strictly adhering to the minimum predictive error criterion, the following model predicts total tons the best among the permutations examined:

$$(3) \text{ } WBC_t(\text{Total}) = \beta_0 + \beta_1 WBC_{t-1} + \beta_2 L15_t + e_t .$$

For the model with tons of food and farm products the best model is:

$$(4) \text{ } WBC_t(\text{Food and Farm}) = \beta_0 + \beta_1 WBC_{t-1} + \beta_2 L25_t + e_t ,$$

and for the model with tons of coal as the dependent variable the best model is:

$$(5) \text{ } WBC_t(\text{Coal}) = \beta_0 + \beta_1 WBC_{t-1} + \beta_2 L27_t + e_t .$$

ESTIMATED MODEL AND FORECASTS

Ordinary least squares was used to estimate the coefficients of the three models (6, 7, 8) shown below. There is no evidence of significant autocorrelations in the estimates.¹⁴

These estimates are for the full sample.¹⁵ As noted above, to examine the ability of this model to predict out-of sample,¹⁶ first estimate the model using data from 1990 through 1996, then use the estimated model to predict the September WBC value for 1997 using the 1997 value of the LPMS lock data released in March. Next, estimate the model using data through 1997 and use the 1998 value from the LPMS to predict the 1998 WBC value, and so on. The following graphs show the outcome of the procedure for the three models.

There are three lines shown in each graph. The line denoted with squares is the actual annual series for the WBC data. The line with diamonds is the in-sample predicted values based upon estimates from the entire sample. The line with triangles is the out-of-sample predictions obtained as described above.

The model predicts the WBC data out-of-sample fairly precisely in the first two cases, less so in the third, but in all three cases the forecast errors are moderate indicating that the March LPMS key lock data capture the essential

$$(6) \text{ } \hat{WBC}(\text{Total}) = 146.1 + 0.763 * WBC_{t-1} + 0.001 * L15$$

(1.14) (4.04*) (0.653)

$$R^2 = .70 \quad D.W. = 2.29$$

$$(7) \text{ } \hat{WBC}(\text{Food and Farm}) = 80.52 + 0.160 * WBC_{t-1} + 0.002 * L25$$

(1.41) (.576) (2.76*)

$$R^2 = .50 \quad D.W. = 2.49$$

$$(8) \text{ } \hat{WBC}(\text{Coal}) = 106.5 + 0.404 * WBC_{t-1} + 0.00001 * L27$$

(1.36) (1.34) (0.161)

$$R^2 = .19 \quad D.W. = 2.17$$

Figure 1: Comparison of Actual WBC Tons to Forecast WBC Tons - Total

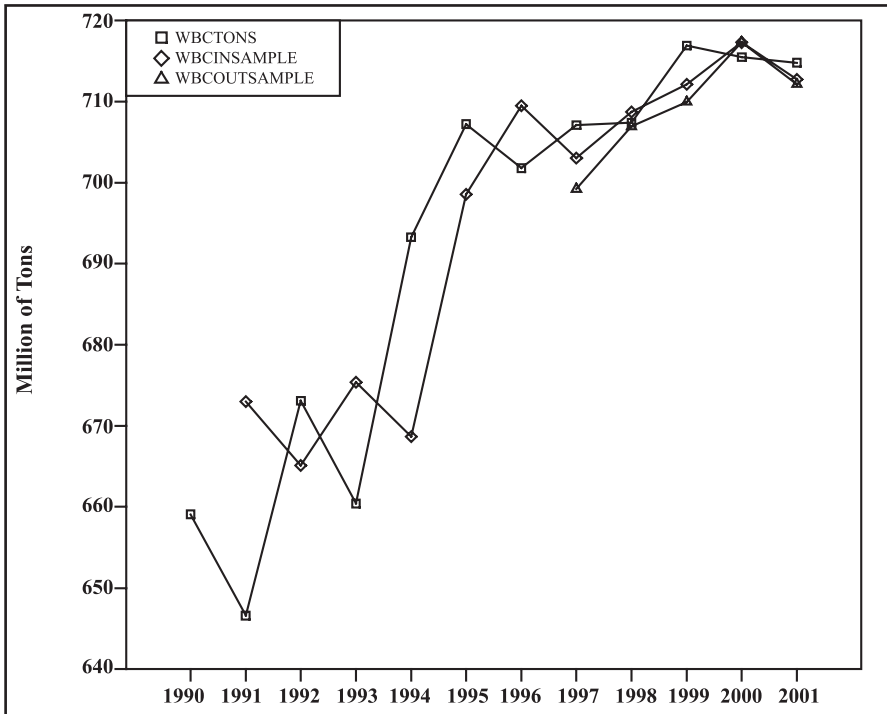


Figure 2: Comparison of Actual WBC Tons to Forecast WBC Tons – Food and Farm

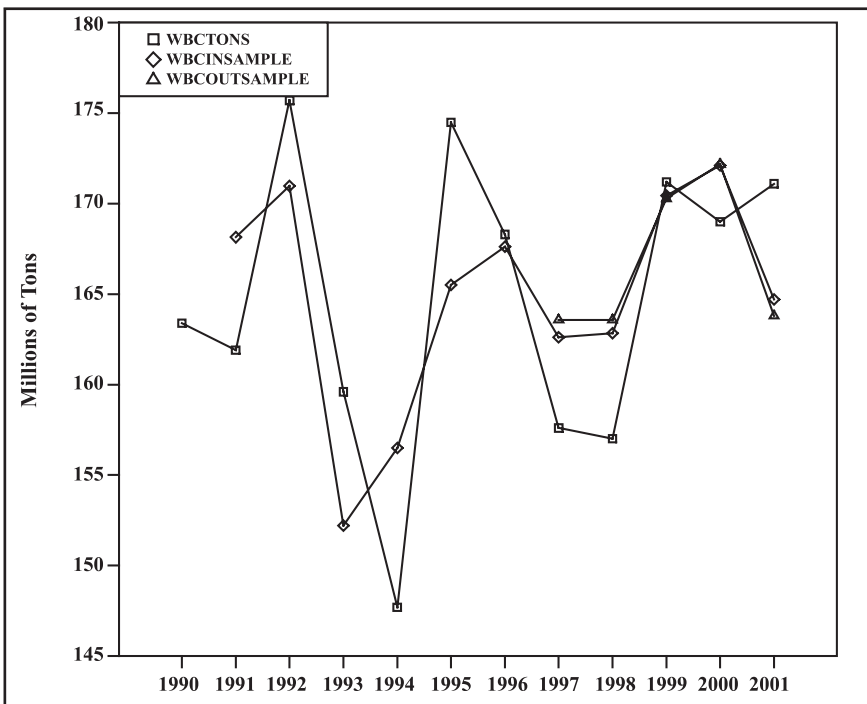
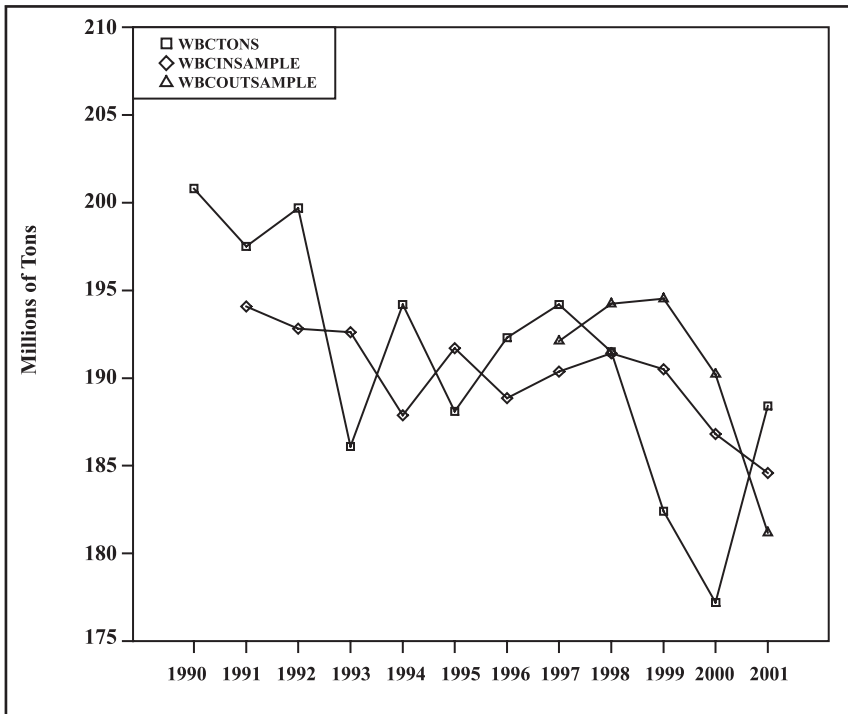


Figure 3: Comparison of Actual WBC Tons to Forecast WBC Tons – Coal



characteristics of the WBC release in September. More formally, the precision of the forecasts can be measured by examining, in percentage terms, the out-of-sample forecast errors. These are shown in Table 2.

For total tons, in all years the forecast error is less than or equal to 1.1% and in three years the error is smaller than 1%.¹⁷ Because the units of measurement are millions of tons, and because the mean value of the WBC variable is between 600 and 700 million tons, a 1% forecast error represents a deviation of 6-7 million tons

out of the 600-700 million tons that is moved. This indicates that the forecasts are relatively precise. The remaining two columns have larger percentage errors, as would be expected when forecasting individual commodities rather than the total, but in no case is the forecast error larger than 7.5% and in all but two cases, both for tons of coal, the error is less than 5%.

For total tons, even though the Lock 15 variable is insignificant in the regression results (the t-statistic is 0.65 in the results given above), dropping this variable causes deterioration in the

Table 2: Forecast Error Percentages and Mean Square Errors

Year	% Error Total	% Error Farm	% Error Coal	Mean Square Error Total	Mean Square Error Farm	Mean Square Error Coal
1997	1.1	-3.8	1.1	1.2	14.4	1.2
1998	0.1	-4.2	-1.4	1.2	32.0	3.2
1999	1	0.5	-6.6	2.2	32.3	47.4
2000	-.3	-1.9	-7.3	2.2	35.8	101.4
2001	0.4	4.3	3.8	2.4	54.0	116.0

mean-squared forecasting error, i.e. the MSE for 2001 increases from 2.4 to 3.3.¹⁸ Dropping the lagged value of the WBC variable from the model causes the MSE for 2001 to increase substantially to 59.1. Thus, both the lagged value of the WBC variable and the LPMS variable are needed to produce the most accurate forecasts. This is true for the individual commodity series as well. As noted above, the inclusion of the lagged WBC value places the prediction at last year's value times the estimated level of regression coefficient towards the unconditional mean (e.g., if the coefficient on the lagged value of WBC is 0.7, then the base value of the forecast this year is 0.7 of the value last year) and the inclusion of the LPMS variable updates this base value to reflect current conditions.

SUMMARY OF PROCEDURE

A summary of the procedure for using the LPMS data to predict the WBC data is:

1) For the river system under consideration, identify key locks. This is accomplished by examining sources such as the *Grain Transportation Report* which lists key locks to track grain movements or the U.S. Army Corps of Engineers Navigation Data Center which also list key locks for various river systems. Then, use additional knowledge of the river, its geography (e.g., if there is an essential tributary, locks nearby could carry important information), and statistical techniques such as the correlation of each lock with the WBC data to identify a set of key locks. If data are limited, the set of key locks may need to be further refined into a small set of essential locks.

2) If there are enough observations to support such an investigation, add additional explanatory variables known to affect tonnage on the river to the empirical model.

3) Given the set of key locks to be included and any additional variables thought to be useful for prediction, specify a model as general as possible

given the available data, e.g., one possibility is equation 2. Then examine permutations of the model, i.e. various combinations of explanatory variables. Determine which model predicts best out-of-sample and has the smallest number of right-hand side variables.

4) When the earlier value of the LPMS data is available, use the model identified in step 3 to predict the WBC value.

CONCLUSION

This paper develops a model to predict WBC data based upon the release of LPMS data earlier in the year. The model developed here is based upon macroeconomic forecasting models involving leading economic indicators to forecast key macroeconomic variables. In the application here, the LPMS data is a leading indicator for WBC data released later in the year and once the variable is interpreted in this way, time-series forecasting techniques used to forecast macroeconomic variables can be employed.

The paper provides a fairly general forecasting model and demonstrates its usefulness by examining a specific example for the Mississippi River. Even with the relatively small number of observations used in the Mississippi River forecasting example, the model produces forecasts with a relatively small out-of-sample mean square error. For the purpose of forecasting tonnage of other rivers, directional traffic, and movements for different commodities, the approach is easily adapted. This type of approach can be used in a wide variety of other settings. Indeed, there are a variety of data elements in transportation that are published, but only after a significant lag. As this paper demonstrates related variables that are available earlier can be used to generate forecasts. Development of such transportation indicators is an area for further research.

Endnotes

1. Waterborne commerce (WBC) data for a given year is scheduled for release in the following September (8 months), but it is not unusual for the WBC data to lag by as much as two years. In other data, the lag for rail waybill statistics can be up to two years. The Commodity Flow Survey is conducted only every five years and is released with an approximate 2 ½ year lag. In addition, the U.S. Department of Transportation (USDOT) collects, processes, and releases lots of other series. Some of these are of higher frequency and are released with lag. To the extent that forecasts of these series are needed, the approach in this paper can be easily modified to other series. Examples are provided at http://www.bts.gov/upcoming_data_releases/.
2. For some rivers, the data are available on-line at <http://www.mvr.usace.army.mil/mvrimi/omni/webrrpts> and provide nearly instantaneous figures.
3. The Leading Indicators from the Conference Board are available at <http://www.conference-board.org/economics/indicators.cfm>.
4. Leading indicators were developed by Burns and Mitchell (1946). For further discussion of leading economic indicators and a list of additional references, see Stock and Watson (1990).
5. Most cost/benefit studies of transportation capital decisions require forecasts of future traffic. For a recent example in the case of waterways, see Jack Fawcett and Associates (2000) at http://www.mvr.usace.army.mil/pdw/nav_study/JFAreport.pdf. Further, the USDOT releases a series of forecasts of transportation data and use forecasts of some indices in the construction of its Transportation Services Index (TSI). The TSI index, methodology etc. can be found in www.bts.gov/xml/tsi/src/index.xml. Other studies have used co-integration techniques to develop long-run forecasts of the inland waterway traffic over an extended period (Thoma and Wilson, 2004a) and use vector autoregressions (VARs) and variance decompositions to develop short-run forecasts using weekly data (Thoma and Wilson (2004b). See <http://www.corpsnets.us>.
6. The LPMS data are recorded at each lock. The WBC data are the result of tow company surveys. The surveys provide detailed information on vessel operations during a month. This information includes the originating and terminating docks. The Waterborne Commerce Center aggregates these data and releases information specific to a river (e.g., Mississippi) and various sub-aggregates, e.g. direction (upstream/downstream) and commodity. The LPMS data record the tons moved through specific locks as well as a number of other variables. The website <http://www.iwr.usace.army.mil/ndc/index.htm> fully describes both the Waterborne Commerce and the Lock Performance Monitoring System data.
7. Such approaches include (i) identifying key locks as those with the most lockages or tonnages passing through, (ii) using pairwise correlations between the WBC data and the lock tonnages, and (iii) regressions of WBC data on all possible sets of LPMS lock data and choosing the set that yields the best forecasts either in or out-of-sample, (iv) step-wise regression.
8. These data are taken directly from Table 3-1 of the website <http://www.iwr.usace.army.mil/ndc/wcsc/pdf/wcusnatl02.pdf>.
9. The WBC data on this website are available from 1983-2002. The LPMS data are only available from 1990-2001.

10. A discussion of autoregressive models and their interpretation is in Hamilton (1994).
11. These may include, for example, factors that shift supply and demand curves such as harvested acres of corn, industrial production, fuel prices, and directional variables, or other variables that are potentially useful in predicting WBC tonnages.
12. A slightly more general version of this model is a VAR model involving the variables in equation 2.
13. Milton Friedman (1953) notes that the only true test of a model is its ability to predict out-of-sample.
14. The t-statistics are in parentheses, and a * represents statistical significance at the 5% or better level.
15. One period is dropped due to the lagged dependent variable so the estimation starts in 1991 and uses 11 observations.
16. Other measures of forecasting performance, e.g. a sign test, can also be examined, but given the small number of observations and the goal of producing a model that can be used to guide the development of more complete models once additional data are available, such measures are not presented here.
17. For comparison purposes, the MSE values when lock 27 is used in place of lock 15 in the forecasting equation are 2.2 for 1997, 2.2 for 1998, 2.7 for 1999, 3.0 for 2000, and 3.0 for 2001 which, though higher in every case than when lock 15 is used, are still relatively small.
18. The 2001 value is presented because it is the average value of all of the out-of-sample forecasts.

References

- Babcock, Michael W. and Xiaohua Lu. "Forecasting Inland Waterway Grain Traffic." *Transportation Research Part E* 38 (1), (2002): 65-74.
- Burns, A.F. and W.C. Mitchell. *Measuring Business Cycles*. National Bureau of Economic Research, New York, 1946.
- Friedman, Milton. "The Methodology of Positive Economics." Milton Friedman ed. *Essays in Positive Economics*. Chicago: University of Chicago Press (1953): 3-43.
- Jack Faucett and Associates. "Review of Historic and Projected Grain Traffic on the Upper Mississippi River and Illinois Waterway: An Addendum. Institute for Water Resources, U.S. Army Corps of Engineers, 2000. Available at http://www.mvr.usace.army.mil/pdw/nav_study/JFAreport.pdf.
- Hamilton, James D. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.
- Mitchell, W.C. and A.F. Burns. *Statistical Indicators of Cyclical Revivals*. National Bureau of Economic Research, New York, 1938.

Stock, James H. and Mark W. Watson. "New Indexes of Coincident and Leading Economic Indicators." NBER Working Paper No. R1380, April 1990.

Tang, Xiuli. "Time Series Forecasting of Quarterly Barge Grain Tonnage on the McClellan-Kerr Arkansas River Navigation System." *Journal of the Transportation Research Forum*. Published in *Transportation Quarterly* 40(3), (2001): 91-108.

Thoma, Mark A. and Wesley W. Wilson. "Market Adjustments Over Transportation Networks: A Time Series Analysis of Grain Movements on the Inland Waterway System." Institute for Water Resources, U.S. Army Corps of Engineers, 2004a. The papers are available at <http://www.corpsnets.us>.

Thoma, Mark A. and Wesley W. Wilson. "Long-Run Forecasts of River Traffic on the Inland Waterway System." Institute for Water Resources, U.S. Army Corps of Engineers, 2004b. The papers are available at <http://www.corpsnets.us>.

US Army Corps of Engineers, Navigation Data Center. *Lock Performance Monitoring System Key Lock Report*, 1990-2001.

United States Department of Agriculture, Transportation Services Branch. *Grain Transportation Report*, 1990-2001.

Acknowledgements

This research was funded by the Navigation Economics Technologies Program (NETS). Research assistance from Paul M. Thoma is gratefully acknowledged as are comments and discussions with the Waterborne Commerce Statistics Center. In particular, we appreciate the comments of Jay Wieriman, Amy Tujague, James Lambert and Keith Hofseth who read earlier drafts of this work. The comments of anonymous referees and the editor are also much appreciated.

Mark Thoma received a Ph.D. in economics from Washington State University in 1985. From 1986 to the present he has been at the University of Oregon, and was department head from 1995-2000. He currently conducts research and teaches in the areas of macroeconomics and time series.

Wesley W. Wilson received a Ph.D. in economics from Washington State University in 1986. From 1986-1989, he was an assistant professor of agricultural economics. Since 1989, he has been at the University of Oregon. He was the president of the agricultural chapter of the TRF in 1986 and of the Transportation and Public Utilities Group of the AEA in 1999-2000. He currently splits his time between the University of Oregon and the Institute for Water Resources.