Ex _____(LAS-5)

FORECASTING SOLID WASTE TONNAGE: TECHNIQUES AND ALTERNATIVES[1]

by
Lisa A. Skumatz, Ph.D.
Principal
Skumatz Economic Research Associates
1511 Third Avenue Suite 1018
Seattle, Washington 98101

I. Introduction

Solid waste agencies require forecasts of the amount of tonnage that will need to be managed over the future in order to support a wide array of activities including:

- comprehensive planning
- program design and planning
- financial and rates calculations
- facilities and staffing planning, and
- other activities.

Historically, solid waste agencies have determined the amount of tonnage they may expect by either assuming last year's tonnage figure, or by "trending" the last few years of tonnage. Recently, however, these estimates of tonnage have become much more important links in determine the need for facilities, programs, and in determining revenues, costs, and cost-effectiveness of waste management options. New, more flexible, robust, and reliable techniques are needed.

This study summarizes the main approaches that can be used for forecasting solid waste tonnage, and provides information on the application of these approaches in communities across the nation. The strengths and weaknesses of the approaches are discussed, as well as issues related to the specification of the model and the variables, performance, and flexibility for the needed approaches.

[1]Some of the work in this paper draws from work included in Skumatz, "Econometric Findings in Solid Waste: Demand, Customer Choice, and Reactions to System Change", Yale Working Paper, draft, December, 1992; Skumatz and Breckinridge, "Handbook for Solid Waste Officials", EPA Document 910/9-90-012a-b, June 1990; and work by the author conducted for Oak Park, Cincinnati, and a solid waste client in the western U.S.

II. Basic Forecasting Methods

There are three basic forecasting approaches that can be applied to solid waste. These include:

- time series modeling: using a time series of *solid waste tonnage alone* to forecast future solid waste tonnage levels.
- "scaling up" or engineering approach: determining a pattern for per-capita or per-employee generation, and then multiplying by forecasts of population or employment to calculate solid waste tonnage.
- econometric methods: developing an equation that includes a number of economic, demographic, weather, or other variables to estimate the solid waste tonnage.

Time Series Approach: In the time series approach, tonnage is estimated as a function of historical tonnage levels. The form of the model allows factors like trends in tonnage over time, monthly or seasonal patterns, and other underlying historical tonnage patterns to provide an estimate of the tonnage in the next periods. It does not include other variables beyond cycles in tonnage -- no variables like economics or price are included in these models. This approach is used in the electric industry for very short term forecasts, and can be implemented thorough Box-Jenkins approaches (which have fairly long data series needs), exponential or *ad hoc* smoothing methods (including Winter's approaches) which have shorter data series needs, and can support models that include separate estimations of a level, trend, and seasonal components. These approaches can be implemented at very low cost[2]. They can provide good short term forecasts, but generally provide no information about the effects of exogenous policy changes or relationships between dependent and independent variables.

"Scaling Up" Method: In the "scaling up" method, a per-capita or per employee waste generation (or disposal) figure is estimated or assumed[3], and then local forecasts of population or employment are used to "scale up" the estimate to the wasteshed. In some cases, differing assumptions are made about per-capita generation for specific subsets of the population (lower income, single family vs. multi-family, etc.), but the basic structure of the forecast approach remains. These methods are by far the most commonly applied by communities estimating waste generation (or disposal) for comprehensive planning documents.

---

[2]Data requirements are relatively low, and many computer packages can implement these approaches practically automatically.

[3]Some communities use the EPA projections.

The "scaling up" approach is simple, logical, and can yield results with limited data. In particular, it requires only a forecast of future population or other readily-obtained data in order to develop projections. Historical series of tonnage and population can be used to tailor or "benchmark" national forecasts of tonnage per capita or per employee. Information from waste composition studies can easily be combined with the models, and forecasting can be done using spreadsheets. Scenario using different assumptions are easily run. However, the approach has several key weaknesses. Explanatory power is very limited. Scenarios can only be run assuming differences in population, or changes in waste generation growth with very little insight into the impact of specific causal factors. Work across different cities has shown little consistency of it, with differences in definition, impacts of seasonal factors, or interpretation and implementation differences a factor. Historically, the approach has been very widely and successfully applied; however, its ability to forecast accurately during times of increasing recycling, education efforts, changing demographics, and other changes to waste generation and disposal habits appears to be breaking down.[4]

Econometric Method: An econometric approach to forecasting assumes an underlying relationship between the tonnage to be forecast and one or more explanatory variables. Generally, the variables include ones that are expected to have an effect on the amount of tonnage disposed or generated (depending on the variable being estimated), including economic factors (including price), demographic factors, and weather information. This approach, although less commonly used in solid waste, has generated a high degree of satisfaction with performance among its users. The equations have generally provided good explanatory power and are intuitive to explain. The method can explicitly account for business cycles and explicitly incorporate specific impacts of weather, prices, and other factors. The results have been reasonably similar across jurisdictions, indicating that the approach is fairly robust. The method supports construction of confidence intervals around forecasts, and is particularly strong for conducting scenario analysis, allowing a wide range of "what if" alternatives. However, the econometric approach has several disadvantages. The data requirements are more extensive, and it may be that some preferred data are not readily available, either historical series, or as independently forecasted series from which to generate tonnage forecasts. A higher level of support in terms of technical staff capabilities and computer equipment may also be needed. In addition, the modeling and the incorporation of judgmental factors into the model is more complicated. These types of models are commonly used in the electric, water, and other utility applications for medium- and long term modeling.

[4]A variation on this approach, end-use or conditional demand approaches is used for models in the electric utility industry. In these models, the amount of energy used per appliance is estimated, and then the percent of homes or businesses with these appliances is estimated and overall forecasts generated.

Strengths and Weaknesses of the Three Approaches

Each of these models provides useful forecasts, and each has strengths and weaknesses.

Time series approach:

Strengths

- very low data requirements
- inexpensive, quick, easy to generate new forecast

Weaknesses

- does not incorporate information about economic cycles, demographic changes, etc.
- little explanatory power or insight -- no information about the effects of exogenous policy changes or dynamic relationships among important variables are provided.
- not appropriate with structural changes
- useful only for short term forecasts

"Scaling up" approach:

Strengths

- allows limited causal explanatory power
- indirectly incorporates business cycles, demographic changes
- relatively low data requirements
- relatively easy to generate forecast
- inexpensive to model
- provides "micro", or "bottom-up" based forecast

Weaknesses

- does not explicitly explain factors underlying per-capita usage rates
- does not incorporate weather or other similar factors

Econometric approach:

Strengths

- good capability to provide explanatory power
- flexible modeling approach that can incorporate array of variables including policy variables or factors that might drive structural changes
- incorporates economic, demographic, weather, and other factors explicitly
- provides good medium-to longer term forecasts -- may not fit short run as well

- provides "top down" forecast or more "macro" approach

Weaknesses

- can be onerous data requirements
- more complicated modeling requirements and data may not be available to provide good fit
- more complicated to generate new forecasts

III. Modeling Issues and Approaches

Estimation of solid waste tonnage provides a number of complicating issues, many of which can be addressed through specific techniques from the literature or from the experience of energy utility forecasting work. Many of the problems below apply to all the modeling approaches, however, the econometric model, with its flexibility, has the greatest ability to deal with the issues, and the topics are, therefore, discussed in that general context.

- Selecting the dependent variable: The dependent variable could be specified as overall tonnage generation, tonnage generation by key sectors (residential, commercial, etc.), or disposal overall or by sectors. Generally, the preferred model would provide multiple models -- one to explain generation, and another to explain tonnage diverted into programs and the difference would provide an estimate of tonnage disposed. This is because it allows the economic variables to be most matched with the point in the customer's waste management decisions. However, most communities lack data on generation, and therefore, most communities develop models to forecast waste disposal.

- Selecting the independent (or explanatory) variables: The modeler will want to consider variables that would be expected to influence the amount of waste generated or disposed. A number of types of drivers have been found to be consistently significant (see table below) across jurisdictions. For example, if the tonnage being modeled is residential, variables such as price, population, household size, income, and weather factors have been found to be important. Non-residential tonnage may logically be related to employment, measures of income or production, housing starts, weather factors, and other variables. In some cases, the precise variable that would be expected to cause a change in tonnage may not be available; appropriate proxy variables that are available can be found in many cases. In addition, the precise form of the variable that is most appropriate may require some judgment. Whether the variable should be "real" or nominal; whether personal or disposal income is best; and other factors may be best judged based on trying alternatives and assessing "fit" in the model.

- Functional form: The models may be linear, in which each variable is assumed to affect the tonnage disposed or generated in a linear fashion. Whether transformations on the variables or the functional form provides a better approach is usually determined based on "fit" of the model and the statistical measures of model performance. In addition, however, log-log models are sometimes used because of the ease of interpretation of the results -- in these models, the coefficients can be directly read as elasticities.[5] Useful transformations include first differences, or logs of first differences which explain the amount of change in the tonnage as a function of the explanatory variables (or changes in the variables). First difference equations can help purge reduce the overwhelming impact that the fact that all these types of variables tend to be increasing over time can have on the model's fit and the patterns of the residuals. In addition, some models have been estimated with the dependent variable normalized as tons per person or tons per household. Selection among alternatives is usually based on fit, ease of interpretation, and performance based on whether the model delivers the expected "signs" of the coefficients of the variables (e.g., you'd expect more persons to provide more tonnage, not reduce the tonnage forecast, all other factors constant).

- Misspecification and multicollinearity: Two frequent sources of model misspecification are omitted variables and serial correlation of the residuals. The amount of bias that your model will have from misspecification depends on the degree of correlation between the omitted variable and the variables that have been included in the equation. Therefore, if the left out variable has patterns that closely resemble those of an included variable, the coefficient on the included variable will include both the impacts for the included variable, and some of that effect that would be more properly attributed to the excluded variable, biasing the effect that you are attributed to the included variable. With serial correlation, the errors from are not independent over time, as the underpinnings of the modeling approach requires. Rather, the errors from one time period are correlated with errors in the performance time period. The Durbin Watson statistic provides a test for this type of first order serial correlation, and using first-difference equations tend to reduce this problem. If the community is using monthly tonnage data, there may be a likelihood of twelfth order serial correlation -- for example, February tonnages may always tend to be low. No standard statistics provide easy indications of this type of problem; knowledge of the correlation would need to be incorporated into the forecasting process (for example, a series of monthly or seasonal dummy variables could help). Strong multicollinearity between variables within the equation (which can be identified by examining the correlation matrix) should generally be avoided. The problem may be diagnosed through instability in the coefficients depending on whether one or both variables are included in the model. For example, if population is included alone, it is significant and the right sign; if population and employment are included, the coefficient on population may lose significance and develop

[5]Elasticities provide a useful summary measure of the impact of changes in the explanatory variable on the dependent variable. For example, an elasticity of .5 would indicate that a doubling of population would lead to a 50% increase in the amount of tonnage disposed.

an unexpected sign. Although two variables may theoretically provide separate explanatory power, it may be important to select only one for inclusion in the model.[6]

- Lagged variables: A period of time frequently passes between the time an explanatory variable changes and the time the dependent variable completes its response. For example, it may take a number of months for a change in rates to have its complete effect on tonnage. Since monthly data are often used (in order to increase the sample size for the estimated data series), and many consumer responses take over a month to occur, it may be necessary to consider including lagged explanatory variables in the model. In fact, it may be appropriate to include variables lagged over a period of time, or approximate the effect with moving averages or weighted (geometric or polynomial distributed) lags. These complicate modeling and examination of the error terms, reduce the degrees of freedom, and complicate correlation between variables, but they can increase the explanatory power or fit the logic of the relationship better than other approaches.

- Seasonality or business cycles: Much economic behavior -- and therefore, tonnage disposal -- follows seasonal patterns. Failure to adequately capture seasonality in the forecasting model leads to a number of estimation problems in the residuals and the estimates of parameters. Seasonality can be treated through dummy variables. Business cycles can be appropriately addressed through the inclusion of employment (in some cases construction employment is particularly appropriate), income, and other economic activity indicators.

- Structural changes: Longer time series of data generally allow for better fitting models -- assuming that there have been no dramatic changes in the relationship between the dependent variable and the explanatory variables. However, if the period includes a time in which dramatic new programs were implemented (which decreased tonnage disposed), or a new price went into effect where none previously existed, changes in the structure of the relationship between tonnage disposed and the explanatory variables may be suspected to have occurred. If a period before and after a suspected structural shift are included, the coefficient for the variables may be an "average" of the before- period relationship and the after-period relationship, and the average may not be appropriate for use in forecasting for the future period. Tests for diagnosing this condition exist, but a logical understanding of the historical conditions in the community are essential to a good fitting model.

[6]In cases where we have found this problem, we have generally retained the variable that showed better performance and/or the one that had forecasts in the further available on a more timely basis.

IV. Case Studies

The following section provides brief summaries of the forecasting approaches being used in different solid waste agencies across the U.S., generally grouped according to the technique applied.

1. Time series approach:

The only example that we have found of the use of a time series approach in solid waste is estimation work SRC conducted for a solid waste agency on the west coast. In this work, we found that a Winters exponential smoothing model tracked the historical data well and produced short term forecasts that performed well. The model decomposed the forecast into a trend and seasonal components, and found that monthly data (rather than seasonal data) performed better. Retaining more data points allowed use of more data, and provided better responsiveness to changes in the tonnage deliveries. This relatively simple approach has not been widely used elsewhere, and it is appropriate only for short term forecasts.

2. Per-capita or per-employee approach

By far the most common technique used by solid waste agencies in projecting future tonnage is to use current or historical waste tonnage to derive a per-capita or per-employee ratio, which can be "scaled up" using forecasts of future population or employment figures. Examples of this approach are provided below.

- Kaui County assumed that growth had come from constructions from both population and businesses forces in the past, and assumed that both factors would continue to be drivers in the future. They used measures of commercial activity, population, a judgment or averaging factor related to tourism, and scaled up the waste stream composition data to derive the projection of disposal. They added disposal plus a recycling percentage that was assumed between 0-3% to derive solid waste generation estimates.

- New York City's study stratified waste composition samples by housing type (measure of density) and income. Then, changes in housing patterns in the future were provided for nine substrata (based on income and density characterizations) and the composition numbers were scaled up by substrata to develop the residential tonnage forecast. For the commercial/industrial sector, the waste stream was broken down into 24 non-residential sources based on SIC code, and the results of the composition studies were then scaled up based on employment projections for the sectors. There is relatively little recycling, so the equations are assumed to estimate generation and disposal, and their disposal numbers are strong because there is one main landfill.

- Los Angeles is taking a multi-faceted approach to developing a forecast. They have surveyed landfills, haulers, and developed a generator-based approach by SIC code. For the non-residential side, they conducted waste sorts for a number of SIC codes, derived estimates of tons per employee by sectors, weighted, and scaled up by forecasts of employment by sector. For the residential side, they are scaling up 1990 tonnage based on a survival model's forecast of population. Forecasts for both "with" and "without" AB939 changes are being estimated.

- Vancouver, BC derives a residential forecast by using information on population within geographic areas to generate forecasts of residential MSW. The non-residential sector forecasts are based on measured data on tons per employee for major SIC codes, and economic forecasts of employment by SIC code are used to scale up to the future.

- Dade County, Florida uses a static generation rate based on the last calculated year, and derive a generation per person per day. They assume a small percentage for source reduction, and use population projections for 20 year to generate forecasts. They are revising their approach to break out separate categories of non-residential customers, and other modifications.

- Waste Composition Study Based Approaches: In Ohio, most of the tonnage forecasting for county solid waste plans draws from waste composition studies, and for the most part, composition numbers are being scaled up by population for the residential sector, and by employment for the non-residential sector. In Ottawa, Canada, the 1991 waste characterization study is being scaled up by population forecasts. Similar approaches are being used in Washington State, Vanderberg County, Indiana, and other locations.

Each of these communities has nuances to the manner in which this forecasting approach is implemented, and each has found that the model performs relatively well. Some, however, are modifying their approach in the future as more data are collected, and more flexible approaches are needed.

3. Econometric Approaches

Several communities have used econometric techniques to derive waste stream forecasts.

- Seattle, Washington estimated separate equations for residential, commercial, and self-haul transfer station waste disposed[7]. The residential model (using a dependent variable

[7]These models are described in Skumatz, "Econometric Findings in Solid Waste: Demand, Customer Choice, and Reactions to System Change", Yale Working Paper, Draft, December, 1992.

of log of first differences in pounds per household) included significant explanatory power from price, income, household size, and recyclable price, but did not find a significant added influence from weather factors. The small commercial self-haul equation derived significant explanatory power from the following variables: Seattle's tipping fee at the transfer station, King County's (nearby competing facility) commercial fees at transfer stations, and employment in the construction sector. Seattle's model for the commercial sector at large derives the disposed tons per employee as a log-log model including price of disposal at the landfill. The equation estimating residential self-haul found explanatory power from per capita income and the price of disposal at the transfer station. These models performed well for rates calculations and program and facility planning activities.

- King County, Washington constructed a log-linear equation estimating aggregate tonnage generated as a function of population and real personal income. The estimated generation series is a constructed series using the disposed tonnage plus estimated recycling in the area. Price was excluded from the model because it was not clear which would be the appropriate price variable, because there was not enough variation in price over the period to justify inclusion, because the County believed that the level of prices was too low to generate reliable estimates of future customer reaction to price, and because there may have been a structural change in the relationship with price in the 1988-1989 time period. The model has performed well for a variety of applications.

- Minnesota is upgrading their forecasting work to develop an equation using population and manufacturing employment as the key drivers. They are likely excluding price from the equation for reasons similar to King County.

- Pooled Equations: Estimation work using monthly data from nine jurisdictions[8] estimated separate residential and commercial equations. The significant variables for the residential model (measured in pounds of refuse per capita per day) included price, household income, mean temperature, precipitation, household size, age distribution of the population, population density, and price for old newspapers. The commercial equation (measured in pounds of refuse discarded per employee per day) included price of disposal, temperature, precipitation, population density, and price for old corrugated containers. The magnitudes for the community specific intercept terms varied widely from community to community.

---

[8]Jenkins, Robin R. Municipal Demand for Solid Waste Disposal Services: The Impact of User Fees, Ph.D. Dissertation Draft, December 1989, University of Maryland, College Park, Maryland.

4. Summary and Conclusions

Generally, the scaling up forecasting approach is most common and is appealing because it is simple and has low data requirements, it meshes well with composition studies, and it is logical and can use a simple spreadsheet for estimation. The econometric approach, although less commonly used, has found communities satisfied with the performance. These models provide good explanatory power, support scenario analysis, provide confidence intervals for estimation, and account for business cycles and weather and other factors in an explicit way.

V. Model Fit and Performance

The econometric models have derived some fairly similar elasticities. An 'X' indicates that some form of the variable was significant in the equations, but the value of the elasticity was not reported.

Variable	King County, WA	Seattle Residential	Seattle Self-haul comm'l	Seattle Comm'l	Seattle Resid'l Self-haul	Pooled Resid'l	Pooled Comm'l	West Coast Agency
Price		-.14	-.55 own, +.15 competing	-.21	-.30	-.1257	-.29	-0.09
Income	+1.14	+.59			-.79	neg		
Employment			+.32					1.3
Population	+2.26							
Household Size		+.72				neg		
Population Density						pos	neg	
Age Distribution						pos		
Price of recyclables		-.03				pos (insig)	pos (insig)	
Housing Starts								0.01-0.1
Precipitation						pos	neg (insig)	0.36
Temperature						pos	pos	0.02

VI. Summary and Conclusions

There are three main approaches to forecasting solid waste tonnage: time series, scaling up, and econometric methods.

SRC staff have applied each of these techniques on different projects for communities. The decision on which type of model to apply has depended on a number of factors, including:

- data availability for both estimation, as well as "feeding" future projections
- whether the required forecast model should emphasize near- or longer term forecasting capabilities
- in-house staff skills to maintain, update the model
- types of decisions the forecast needs to support -- long-term policy or structural and scenario modeling, medium-term operations or facilities planning, near-term rate planning, etc.
- "fit" and performance of the estimated models.

In this work, we have learned a number of lessons from this forecasting experience:

- Data collection is an important foundation to good forecasts: Important variables include not only own price, but competitors price or forms of relative prices or price differentials. Data on recycling, program availability, and "leakage" are important for a strong equation that can perform over time and take account of programmatic changes. It is important for communities to require reporting of consistent, well-defined series, and for historical data to be reconstructed to the extent possible. Communities need to put some emphasis on data collection, including tonnage, price, program and diversion tons; and find good sources for forecasts of economics, demographics, and other key variables.
- Simple models can be useful tools: Underlying cycles, based on only tonnage data, can be very useful for projecting near term tonnage -- cycles may dominate the short term -- and projections can be obtained at very low cost. "Scaling up" methods can provide reliable, short- to medium-term and integrate well with composition studies.

- Econometric models provide excellent models for scenario analysis: These models have performed well in an array of applications, are flexible, and can provide strong explanatory power.

In our work and in our review of the literature and case studies, we have found that in several areas, there is general agreement on model structure, key variables, and even on "elasticities" or impacts of key variables. However, more work is needed to examine performance and comparability across jurisdictions and the robustness of approaches over time.

Unfortunately, key data series are lacking in many communities to allow similar estimation work. The results of this review may provide alternatives that could be used in these communities until needed data series are collected, as the two approaches that are not data intensive -- time series or "scaling up" approaches -- can provide some help for near term forecasts. Or possibly, for those areas with general agreement, communities could adopt (and adapt) models estimated for other areas until their own data can support specific estimations for the communities.

Finally, agencies may find that, as in the electric industry, multiple types of models may be appropriate, depending on the applications needed.