SHAH, Kanti L., 1935-
TECHNIQUES FOR ESTIMATING CONSTRUCTION COSTS OF WASTE TREATMENT PLANTS.

The University of Oklahoma, Ph.D., 1969
Engineering, sanitary and municipal

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

TECHNIQUES FOR ESTIMATING CONSTRUCTION COSTS OF WASTE TREATMENT PLANTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

KANTI L. SHAH

Norman, Oklahoma

1969

TECHNIQUES FOR ESTIMATING CONSTRUCTION COSTS

OF WASTE TREATMENT PLANTS

APPROVED BY

[illegible]

[illegible]

James M. Robertson

[illegible]

DISSERTATION COMMITTEE

ACKNOWLEDGEMENT

I wish to express my deep appreciation to the members of my graduate committee, and in particular to Professor George W. Reid, who served as chairman of the committee. Professor Reid's helpful suggestions, guidance, and interest were of great encouragement to me.

The financial aid from the United States Public Health Service through the Department of Civil Engineering, University of Oklahoma, throughout my graduate studies, is gratefully acknowledged.

I express my thanks and appreciation to my parents, Dr. and Mrs. S. M. Shah, for their monetary help from time to time, and encouragement in all my endeavors.

For their time and understanding I thank my wife and children.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

TECHNIQUES FOR ESTIMATING CONSTRUCTION COSTS

OF WASTE TREATMENT PLANTS

CHAPTER I

INTRODUCTION

General

The increasing size and density of population, and its activity has made water-quality management more and more important. This point is stressed by The Select Committee on National Water Resources (1):

> Because the Nation will presently reach the limits of its capturable supplies while water demands are increasing sharply with no let up in sight, it becomes abundantly clear that for the foreseeable future, water needs can only be met by using the available supply over and over again. This has already happened in some of the more heavily populated and industrialized areas of the country. Thus, during periods of low flow it has been estimated that the water of the Ohio is reused almost four times before it reaches the Mississippi. In the future, the waters of many of our major rivers may have to be reused many more times than this.
>
> To be reusable, water must be of the right quality. Used water is always polluted and pollution degrades the physical, chemical, biological, bacterial, and/or esthetic qualities of water; the degree depending upon the kinds and amounts of pollution in relation to the extent and kind of planned reuse. Pollution can be just as effective in reducing or eliminating a water resource as a drought or consumptive withdrawal. Because of the rapid increase in the kinds and amounts of pollution occurring in water, water quality management

may now be given the highest priority if we are to meet future water requirements.

For the efficient development and management of any water resources region, it is desirable that all the variables and their interactions be well defined and their effects be well understood. These variables are: number of people within a water resources region, type and number of industries, area and topographical nature of the region, hydrology of the region, and cost of treating water and wastes.

It is then necessary that means be available for estimating the future population and the probable number and type of industries, both current and future, and the variability and alterability of the topography and hydrology of the region under study.

Other two variables are the costs of treating water and wastes which are functions of scale, time and pollutants. There is also a relationship between the cost of treating the water and the wastes. Minimal cost of processing and cooling water is a factor in industrial plant location and the cost of production of certain types of goods depends on the cost of treating the water and the wastes. Therefore, a knowledge of the above two variables is very essential and a prerequisite in the development and operation of a water resources region.

There are many reasons why engineers and economists concerned with the development and management of water resource regions should be conscious of costs. First, the public and the elected officials often prefer low cost, short term benefits, but engineering ethics puts sound engineering over cost and generally prefers long term over short term benefits. Second, economics is a factor in sanitary and water resources engineering at the local level because of public apathy toward pollution. The public likes to enjoy the benefits of clean water, but it is often reluctant to pay for the renovation of waste waters. Naturally, the attitude of the public is reflected in elected officials who do not cherish the idea of spending any more public funds on a waste treatment plant than is barely necessary. However, public policies related to water quality objectives are becoming more and more restrictive and pollution abatement laws exist at state and federal levels. Thus adequate and proper treatment of municipal and industrial wastes is necessary from both the legal and moral aspects. It is essential therefore, that there is available in the planning stages of the development of water resources activities a reliable estimate of the cost of treatment of wastes to a desired degree within any basin.

Problem

The problem of this dissertation arises from the need of reliable unit cost estimates for construction of plants of major treatment categories. Present cost estimating data, though numerous, vary widely in format from author to author. Much of this data is reported without sufficient and representative cost information. Many authors did not take into account regional differences. Very few considered unit construction cost in terms of population equivalency [PE]. Most authors estimated unit construction costs in terms of population or flow. An intensive search of past and present literature failed to find a single article or paper which considered all the significant factors or variables with enough data to develop a reasonable mathematical model for estimating the cost of municipal sewage treatment plants. <u>The terms "enough" and "reasonable" as used herein are meant in a practical and statistical sense.</u>

Reliable input data for use in the models is needed to make them workable or to get a better output. The authenticity of any cost estimation data is dependent upon the number of projects involved in establishing criteria, the inclusiveness of the types of projects, and local conditions under which the projects have been constructed. A few mathematical models for maximizing net benefit in the development of a water resources region have been proposed. A reliable estimate of the cost of waste treatment is one of the variables which is needed to make a mathematical model operable.

Objective

The purpose of this study is to provide estimates of the cost of treating municipal and industrial wastes. The study will use statistical tools to establish a model for estimating the cost of municipal waste treatment plants. In the case of industrial wastes many statistical procedures could not be followed because of the lack of data on various types of industrial wastes.

By application of the data obtained as a result of the mail survey a mathematical model will be developed to provide means for estimating the construction cost of waste treatment as a function of several variables such as, population equivalency, degree of treatment, types of processes, and flow. The model is to be simple and future independent variables reasonably easy to observe so that it can be readily applied to any region, for any major process, and at any scale, considered in the study. The general form of the equation desired is:

$$Y = \beta_o + \sum_{i=1}^{n} \beta_i X_i \quad .... \qquad (1)$$

Although a fairly large volume of data has been obtained, additional and more representative data as available should be used to refine the model. The variables in the model, however, will remain unchanged.

This study will give separate mathematical equations for estimating the unit construction cost for different categories of current treatment works commonly constructed for treating municipal wastes.

Need for Study

According to the Bureau of Census estimates (2,3), the population of United States may reach 274 million by the year 1980 and exceed 420 million by the year 2000. The increase in population will increase water consumption. The U.S. Public Health Service (USPHS) has estimated that by 1980 total withdrawals of water for all purposes might equal 650 billion gallons per day (4). Municipal demands alone in the year 2000 could reach five times the present domestic requirements or about 90 billion gallons per day.

The sanitary and water resources engineers are thus not only concerned with the problem of meeting the ever increasing demand for suitable water, but also with the fact that water as it is used becomes polluted and unfit for further use unless adequately treated.

It is thus clear, that one of the principal requirements for water in the future is for the dilution of effluent from the treatment plant and disposal of wastes into streams. Dilution requirements, in relation to the degree of treatment, for maintenance of four parts per million of dissolved oxygen in the stream have been estimated by Reid (5). This model trades treatment costs off against impoundment costs for the projected waste discharges (6). Impoundment costs are available. The optimal solution, of course, can be found by the relationship:

$$\text{Minimum cost per Unit Vol. (\$/MGD) of Wastes} = (\$)_T \times \text{Vol. Treated} + (\$)_S \times \text{Vol. for Dilution}$$

where $\$_T$ and $\$_S$ are costs of waste water treatment and relatively pure water storage for dilution, respectively.

Reid (5) has predicted that in the period 1980 to 2000 approximately 64 per cent of the required stream flow for all purposes will be needed for dilution of wastes. Table 1 shows the distribution of predicted required stream flow.

Increasing requirements for relatively pure water for waste dilution make it necessary to select the most economical and efficient combination of water storage facilities and new, highly efficient waste treatment facilities. The solution lies in engineered systems of water quality management. There are two main approaches to such systems:

1. Water reclamation, i.e. removal of water from waste-water return flows.
2. Water renovation, i.e. removal of the quality factors or upgrading of the water quality.

Needless to say, water renovation is the most common approach today and will continue in the future. This approach is generally known as water treatment; municipal waste treatment, or purification; and industrial waste treatment. It follows then that new and efficient treatment processes, requiring less dilution water and treating the wastes to a degree so that reuse of the water is possible,

TABLE 1

DISTRIBUTION OF REQUIRED STREAM FLOW BY USES, UNITED STATES, 1980 AND 2000

Use	Per Cent of Total Flow 1980	Per Cent of Total Flow 2000
Agriculture	20.0	18.1
Mining	0.1	0.1
Manufacturing	1.7	3.0
Thermal Power	0.3	0.4
Municipal	0.7	0.8
Land Treatment	0.8	1.0
Fish and Wildlife Habitat	12.8	12.8
Sub-Total	36.4	36.2
Waste Dilution Flow	63.6	63.8
TOTAL	100.0	100.0

Source: Water Requirements for Pollution Abatement, Committee Print No. 29, Water Resources Activities in United States, U.S. Senate Select Committee on National Water Resources, July 1960.

must be developed. A study of current waste treatment costs can provide guidelines for extrapolating costs of treating wastes by new and improved waste treatment systems.

Such a study must be kept up to date by research and new data. This study is concerned with the development of mathematical and statistical techniques for estimating the current cost of construction of the waste treatment plants and also provide estimates of expected costs in future. Such estimates would take into consideration the developments in the waste treatment and construction technology.

Justification

Grants-in-aid for municipal waste treatment works construction were first authorized under the Federal Water Pollution Control Act of 1956 (7), and widely liberalized by the congress in the decade which followed. Commensurately, the dollar volume of the construction contracts awarded for these facilities has steadily increased. The annual investment in municipal waste treatment construction has more than tripled in the past 7 years (8). This, plus the many changes which have evolved in treatment methods such as modifications in the activated sludge treatment process and increased use of stabilization ponds have increased the need for shortcut reliable estimating methods.

It is necessary for the public officials charged with the responsibility of providing sewage treatment facilities; for the consulting engineers whose task is to plan and design the treatment works; for the investors who need to make bond purchase decisions; and last but not the least for the regulatory agencies who are vested with the authority of approving or disapproving the methods and physical processes, to have reliable information on how much the treatment facility would cost.

It is not possible to have actual project costs in the initial stages of planning. Actual costs are available only after plans and specifications are complete, bids for construction work, and materials and equipment are received, and contracts awarded. But before this long and often cumbersome procedure, there is a real need for some method which will provide preliminary cost estimates.

Emphasis must be placed on the actual cost information; there is no substitute for it. However, cost estimates play a very important role in the initial stages of sewage treatment works planning. Of course the decisions should be based on water quality requirements and not on availability of funds (8). The size of the project is established by organic and hydraulic loadings on the proposed plant. Yet, information supplied by preliminary cost estimates of the treatment plants will be valuable (8) in the following ways:

a. One can get some indication whether the project should be constructed in stages rather than full-scale works on a one time basis. Plants are designed for 10-20 years in future depending on various considerations including financing at the project time. It is possible to estimate the costs of a waste treatment plant at various stages of construction and come up with minimum total cost of balancing between economies of scale and the cost of amortization of the original capital outlay.

b. It will help municipal officials develop master plans for sewage works facilities and estimate future financial needs.

c. It may indicate the future period for which actual construction will be scheduled on a long range plan.

d. Such estimates can guide bond issue referenda.

The above examples of the usefulness of construction cost estimates demonstrate the definite need for reliable estimating tools. Cost-estimating guidelines based on valid cost statistics and their intelligent interpretation when used with caution are very useful tools for consulting engineers, municipal officials, developers, and the agencies charged with the development and management of the water resources region.

Once more, it must be emphasized that estimates are just estimates and are in no way substitutes for actual cost figures. They may be used by experienced persons with a thorough knowledge of the specific projects.

CHAPTER II

PREVIOUS WORK

One of the first attempts to establish a relationship between the costs and various methods of municipal waste treatment and the different processes involved within a particular method, was made by Schroepfer (9). Based on observations and statistics gathered on a nationwide basis, he analyzed both construction and operation costs of treatment in terms of waste reduction. The parameters for the waste water quality were suspended solids and biochemical oxygen demand (BOD). He also gave construction costs for various plant capacities in terms of flow measured as million gallons daily (MGD).

The cost data of this study are of little value today since they were derived from a rather limited number of plants. Furthermore, this study, carried out before the development of high rate trickling filters and nationwide standards of operating efficiency, showed widespread variations in cost. Its main significance lies in the general ideas and concepts presented. For example, Schroepfer derived functions describing the costs of waste treatment for different sizes of treatment plants and for various degrees of treatment. Ever since Schroepfer's work, various authors have attempted to find a relationship between the costs of construction and operation, with various design parameters.

Velz (10), collected information on construction costs of treatment plants from literature and questionnaires made a statistical analysis of the data. He related the unit construction cost of waste treatment works per MGD to the size of plant in MGD. The cost data he used were bid prices or final payments to the contractor. Preliminary expenses, land, engineering, legal, and interest charges were excluded. The author estimated that these bid prices or the final payments represented about 80 to 85 per cent of the total costs of the treatment plants. All plant costs were referred to 1926 as the base year of construction, adjusted by means of United States Average Engineering News Record Construction Cost Index (ENR-C Index). Velz used the data on 185 different plants. The costs were referred to 100 per cent efficiency in the removal of the BOD by considering primary, chemical coagulation, trickling filter, and activated sludge systems as effecting 35, 65, 85 and 90 per cent BOD removal, respectively. A plot of the data on double-logarithmic paper suggested a linear relationship between the plant size and the construction costs per MGD; the equation of the line was then determined by using least squares method.

In the above study by Velz, no regional price differentials were considered. The estimation equation used was of the form:

$$y = ax^b \tag{2}$$

where y is the unit cost per MGD, and x is size of plant in MGD. The study was based on a sample of waste treatment plants in northeastern and central United States, and as such cannot be considered valid for the rest of the regions.

In 1954 Morgan and Baumann (11) studied the costs of trickling filter plants in Iowa. Since there were some industries the authors based costs on population equivalents. Whether these population equivalents were based on a hydraulic flow or BOD loading basis is not mentioned.

Diachishin in 1957 (12), attempted to refine and up date the work of Velz. He analyzed the data of 154 plants and derived two estimating curves, one for primary treatment and one for secondary treatment. All plant costs were referred to 1913 as the base year of construction, adjusted by means of the ENR-C Index which has a value of 100 for the year 1913. Diachishin's data formed a curve parallel to that of Velz but with roughly 30 per cent lower costs. He concluded, that increased efficiency of design, better equipment and new and better methods developed as a result of research, have decreased the costs of sewage treatment.

In Diachishin's study, the cost curves were based on information for the treatment plants constructed since 1947. However, no attempt was made to account for regional differences in the estimation equation. Like Velz, he also related the unit construction cost of waste treatment works per MGD to size of plant in MGD. He also did not consider other variables such as, PE and efficiency, which may influence the cost of construction.

The same conclusion was reached by Mau (13), who analyzed treatment plant costs in the State of Kansas. He indicated that the costs of treatment plants built in Kansas between 1950 and 1956 were approximately 70 per cent of the costs of the plants which Velz analyzed in 1948. The author found that the design criteria for treatment plants built before 1939 were much different from the current design policies.

Recognizing the need for some sort of preliminary cost estimating aids, the USPHS, began in 1958 a study of the construction costs of sewage treatment facilities. Such cost estimating aids, which may be called cost "indicators," were designed to be of value to municipal officials, consulting engineers, and agencies involved in sewage treatment works. Since then many studies have been made by the USPHS mainly to refine and update its previous studies.

Howells and Dubois (14), made the first of these investigations shortly after a Federal grants program was authorized by PL-660. This study, based on the analysis of twenty small secondary sewage treatment plants in the upper midwest, concerned itself with engineering design practices, cost of construction, and estimated operation and maintenance costs for projects assisted under the program.

The above construction cost study drew its data from actual contract prices. It did not include the cost of land, or engineering, administrative and legal services. The design population range was from 600-12,500 and design data were given for primary sedimentation, activated sludge, trickling filters, secondary sedimentation, sludge digestion and sludge drying beds. The cost data from these twenty projects were evaluated against the parameters of design population, population equivalent, and design flow to establish unit cost curves for each parameter.

Access to data on the increasing number of treatment plants receiving Federal construction grants led to another cost study conducted by the USPHS. These data indicated probable construction costs and did not include land, engineering, administration and legal costs. Interceptor and outfall sewers, pumping stations and similar works were also excluded.

The study made by Thoman and Jenkins (15), in 1958, presented construction cost estimating equations and curves, one for primary treatment plants, one for secondary treatment plants and one for stabilization ponds. Construction costs were adjusted to the ENR-C Index base year 1913. The equations were computed for estimating cost per capita as a function of design population. In an effort to account for regional differences in construction costs the United States was partitioned into twenty regions on a county line basis. Each region corresponded to one of the twenty cities used in obtaining the United States average ENR-C Index. A treatment plant selected within any region was assigned the ENR-C Index for the ENR City. The form of the equation was,

$$y = ax^b \tag{3}$$

where y is cost per capita, x is design population.

In 1960, the USPHS reported a study which expanded and refined the preceding investigations (16). This study updated specifically the previous USPHS cost studies. All cost data were converted into 1913 dollars using the ENR-C Index. It evaluated costs for six specific types of treatment: Imhoff tank, conventional primary treatment with separate sludge digestion, activated sludge, trickling filter with separate sludge digestion, trickling filters with Imhoff type treatment, and stabilization ponds. The data were analyzed by the statistical method of least squares. Estimation equations were derived relating costs per capita to design population.

The above study by Rowan, <u>et al</u>. was an improvement over the previous studies, although much was left to be desired. The authors used ENR-C Index to convert all cost data, nationally, to a "common denominator" of cost. In Chapter V, of this study, in the discussion of the cost index it has been pointed out that the Federal Water Pollution Control Administration Sewage Treatment Plant Construction Cost Index (WPC-STP Index) has certain advantages over the ENR-C Index. Also, Rowan, et al. did not consider either the flow or population equivalency.

Logan, et al. (17) investigated the application of systems-analysis techniques to the preliminary design of waste water treatment plants. Since the unit-process data was not available, construction and operation costs were obtained by visits to the plants. Equations for estimating cost per MGD as a function of design capacity in MGD were derived for unit processes in primary, high-rate trickling filter, standard-rate trickling filter, and activated sludge treatment plants.

Because of the inconsistencies in the field data, and the resulting difficulties involved in establishing parameters from actual costs of the existing plants the authors attempted to make cost estimates of a series of theoretical designs predicted on idealized conditions.

These estimates were based on St. Louis, Mo., conditions. ENR-C Index prevailing in 1960 was used. While this study was well planned the idealized conditions on which the authors based their theoretical designs seldom exist. The attempt to analyze the costs for each of the unit processes, however seemed to be a better approach.

Wollman (18) was the first to use a multiple regression model for the estimation of operation and maintenance costs. The estimation equation used was of the form

$$y = b_o + b_1x_1 + b_2x_2 + b_3x_3 \tag{4}$$

where

y = the annual operation and maintenance cost per daily population equivalency (PE),

x_1 = treatment level in per cent of BOD removed,

x_2 = per cent of total waste that is industrial, and

x_3 = population served by sewage system.

Park (19) approached the problem of estimating sewage plant construction costs by taking both hydraulic and biological loadings. He based the primary treatment plant construction costs on hydraulic loading capacity, dollars per gallon per day and secondary plants on organic loading capacity, dollars per lb per day. The author took into consideration both the hydraulic and organic loadings for estimating the cost of, what he termed a complete treatment. He deducted the unit costs of primary treatment plants from the unit costs of the corresponding complete treatment plants to get the incremental costs. Incremental costs were attributed to the organic characteristics of wastes.

Park converted the unit cost per capita to a unit cost per lb of BOD by assuming 0.2 lb of 5 day BOD per person per day. Similarly, the unit construction costs of primary treatment plants were converted to a unit cost per MGD, using 100 gallons per capita per day as the conversion factor. Such flat assumptions with regard to flow or the organic waste are questionable, because the nature of waste varies with the nature of the area, the number and type of industries contributing wastes and the amount of infiltration into the sewers.

In 1963 Assenzo (20), using multiple linear regression techniques analyzed the cost data on the basis of nine regions. It may be recalled that the USPHS divided the country into nine regions. He studied the construction and operation costs under the categories of primary and the secondary plants.

Assenzo's technique for estimating the cost of construction of the waste treatment plants was a definite improvement over previous methods. However, like other workers, he based his work on the ENR-C Index instead of the WPC-STP Index. Also, he considered one single equation for all types of treatment plants. The estimate of variance which he used to calculate the sample size was based on the studies made by the USPHS. The data from which the variance was estimated was not representative enough. This has been discussed in Chapter IV.

Smith in his report (21) made an effort to bring together in one computation scheme the significant cost and performance relationships for a limited group of processes and then attempted to calculate the performance and cost of the system as a whole, based on relationships which had been developed for the processes individually. The processes were: primary settler, aerator, final settler, thickener, digester, elutriation, vacuum filter, and incinerator. In Smith's study the program has been slanted towards preliminary design rather than towards simulation.

Very limited data is available on construction costs under different categories of industrial waste treatment plants. An effort was made by Eckenfelder (22) to put together the costs of constructing and operating industrial waste treatment plants. Graphs for estimating the cost of construction were presented on the basis of very limited data. No equations were given.

It is believed that this study will lead to the establishment of a mathematical model which will give better predictions than some of the previous models. This improvement will be achieved in various ways. For example, separate equations will be provided for estimating the construction costs of activated sludge plants, and high-rate and standard-rate trickling filter plants rather than including all of them in a broad category of secondary treatment plants. The use of the WPC-STP Index rather than the

ENR-C Index to bring all costs to a common base will provide a better evaluation of waste treatment plant costs. The WPC-STP Index is believed to be representative of cost changes peculiar to municipal sewage treatment plant construction and is more responsive in elimination of variations in treatment plants costs due to geographic location and time differences. Consideration of more than one independent variable might help in determining the significant variables as far as cost of construction is concerned. Also an attempt will be made to collect representative data from all over the United States.

Therefore, it is hoped that the study will update, refine, and improve waste treatment construction cost studies, and will also provide good predictive model.

CHAPTER III

MATHEMATICAL MODEL BUILDING

In many engineering and scientific studies the clear concept of cause and effect is obscure. This situation, tragic as it may seem, is due to perhaps unidentifiable variables because of lack of knowledge or some variables may be unquantified. If the causes of an effect are not known then it is not possible to predict the effect exactly, but there may be independent variables which can be observed and which can be used for predicting the effect, the dependent variable or response.

One of the important decisions in planning and development of a mathematical model is selecting the variables to be tested. It is desirable to represent the relationships adequately with a minimum and most significant number of explanatory variables. Only those independent variables which are thought to add materially to the significance of the regression should be included. Also, independent variables that are readily measurable should be selected. Experience and previous work in the area of study play an important role in selecting the appropriate variables.

In general, the explanatory variables in the model should not be correlated with some other unincluded variable or variables. Depending on the degree of correlation, nature of the study, and type of the model, the correlationship among the variables may mask the true relationship. Only those independent variables which account for the correlation with the dependent variable in the model must be selected. If the explanatory variables included in the model are interdependent a problem called multicollinearity arises. Johnston (23) has indicated that this problem of intercorrelation is generally not serious in predictive type models, the type of model involved in this study, if the interdependency may be expected to continue. However, coefficients of determination (R^2) may be misleading due to the interdependency of the explanatory variables.

The forementioned criterion for selecting variables were kept in mind in the study. Also, experience and previous work in the field were taken into consideration. Statistical procedures in the establishment of the model are discussed in Chapter V. Details of the data collection, procedures, and distribution of the data are given in Chapter IV. Stepwise multiple regression technique (24) was employed for establishing a model and is discussed in Chapter V. Only those variables which seemed to be meaningful and significant were included in the model.

Variables

From the theoretical considerations it is logical to relate unit costs of primary treatment plants to hydraulic loadings, and of secondary treatment plants to hydraulic and organic loadings. In designing the biological treatment unit process of the secondary treatment plant consideration to the organic loading is given.

A good measure of organic loading is the PE of the waste, reflecting the contribution to the organic loading from all sources within the community, for example, domestic and industrial. The population equivalency of municipal wastes can be computed,

$$P.E. = \frac{8.33\ QL}{b} \tag{5}$$

8.33 is a conversion constant, Q is average waste inflow to the treatment plant in MGD, L is average 5-day BOD of the waste in milligrams per liter (MG/L), and b is generally assumed to be, 0.17 to 0.2 lb of BOD per capita per day. In this study the figure of 0.17 lb per capita per day was used. Very few workers have used PE as a variable in the development of unit cost estimation equations. The word sufficient as used here, means volume of the data which would have given statistical precision. In addition to PE, other variables which may be important are: flow, BOD of the influent and efficiency.

Two methods can be used for establishing the equations for estimating the unit costs of construction of different types of treatment plants. One of the methods used by Rowan, et al. (16), is computing different equations, one for each type of treatment plant. Another possible method is to quantify the types of plants on some rational basis. It would then be possible to have more observations to use in the development of the regression equation for estimating the unit cost of construction of the plants. Both methods were tried in this study. It was thought that use of both possible methods of computing the equations would at least provide a good comparison.

No satisfactory method concerning the quantification of the treatment plants could be found. It was decided by the author to use dummy variables to represent the different types of secondary treatment plants. Their use in the study is expounded in Chapter V. Discussion concerning the use of the dummy variables can be found in Johnston (25) and Draper and Smith (26).

In previous studies (9,10,12,15,16,17), the dimension of the dependent variable was either dollars per capita or dollars per MGD. To compute the total cost one computed the estimated value of the dependent variable from either the design population or design flow and multiplied the value by either design population or flow, depending on the equation one chose. In this study, equations will be provided which predict unit costs of construction in terms of dollars per PE for which the plant was designed and dollars per MGD (design flow). In the case of the primary treatment plants a single equation will be presented estimating unit cost in terms of waste volume. The reason for providing the unit cost estimating equation in terms of waste volume is, that generally, primary plants are designed on the basis of volume of waste flow. Costs of primary treatment plants can therefore justifiably be estimated according to its hydraulic loading.

For industrial waste treatment plants it is desirable to consider chemical oxygen demand (COD) as one of the variables. Unfortunately not enough data was available to compute a regression equation for estimating the unit costs with reasonable precision. Dummy variables were used to quantify the different types of industrial wastes treated.

The expected strength of the influent and the effulent will give a measure of the overall efficiency required of a treatment plant. The type of treatment plant is accordingly decided upon, and its cost of construction estimated.

CHAPTER IV

SAMPLING PROCEDURE

In order to estimate the parameters of a predictive model for estimating the construction costs of the waste treatment works, a study of already constructed plants was conducted. Generally, when data on a large number of sampling units is needed, such as in this study, a sample survey will provide desired precision. In planning such a sample survey the following information must be known at some stage (27).

(a) Objective of the survey,

(b) The population to be sampled,

(c) The information desired concerning this population,

(d) The degree of precision desired in the results.

There are two kinds of sampling methods--probability and nonprobability sampling. In probability sampling, each unit is drawn with known probability. Methods of selecting the sample based on the theory of probability provide a measure of precision (28). The precision of the estimate based on a certain sample size can be computed if the sample is selected, its size calculated, and the estimate obtained by methods based on the theory of probability.

As indicated earlier, this study is concerned with the establishment of a mathematical model for estimating waste treatment plant construction costs in any part of the United States. Therefore, the sample should be representative of all types and sizes of plants from all over the country. Since the scope of inference was to be all plants in the United States, a mail survey was conducted. As discussed below, this was possible for municipal waste treatment facilities but not for industrial waste treatment facilities.

In probability sampling one must have a list of treatment plants in order to assign equal probability of selection to each sampling unit. In this study a list of the types and locations of all completed treatment facilities was needed. The Federal Water Pollution Control Agency (FWPCA), Department of the Interior (29) kindly provided lists of all municipal waste treatment facilities constructed during 1957-1967. However, no list was available for industrial waste treatment facilities and therefore use of the theory of probability in this case was not possible. Thus no precision could be attached to the predictive equations for the unit costs of construction of the industrial waste treatment plants.

For municipal waste treatment plants, separate lists for each of the five categories of plants were available (29). These categories of the plants were: activated sludge, high-rate trickling filter, standard rate-trickling filter, stabilization pond and primary treatment.

To account for the economies of scale in the construction costs, if any, each list was stratified according to the size. The stratification was based on designed population equivalency and was arbitrary. Populations divided into non-overlapping subpopulations are called strata. The sample is chosen by selecting simple random sample of elements from each stratum. The process of breaking down the population into strata, selecting simple random samples from each stratum, and then combining these samples into a single sample to estimate the population parameters is called stratified random sampling. In a simple random sampling procedure a sample is so drawn that every element has the same chance of being drawn. To be sure that each element has an equal probability of being selected, a table of random numbers was used.

It was decided at the outset to exclude the states of Hawaii and Alaska, and the Territories from the sampling plan, but to include the District of Columbia. No list of treatment plants was available for these states. It was also decided to eliminate septic tank treatment facilities.

The population to be sampled should coincide with the population about which the information is wanted. The population defined, the information desired concerning this population was determined. The variables that are to be studied have been indicated in Chapter III.

The required precision, in statistical terms, is difficult to specify and also difficult to obtain. To compute a sample size, n, for a specified precision, one must have an estimate of the population variance. The precision is generally specified by defining the expected width of a specified confidence interval. For example, it may be specified that the expected width of a 95 per cent confidence interval for the population mean be of a certain size. The confidence interval which can be calculated for any statistic is of considerable importance as it expresses the reliability of the estimate of a parameter. The narrower the interval, the more precise is the estimate.

In order to estimate the sample size within each stratum of the stratified random sampling scheme it is necessary to know or have an estimate of the variance within each stratum. A random sample of reasonable size (greater than 30 wherever possible) was taken and an estimate of the variance was obtained for each stratum. With the help of this estimated variance, the required sample size within each stratum was calculated by Neyman's allocation (30).

Neyman's allocation allows the assumption of equal cost per sampling unit for all strata. This method of allocation was therefore considered appropriate for this study. Hence, the problem of allocating the sample size among the strata can be put in mathematical terminology. The sample size n_s was determined using the relationship (30):

$$n_s = \frac{N_s\ S_s}{\Sigma N_s S_s} \cdot n \ \ldots\ldots\ldots \qquad (6)$$

where n_s is the sample size required of s^{th} stratum, N_s is the size of the s^{th} stratum, S_s is the sample estimate of the standard deviation, and n is the number of observations required. The required sample size can be computed by the formula.

$$n = \frac{(\Sigma N_s S_s)^2}{\Sigma N_s S_s^2 + N^2 V^2} \ \ldots\ldots \qquad (7)$$

where N is the total population size, and V^2 is the desired variance, which is:

$$\frac{d^2}{t^2},$$

where d is the half width of the required confidence interval, and t is the level of reliability, d the half width of the required conficence is approximated by the relationship (31);

$$E = \frac{tc}{\sqrt{n_o}} \qquad (8)$$

where E is the maximum error of the average of the sample in per cent, t is "students" t for (n-1) degrees of freedom at a specific probability level, n_o is number of observations, and c is called the coefficient of variation which is the standard deviation expressed as a percentage of the average $\overline{X}$, and is equal $\frac{100\ S_o}{\overline{X}}$ where S_o is estimated total variance, within and between the strata.

Considering the procedures used in the previous study (20) and the proposed statistical analyses to be used in this study, a 95 per cent confidence interval width of 10 to 20 per cent of the estimated mean cost in dollars per PE was chosen. It should be pointed out that these percentages of the estimated mean cost in dollars per PE represent the interval sizes for the mean of the dependent variable when the independent variables take on their mean values. Using the required precision and the estimates of the variances, the number of observations required for each type of treatment plant was computed using the relationship in Equation 7. The total required sample size thus computed was 616.

Having computed the required sample size (Table 2) for each type of treatment plant, Neyman's allocation using Equation 6 was applied to compute the required sample size for each stratum. For example, the total number of observations required for activated sludge plants was 135. Neyman's allocation indicated that a sample size of 92 of the

TABLE 2

DISTRIBUTION OF SAMPLE SIZES BY TREATMENT TYPES

Sample Size Needed	Sample Size Requested	Sample Size Received	Difference Between Size Needed and Received
		Primary Municipal	
93	103	102	+ 9
		Stabilization Ponds	
194	232	157	-37
		Standard-Rate Trickling Filter	
83	98	67	-16
		High-Rate Trickling Filter	
111	155	122	+11
		Activated Sludge	
135	172	115	-20

activated sludge treatment plants designed for population equivalency of 2,500 or under was needed. Such computations were made for the five different types of treatment plants considered in this study.

In order to obtain a representative sample of different types of municipal waste treatment plants (types considered in this study) in each state within each stratum proportionate stratified sampling was used. The lists of the treatment plants (29) furnished by the FWPCA were based on states. For example, the list of activated sludge treatment plants indicated 14 such plants within the range of 1-2,500 design PE were located in the state of Illinois and therefore on a proportionate basis the required sample size of 6 was computed. In other words, to meet the statistical requirements for a given precision, and to be representative, information concerning 6 activated sludge treatment plants for the stratum (1-2,500) was needed from the state of Illinois.

As indicated earlier, a mail survey was deemed feasible and was carefully designed. The questions were formulated such that the answers to the questions either gave values of the variables for which the information was desired or gave the data from which variables could be computed. The questions should be clear, concise and easy to answer as this brings a better response. This was kept in mind when designing the questionnaire for this study. For example, the population equivalency was not asked, instead,

TABLE 3

DISTRIBUTION OF SAMPLES RECEIVED BY THE DESIGN CAPACITY IN TERMS OF POPULATION EQUIVALENCY (PE)

1-2,500	2,501-10,000	10,001-25,000	25,001-50,000	50,001-100,000	Above 100,000
		Primary Municipal*			
17	29	21	14	12	9
		Standard-Rate Trickling Filters			
24	26	10	2	3	2
		High-Rate Trickling Filters			
18	56	26	11	6	5
		Activated Sludge Treatment Plants			
58	21	15	8	6	7

1-1,000	1,001-2,500	2,501-10,000	10,001-25,000	25,001-50,000	50,001-100,000	Above 100,000
			Stabilization Ponds			
91	32	20	8	2	3	1

*Distribution based on population.

flow and BOD were asked. This enabled the author to calculate PE. Appendix A is a copy of the questionnaire used in this study. This questionnaire along with the instructions for filling it and a cover letter from Professor George W. Reid, Director, Bureau of Water Resources Research, University of Oklahoma, and a stamped return envelope was mailed to the Superintendent of the Sewage Works of the community whose treatment plant was selected by the random sampling scheme. In addition, a questionnaire requesting the same information was sent to each State Health Director, with a list of communities whose treatment plants were selected within their respective jurisdictions. This was done to ensure a better response and also to check the reliability of the data. The purpose was achieved to some extent.

It was indicated earlier that the required sample size computed was 616. However, 760 questionnaires were mailed. These additional questionnaires were sent for two reasons: mail surveys usually result in a low response rate and therefore it was hoped that more questionnaires than required would bring a response closer to the sample size required, and secondly, rounding of fractions was made on the positive side. For example, suppose on a proportional basis if the computed sample size for a stratum in a certain state was 1.4, it was rounded to 2.0.

Eight hundred and nine letters with questionnaire forms were mailed on March 1, 1969, including 49 letters to

State Health Directors. The addresses were requested to return the completed questionnaires by March 22. On March 26 the questionnaires received numbered 370 including the information received from the various State Health Departments. Since the sample size received was not enough, it was decided to send reminders to State Health Departments. The reminders were not sent to Superintendents of the individual plants because of the large number of letters involved. On March 30, 1969, a personal letter from Professor Reid was sent to the State Health Directors who had not responded to the original questionnaire. The reminders requested the Directors to return the completed questionnaire by April 7. At that time it was decided that the processing of the data would be started on April 11, and any questionnaires received after that date would be discarded. The total sample size obtained by April 11 was 563, 53 less than required. Distribution of the stratum sample size received is given in Table 2 for each type of treatment plant.

CHAPTER V

METHODOLOGY

This chapter discusses the selection, employment and justification of certain procedures and techniques.

Procedures for Accounting Regional and Time Differentials

In order to account for the regional differences in the cost of waste treatment plants which were noted in several studies (12,15,16,17,18), the United States was partitioned into several regions. The question arose as to how many regions and what basis should be used for dividing the United States for this study. It was first thought that the United States should be divided into 22 regions based on the major water resources regions (5,6). However, the lists of the municipal waste treatment plants supplied by the FWPCA were not based on the water resources regions. It was concluded that it would be very difficult to construct 22 lists based on water resources regions. The assignment of several plants to any one region might be questionable. It was finally decided that it would be more meaningful if political boundaries were used for dividing the country. The state boundaries were used for obtaining the samples as discussed in Chapter IV. To account for possible regional differences the United States was partitioned into 20 regions on a county line basis. Each region corresponded to one of the 20 cities used in obtaining the United States Average Engineering News Record (E.N.R.) Construction Cost Index (7). Figure 1 shows the twenty regions.

Cost indices thus developed for the areas of influence of 20 major trade centers are helpful in comparing the costs on a regional or time basis. These 20 regional indices can be used to evaluate the costs in any of the 22 river basins by considering the index or indices which influence the area within a particular basin.

Since waste treatment works were generally considered to follow the ENR-C Index, a comparison of this index to the WPC-STP Index was undertaken by the FWPCA (7). In order to compare the indexes both were converted to a base of 1930 = 100. The base of 1930 = 100 means that the average plant cost for each of the 20 cities was determined for the year 1930 and the sum of these 20 average values was the base for the index and equals 100. In Figure 2, the converted WPC-STP Index is shown graphically. Appendix B presents the converted WPC-STP Index.

It is clear from the graph that since 1940 there has been a widening divergence between the WPC-STP Index and the ENR-C Index. In order to provide a comparable base to other

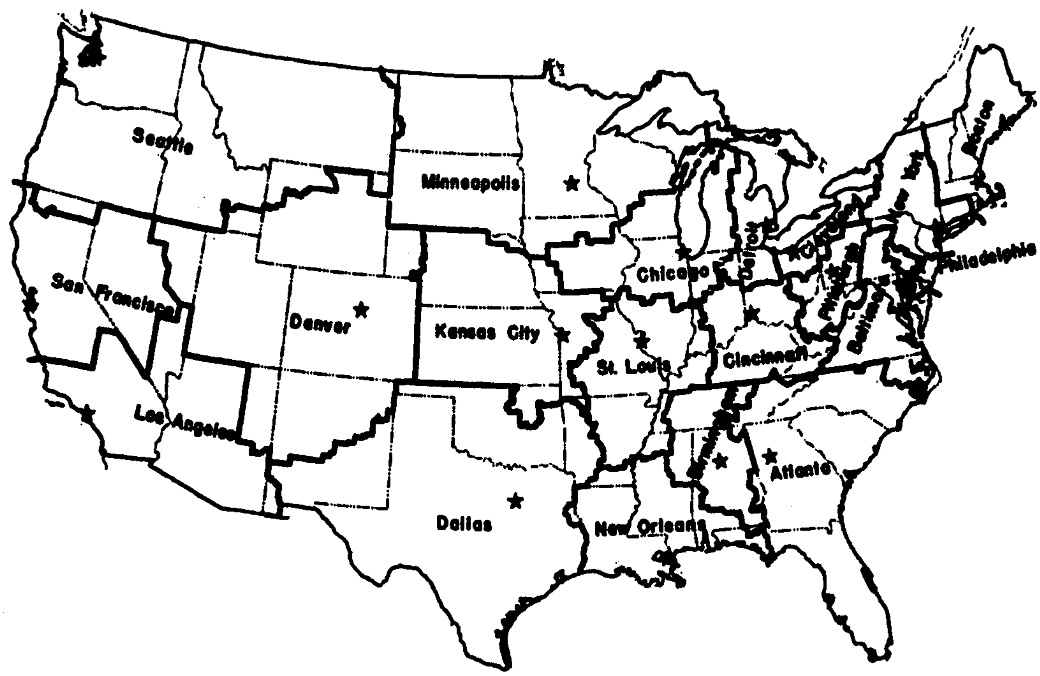

Figure 1. Map of 20 Index Cities and Their Assigned Areas of Influence. After Federal Water Pollution Control Administration.

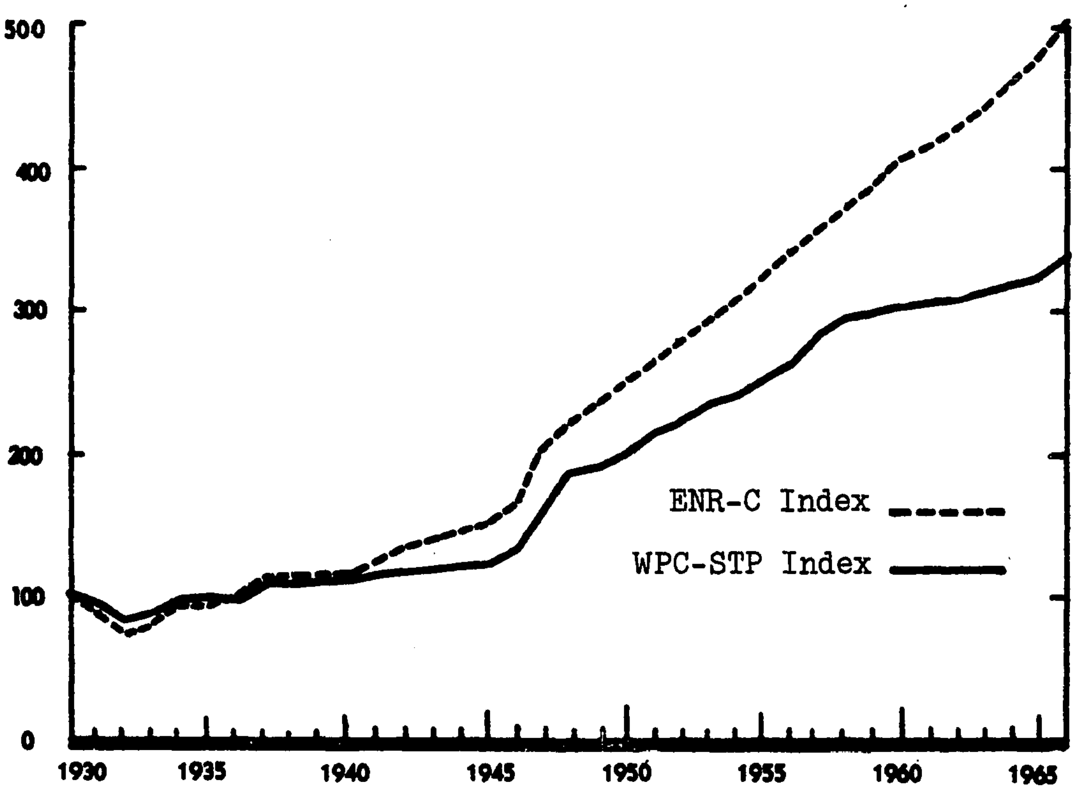

Figure 2. Sewage Treatment Plant Construction Cost Indexes, 1930=100. After Federal Water Pollution Control Administration.

Federal indexes, the FWPCA selected a 36-month period from January 1, 1957, to December 31, 1959, as a base for the cost index. This period conforms to the postwar base period for Federal indexes numbers as defined by the Bureau of Budget. This index may be used to evaluate past construction activity and to estimate future sewage treatment plant construction cost requirements for the 20 areas of influence in the United States as well as for the nation as a whole.

Because the WPC-STP Index is based on information peculiar to sewage treatment plant construction its usefulness for evaluating the construction costs of waste treatment plants is believed to be much greater than the ENR-C Index. The FWPCA has indicated (7) that revaluation of the index base will be undertaken at intervals of about 10 years. It is expected that the WPC-STP Index will reflect new developments in the field of waste treatment. Therefore, in this study it was decided to use WPC-STP Index to convert the cost data of treatment plants obtained from the various parts of the country to a common base. For example, a treatment plant in Virginia is in the area of influence of Baltimore. Thus, the cost index (1957-59=100) prevailing in Baltimore at the time of contract award was used to convert the plant's construction cost to a common base. The statistical procedures and techniques used for analyzing the data in Chapter VI, are discussed below.

Correlation Coefficients

The correlation coefficient between two random variables x and y which are following some unknown bivariate distribution is defined as:

$$r = \frac{\text{covariance}(x,y)}{[\text{variance}(x)\text{variance}(y)]^{1/2}} \tag{9}$$

The range of values of the correlation coefficient is from -1 to +1. A non zero value of the correlation coefficient indicates the association between the observed values of two variables but this does not necessarily imply that there is a relationship between the two variables. A zero value of the correlation coefficient implies that the variables are uncorrelated. However, a zero correlation coefficient can exist between dependent variables. This is possible because only the linear relationship is explained by the coefficient of correlation. A simple correlation coefficient may indicate a correlation between two variables because of the common relationship with another variable and not a relationship between each other.

If one has to deal with more than two variables at a time the partial correlation coefficient is used as a measure of linearity between observations of two variables with others held constant. Thus, the partial coefficient removes the influence of the other variables.

A partial correlation coefficient in terms of simple correlation coefficient is expressed as follows:

$$r_{12.3} = r_{21.3} = \frac{r_{12}-r_{13}r_{23}}{[(1-r_{13}^2)(1-r_{23}^2)]} \qquad (10)$$

where $r_{12.3}$ and $r_{21.3}$ have the same meaning and are the partial correlation coefficients of x_1 and x_2 with x_3 held constant. The right hand side of the equation has the simple correlation coefficients of the variables which are identified by subscripts.

Correlation coefficients were used as one of the procedures for screening to select the variables. Only those variables which significantly explained the variation in the response variable (dollars per MGD or dollars per PE) were selected. Correlation coefficients also were used to indicate which independent variables had a high correlation existing between their respective values and therefore the inclusion of either variable in the regression equation would give similar parameter values.

As more data on the cost of waste treatment systems becomes available, calculation of correlation coefficients would provide further insight into the variables. Correlation coefficients would help in determining which independent variables best explain changes in the response variable (unit cost), which variables do not appear to influence unit cost, and which variables may only appear to be significant in explaining the magnitude of unit cost because of its apparent high association with a variable that actually contributes to the regression of the response or the dependent variable.

Multiple Regression

A great deal of mathematical and statistical theory has been developed using linear equations. The examination of the data from a previous study (20) indicated the use of a multiple regression model in this study.

The linear model is an equation that involves dependent and independent variables, and parameters. It is an equation that is linear in the random variables and in the parameters.

The form that is used in this study is

$$Y = \beta_0+\beta_1X_1+\beta_2X_2 + \ldots\ldots +\beta_kX_k+e \qquad (11)$$

For the definition of the variables and parameters, see Equations 15 through 19.

The problem of best fitting a hyper plane to a set of joint observations on a dependent variable which is a linear function of several independent variables can be accomplished by the least-squares principle. Least-squares can also be applied to a number of more complex trend types such as certain types of transformations. For any linear model, least-squares minimizes the residual sum of squares and provides an unbiased linear estimate with minimum variance of the parameters.

It becomes tedious and complicated to analyze a model as the number of variables and observations increase. Fortunately, the use of matrices provides a compact way of treating the equations. The use of a digital computer is essential if investigation of many possible predictive equations is desirable.

The n equations can be set out compactly in a matrix notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{12}$$

where $\mathbf{Y}$ is a n by 1 vector of observations of a dependent variable, $\mathbf{X}$ is a n by k matrix of independent variables which explains the dependent variable's value, $\boldsymbol{\beta}$ is a k by 1 vector of unknown parameters, and $\mathbf{e}$ is a n by 1 vector of residuals. β_1 is the intercept term and indicates that each element of the first column of the matrix $\mathbf{X}$, equals unity. A sample of n sets of observations on y and k values of x can be represented by matrices as below.

$$\mathbf{Y} = \begin{matrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{matrix} \qquad \mathbf{X} = \begin{matrix} 1 & x_{21} & \cdots & x_{k1} \\ 1 & x_{22} & \cdots & x_{k2} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 1 & x_{2n} & \cdots & x_{kn} \end{matrix} \qquad \boldsymbol{\beta} = \begin{matrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{matrix} \qquad \mathbf{e} = \begin{matrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{matrix}$$

The method of least-squares minimizes the sum of the squared residuals $\mathbf{e}'\mathbf{e}$ in matrix form as $\sum_{i=1}^{n} e_i^2$. From Equation 12 we have $\mathbf{e} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$.

Therefore,

$$\mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \qquad (13)$$
$$= \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

The least-squares estimate of $\boldsymbol{\beta}$ is b. This when substituted in the above equation minimizes $\mathbf{e}'\mathbf{e}$. Differentiating and setting the resultant matrix equation to zero gives the normal equation,

$$(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \qquad (14)$$

For a review of the derivation of the distribution of pertinent statistics needed for estimation of the parameters in this model and for testing hypotheses about them and for the necessary assumptions see Johnston (32).

An inspection of the data indicated the inappropriateness of fitting a linear relationship. In such cases the two possibilities are to try an appropriate nonlinear fit directly to the data or else to make an initial transformation of the data such that the relationship between the transformed data is almost linear and principle of least-squares can be applied. The most commonly used transformations used to reduce complex models to linear ones are the logarithmic and the reciprocal. Draper and Smith (26) warn that when transformation involve the dependent variable special care must be taken to check that least-squares assumptions are not violated. These assumptions and their verifications will be discussed in the Section (Examination of Residuals) below.

An examination of the data obtained as a result of the sampling of municipal waste treatment plants and from the previous studies (16,20) indicated that a transformation of variables would be necessary. A frequency polygon of y values resulted in skewed frequency distributions of the sample to the right. A transformation commonly used for this type of distribution is the logarithmic transformation. While the forementioned procedure indicated the need for a transformation in the response or dependent variable it was not possible to determine the necessity for a transformation of the independent variable. Consequently, using the same sample data partial regression coefficients for the following linear equations would be computed. The form which gave the best fit was to be used as the estimation equation:

$$\ln Y = b_o + \sum_{i=1}^{k} b_i \ln X_i \qquad (i=1,2,\ldots n) \qquad (15)$$

$$\frac{1}{\ln Y} = b_o + \sum_{i=1}^{k} b_i \ln X_i \qquad \ldots\ldots\ldots\ldots \qquad (16)$$

$$\ln Y = b_o + \sum_{i=1}^{l} b_i X_i \qquad \ldots\ldots\ldots\ldots \qquad (17)$$

$$\frac{1}{Y} = b_o + \sum_{i=1}^{k} b_i X_i \qquad \ldots\ldots\ldots\ldots \qquad (18)$$

$$Y = b_o + \sum_{i=1}^{k} b_i X_i \quad \ldots\ldots\ldots\ldots\ldots \quad (19)$$

where Y is the dependent variable, X_i are the independent variables, and b_i are partial regression coefficients and are the estimates of the unknown parameters β_i in equation.

A modified stepwise regression procedure as opposed to typical stepwise regression was used in this study. A typical stepwise regression procedure uses a simple correlation matrix and enters into regression the independent variable with the largest absolute value correlation coefficient with the response or the dependent variable. The subsequent variables in the typical stepwise procedure are selected from the remaining independent variables by selecting the variable having the highest partial correlation coefficient with the response. The decision of acceptance or rejection of each new variable is based on the results of an overall and a partial F test. Often, the partial F criterion for each variable in the regression at any stage of calculation is evaluated and compared with a preselected F value. Any variable which provides a nonsignificant contribution is removed from the model. Given the regression equation, $Y = f(X_1, X_2)$, the stepwise regression examines the contribution previously added variable X_1 would have made if the newly added variable X_2 had been entered first. A variable once accepted into the regression may be rejected by this method.

For this study, the modification made to the typical stepwise regression procedure was that the variable's order of entry was determined by previous studies and the author's prior study and not the correlation matrix alone.

Selection of Best Equation

The square of the multiple correlation coefficient or the coefficient of multiple determination (R^2), the ratio of the sum of squares due to the regression to the total sum of squares, is one possible criterion for selection of the best equation. When one uses the same data to compute several forms of linear equations, the procedure for selecting the form that fits "best" is to choose the form which gives the highest coefficient of determination, R^2, or the highest R, the coefficient of multiple correlation. However, the significance can be misleading. This is possible particularly when only a small number of observations are used in computing the parameters of an equation. The increase in the value of R^2 may be more influenced by the increase in the number of independent variables rather than the related explanation contributed by the variables. The addition of another variable to a regression equation will either cause the sum of squares to increase or remain the same.

Draper and Smith (33) have indicated that if a set of observations on a dependent variable has only four different values a four parameter model will provide a perfect fit. However, the number of observations in this study is much larger than the number of parameters involved in the equations. Therefore, the R^2 values in all probability will reflect the explanation contributed by the variables.

The standard error of estimate is defined as the square root of the residual mean square. It takes into consideration the degrees of freedom of the residual and, therefore, is also used as indices for evaluating alternative regression equations. In previous studies (15,16) reviewed in Chapter II, the smallest error of estimate was the criterion used for selecting the form of the regression equation. However, the standard error of estimate does not give a measure of the proportion of the variation in the dependent variable which can be explained by, or is associated with, variation in the independent factor. The standard error of estimate is indicative of the closeness with which estimated values of the dependent variable agree with the original values which are used to determine regression coefficients.

Either standard error of estimate or the multiple correlation coefficient can be used as the test criterion for "best" fit if one has determined which set of independent variables provide such a fit. Both provide a comparison of the residual variation for each set of independent variables with the same standard deviation. In this study, at this stage, the problem is not one of determining which set of independent variables to choose so as to provide the "best" fit but rather to determine which transformation of the same set of variables gives the best equation. After transformation, the standard deviation of the response may change and, therefore, the standard error of estimate may not necessarily decrease with the increase in correlation. The former is an absolute measure whereas the latter is a relative measure. Hence, in this study the multiple correlation coefficient was used as the criterion for selecting the best equation.

A significant F-value, the ratio of the regression mean square to the residual mean square, indicates that the regression coefficients explain more of the variation in the data than expected by chance alone, under identical conditions, a certain percentage of time. A reference to the work of J. M. Wetz made by Draper and Smith (33) suggests that an equation should be regarded as a satisfactory predictor if the observed F-ratio should exceed by about four times the selected percentage point of the F-distribution. For example, for a ninety-five per cent confidence level F-value from the distribution tables for 1 and 22 degrees of freedom is 4.30. Therefore, F-ratio should be about 17.20 (4x4.30) or more for the fitted equation to be a satisfactory prediction model. As discussed in the next section, a necessary condition for the F-test is that the residuals be normally distributed. However, normal distribution of the residuals is not necessary condition for regression analysis.

The partial F-test is the ratio of the mean square corresponding to the addition of a new variable (equal to the sum of squares since it has one degree freedom), to the residual mean square. In this study, the partial or sequential F-test was used to determine if the addition of a new variable to regression explained more of the variation then would be expected by chance. The F-tests were used at a 5 per cent level of significance.

Examinations of Residuals

Residual is the difference between the observed and regression equation value of the dependent variables. As indicated above, there are certain basic assumptions made about the residuals when using least-squares regression analysis. These assumptions are: that the residuals are independent, have a constant variance and zero mean and if an F-test is used that they follow a normal distribution. Residuals, therefore, should be examined for verifying the forementioned assumptions.

A fitted model may be regarded as correct if the above assumptions do not appear to be violated on the basis of the data seen (34). This does not mean that one is concluding that the assumptions are correct. It means that on the basis of the data from the sample observations, it is seen that there is no reason to say that they are incorrect.

To test for independence of the successive disturbances "Durbin-Watson d Statistic" is used. Let

$s_p(p=1,2,...n)$ denote the residuals from a fitted least squares regression. Statistic d is defined

$$d = \frac{\sum_{p=2}^{n} (s_p - s_{p-1})^2}{\sum_{p=1}^{n} s_p^2} \tag{20}$$

Durbin and Watson have tabulated (35) lower and upper bounds, d_L and d_u, for various values of n and k (= number of independent variables). If $d < d_L$ one concludes that there is a positive correlation; if $d > d_u$ the assumption of independent residuals holds. However, if $d_L < d < d_u$ the test is inconclusive and further observations are needed. The tests on the sample observations did not give any evidence of correlation between the errors. Appendix C gives the values of d, d_L and d_u.

A plot of the estimated values of the response versus the corresponding residuals yields information on any abnormality, such as, variation in variance and systematic departure from the fitted equation. In the latter case low and high fitted values of the dependent variable produce negative and positive residuals, respectively. A plot indicated approximately a horizontal band and, therefore, abnormality was not suspected.

In order to check the assumption that the residuals are normally distributed these residuals are transformed into a unit normal deviate form. The resulting form of the residuals can be compared to an N(0,1) distribution. Using this technique approximately 95 per cent of the unit normal deviates would be expected to be within -1.96 to +1.96 or between the limits (-2,2). A sample computer print-out (Appendix D) clearly indicates a normal distribution of the residuals.

No checks can be made for the assumption that the residuals have a zero mean because the method of least-squares necessarily gives a zero sum of the residuals.

The runs test used to examine the pattern of residuals' signs to determine if the arrangement was unusual. This is done by comparing the observed values to the overall average. When the number of observations is greater than 20 a normal approximation of the underlying distribution is suggested by Draper and Smith (34) where:

$$u = \frac{2n_1n_2}{n_1+n_2} + 1 \qquad (21)$$

$$\sigma^2 = \frac{2n_1n_2[2n_1n_2-(n_1+n_2)]}{(n_1+n_2)^2(n_1+n_2-1)}$$

$$Z = \frac{(u-u+1/2)}{\sigma} \qquad (23)$$

with n_1 representing the number of positive residuals and n_2 the number of negative residuals.

u and σ^2 are the mean and variance of the discrete distribution; u, the number of runs. Z approximates the unit normal deviate.

The sample size was greater than 20 for each type of treatment plant and therefore the above approximation was made. Since no unusual statistical arrangement was found, it was concluded that no unconsidered variables changed levels within the sample range.

Dummy Variables

The variables considered in regression equations usually can take on values over a continuous range. In order to introduce a factor which has two or more distinct levels one can assign to these variables different levels to account for possible separate deterministic effects on the response. Such variables are called dummy variables. They can be used to represent temporal effects as well as qualitative variables (25). In the study an attempt was made to use the dummy variables for the quantification of the type of treatment plants. Results obtained as a result of the use of these dummy variables are discussed in the next chapter. Tables 9 and 10 give the (R^2) values obtained for the regression equations with dummy variables.

CHAPTER VI

RESULTS OF THE ANALYSES OF THE DATA

Having received the data as a result of a mail survey, analyses were made. It was indicated in Chapter IV that in an attempt to get a better response and to provide a check similar questionnaires were sent simultaneously to the Superintendents of the Sewage Treatment Plants and to the State Health Departments of the country.

Some disagreements were noted between the data reported by the Superintendents of the plants and the State Health Departments. Almost all differences noted were in the contract costs and effluent BOD values. On comparing the cost figures with those furnished by the FWPCA (29), agreement was noted between the State Health Department and the FWPCA figures. There was considerable difference between the cost figures furnished by the FWPCA and the ones reported by some of the Superintendents. Since the treatment plants were sampled from a list of plants constructed under PL-660 grants given by FWPCA, the cost data furnished by the latter was considered reliable. The Superintendents of the Treatment Plants generally have very little to do with the cost of construction of the plants. Therefore, some of them do not keep up-to-date and complete records of the plant costs and the cost figure reported in some instances might have been an approximation. This thinking is supported by the fact that some of the questionnaires were returned without the cost information or with the remark that the information was not available.

In a few instances, disagreement was also noted on the effluent BOD values. The operators of some of the treatment plants reported lower effluent BOD values than those reported by the State Health Departments. Here again, the figures reported by the State Health Departments would appear to be more reliable because most State Health Departments being the regulatory agencies, keep complete plant performance data reported by their District Engineers.

From the questionnaires received, the author noted that many treatment plants still do not have laboratory facilities to evaluate plant performance and/or do not have reliable and up-to-date records. One of the items in the questionnaire mailed concerned the current influent and effluent BOD values. It was surprising to note that out of 563 completed questionnaires received only 189 provided information on the influent and effluent BOD. The remaining questionnaires, 427 of them were returned either with the remark that no laboratory facilities were available or that no records were kept on the influent and effluent analysis.

Even the State Health Departments were unable to furnish this information in most cases.

Regression Equations

Regression equations using all possible and reasonable combinations of explanatory variables were developed to explain the response, unit cost of construction. Variables used in the regression equations were:

Y' = Construction cost per design PE, in 1957-59 dollars.

Y'' = Construction cost per design MGD, in 1957-59 dollars.

Y'_E of Y''_E indicate that explanatory variable X_4, the efficiency has been included in the equation, Y' and Y'' defined as above.

Y''_P = Construction cost per design MGD of the primary treatment plants treating industrial wastes, in 1957-59 dollars.

X_1 = Design PE.

X_2 = Design flow in MGD

X_3 = Design BOD of the influent in MG/L.

X_4 = BOD removal efficiency $\dfrac{\text{BOD influent - BOD effluent X 100}}{\text{BOD influent}}$

Y''_S = Construction cost per design MGD of the secondary treatment plants treating industrial wastes, in 1957-59 dollars.

D_1 and D_2 are dummy variables and have been explained in Tables 9 and 10.

X_5 = Effluent BOD in MG/L.

The procedures and criteria discussed in Chapter V were used to develop and evaluate the regression equations. Data on each type of treatment were used to derive several forms of linear equations. Typical regression equations using the above variables are presented in Tables 4 through 10. A discussion of the equations derived for each type of plant with different combinations of variables is presented below. As discussed in Chapter V, the criterion for the selection of the form that fits "best" is to choose the form which gives the highest coefficient of determination, R^2, or the highest R, the coefficient of correlation. The sequential F-test, using a 5 per cent significance level, was used to justify the acceptance of each variable into the regression equations.

Primary Treatment Plants

Data on 102 primary treatment plants were obtained from all over the country for the study. Based on the argument presented in Chapter III, the variables cost per PE (Y'), and the design PE (X_1), were not used to derive the regression equations. It was believed that it would be more meaningful to compute an equation which would give unit construction cost in terms of flow. Therefore, the relationship between the variables cost per MGD (Y''), and the flow in MGD (X_2), was studied. Several forms of the linear equation were derived. Equation 24 has the highest R value and is preferred over other equations for estimating the unit construction cost of the primary treatment plants. Table 4 presents the results of the analysis.

Stabilization Ponds

Equations relating response variables Y' and Y" and the explanatory variables X_1, X_2, and X_3, and X_1, X_2, and X_4 were derived in various forms as discussed in Chapter V. Only 12 of the 157 plants reported the current performance of the treatment plant (in terms of the BOD removal). So only 12 observations were available on the variable X_4. Therefore, equations with the explanatory variable X_4 were based on a sample size of only 12. The screening of the correlation matrix indicated a high correlation between X_1 and X_3. Since BOD is one of the parameters used in computing the PE the high correlation between the two was not surprising. X_3 was therefore dropped altogether from the regression equations. A sequential F-test indicated the nonsignificance of the variable X_4, and all equations with the variable X_4 were rejected.

The regression equations with explanatory variables X_1 and X_2 were developed. Of these, Equations 34 and 39 have the highest R values. These equations are therefore suggested for estimating the unit cost of construction and are presented in Table 5.

TABLE 4

EQUATIONS CONSIDERED FOR ESTIMATING THE UNIT CONSTRUCTION COST OF THE MUNICIPAL PRIMARY TREATMENT PLANTS

Equations		R	R^2	
$\underline{\ln Y''=12.42265-0.38520\ln X_2}$	(**)	$\underline{0.783}$	0.613	(24)
$\frac{1}{\ln Y''}=0.08079+0.00262\ln X_2$	(**)	0.778	0.605	(25)
$\ln Y''=12.42463-0.05052X_2$	(**)	0.617	0.380	(26)
$Y''=307258.31250-5345.71875X_2$	(**)	0.309	0.092	(27)
$\frac{1}{Y''}=0.00001+0.00003X_2$	(**)	0.568	0.322	(28)

(**) = Satisfied sequential F-test criterion and indicates no evidence of residual correlation.

TABLE 5

EQUATIONS CONSIDERED FOR STABILIZATION PONDS

Equations		R	R^2	
$\ln Y' = 5.98454 - 0.41688 \ln X_1 + 0.00534 \ln X_2$		0.560	0.313	(29)
$\ln Y'' = 5.98301 + 0.5833 \ln X_1 - 0.99483 \ln X_2$	(*)	0.640	0.409	(30)
$\ln Y'_E = 0.48488 - 1.1921 \ln X_1 + 0.47956 \ln X_2 + 2.8314 \ln X_4$		0.807	0.651	(31)
$\ln Y''_E = 0.48487 - 0.19215 \ln X_1 - 0.52043 \ln X_2 + 2.8314 \ln X_4$		0.807	0.651	(32)
$\frac{1}{\ln Y'} = 2.52594 - 0.23079 \ln X_1 + 0.24925 \ln X_2$		0.327	0.106	(33)
$\frac{1}{\ln Y''} = \underline{0.12918 - 0.00441 \ln X_1 + 0.00732 \ln X_2}$	(*)	$\underline{0.647}$	0.418	(34)
$\frac{1}{\ln Y'_E} = 1.1544 + 0.17780 \ln X_1 - .05035 \ln X_2 - 0.50412 \ln X_4$		0.734	0.538	(35)
$\frac{1}{\ln Y''_E} = 0.1575 + 0.0021 \ln X_1 + 0.0030 \ln X_2 - 0.01989 \ln X_4$		0.860	0.739	(36)
$\ln Y' = 3.1955 - 0.00005 X_1 + 0.03394 X_2$		0.474	0.224	(37)
$\ln Y'' = 12.42808 - 0.00001 X_1 - 0.34552 X_2$		0.363	0.131	(38)

TABLE 5--<u>Continued</u>

Equations		R	R^2	
$\underline{\frac{1}{Y'}=0.0511+0.00001X_1-0.06403X_2}$	(*)	$\underline{0.760}$	0.577	(39)
$\frac{1}{Y''}=0.00001-0.00000X_1+0.00001X_2$		0.248	0.061	(40)
$Y'=32.7355-0.00067X_1-1.44013X_2$		0.291	0.084	(41)
$Y''=362813.6875-2.39894X_1-59054.9375X_2$		0.236	0.055	(42)

(*) = Satisfies sequential F-test criterion but there is some evidence of residual correlation.

TABLE 6

EQUATIONS CONSIDERED FOR STANDARD-RATE TRICKLING FILTER

Equations		R	R^2	
$\ln Y'=7.90480-049927\ln X_1+0.04317\ln X_2$		.806	0.650	(43)
$\underline{\ln Y''=7.9049+0.50072\ln X_1-0.95682\ln X_2}$	(**)	$\underline{.867}$	0.751	(44)
$\ln Y'_E=15.6896-1.34738\ln X_1+0.93717\ln X_2+0.04231\ln X_4$		0.871	0.758	(45)
$\ln Y''_E=15.6895-0.34737\ln X_1-.06284\ln X_2+0.04231\ln X_4$		0.830	0.698	(46)
$\frac{1}{\ln Y'}=-0.09277+0.04424\ln X_1-0.00385\ln X_2$		0.759	0.576	(47)
$\frac{1}{\ln Y''}=0.11458-0.00374\ln X_1+0.00655\ln X_2$	(*.)	0.860	0.739	(48)
$\frac{1}{\ln Y'_E}=-0.4160+0.07706\ln X_1-0.04653\ln X_2+0.00046\ln X_4$		0.832	0.692	(49)
$\frac{1}{\ln Y''_E}=0.06319+0.00181\ln X_1+0.00065\ln X_2-0.00020\ln X_4$		0.825	0.680	(50)
$\ln Y'=3.9739-0.00004X_1+0.01387X_2$		0.661	0.436	(51)
$\ln Y''=13.1956+0.00002X_1-0.48421X_2$		0.681	0.463	(52)

TABLE 6--Continued

Equations	R	R^2	
$\ln Y'_E=4.05163-0.00014X_1+0.97866X_2+$			
$0.00125X_4$	0.641	0.410	(53)
$\ln Y''_E=13.17069-0.00003X_1-$			
$0.09898X_2+0.00217X_4$	0.679	0.461	(54)
$\frac{1}{Y'}=0.0183+0.00000X_1-0.00557X_2$	0.780	0.608	(55)
$\frac{1}{Y''}=0.00000-0.00000X_1+0.00000X_2$	0.507	0.257	(56)
$\frac{1}{Y'_E}=0.0204+0.0000X_1-0.01581X_2-$			
$0.00003X_4$	0.638	0.407	(57)
$\frac{1}{Y''_E}=0.0000-0.0000X_1+0.0000X_2-$			
$0.0000X_4$	0.656	0.430	(58)
$Y'=57.972-0.00094X_1-1.96928X_2$	0.629	0.395	(59)
$Y''=597357.68+0.41803X_1-$			
$116894.9375X_2$	0.551	0.304	(60)
$Y'_E=56.933-0.00877X_1+67.737X_2+$			
$0.19194X_4$	0.511	0.261	(61)
$Y''_E=540436.6875-32.319X_1+$			
$150139.625X_2+1538.3889X_4$	0.608	0.369	(62)

(**) = Satisfies sequential F-test criterion and indicates no evidence of residual correlation.

(*.) = Satisfied sequential F-test criterion but test to check the correlation of the residuals was not made.

TABLE 7

EQUATIONS CONSIDERED FOR HIGH-RATE TRICKLING FILTERS

Equations		R	R^2	
$\ln Y' = 9.39389 - 0.64431 \ln X_1 + 0.35570 \ln X_2$	(**)	0.673	0.453	(63)
$\ln Y'' = 9.39381 + 0.35571 \ln X_1 - 0.64431 \ln X_2$	(**)	0.731	0.534	(64)
$\ln Y'_E = 12.13465 - 0.84553 \ln X_1 + 0.53106 \ln X_2 - 0.20051 \ln X_4$		0.707	0.499	(65)
$\ln Y''_E = 12.13470 + 0.15446 \ln X_1 - 0.46894 \ln X_2 - 0.20051 \ln X_4$		0.684	0.468	(66)
$\frac{1}{\ln Y'} = -0.20818 + 0.05497 \ln X_1 - 0.02935 \ln X_2$	(*.)	0.644	0.414	(67)
$\frac{1}{\ln Y''} = 0.10049 - 0.00232 \ln X_1 + 0.00414 \ln X_2$	(*.)	0.728	0.529	(68)
$\frac{1}{\ln Y'_E} = -0.44116 + 0.07195 \ln X_1 - 0.04470 \ln X_2 + 0.01792 \ln X_4$		0.657	0.431	(69)
$\frac{1}{\ln Y''_E} = 0.08416 - 0.00109 \ln X_1 + 0.00308 \ln X_2 + 0.00114 \ln X_4$		0.681	0.463	(70)
$\ln Y' = 3.75422 - 0.00003 X_1 + 0.09126 X_2$	(*.)	0.573	0.328	(71)
$\ln Y'' = 12.93504 - 0.00001 X_1 - 0.07624 X_2$		0.503	0.252	(72)

TABLE 7--Continued

Equations		R	R^2	
$\ln Y'_E=4.56643-0.00003X_1+$ $0.08663X_2-0.00991X_4$		0.588	0.345	(73)
$\ln Y''_E=13.313-0.00001X_1-0.06469X_2-$ $0.00465X_4$		0.521	0.271	(74)
$\frac{1}{Y'}=0.02661+0.0000X_1-0.0024X_2$	(*•)	0.50	0.250	(75)
$\frac{1}{Y''}=0.0000-0.000X_1+0.0000X_2$		0.49	0.240	(76)
$\frac{1}{Y'_E}=0.00015+0.0000X_1-0.0022X_2+$ $0.00036X_4$		0.441	0.194	(77)
$\frac{1}{Y''_E}=0.0000+0.0000X_1+0.0000X_2+$ $0.0000X_4$		0.458	0.209	(78)
$Y'=47.935-0.00127X_1+6.3517X_2$	(*•)	0.629	0.395	(79)
$Y''=470924.06-1.6995X_1-$ $22850.875X_2$		0.370	0.136	(80)
$Y'_E=92.549-0.00132X_1+6.309X_2-$ $0.5110X_4$		0.685	0.469	(81)
$Y''_E=666210.56-2.7983X_1-$ $17703.152X_2-2339.193X_4$		0.431	0.185	(82)

(**) = Satisfies sequential F-test criterion and indicates no evidence of residual correlation.

(*•) = Satisfies sequential F-test criterion but test to check the correlation of the residuals was not made.

TABLE 8

EQUATIONS CONSIDERED FOR ACTIVATED SLUDGE TREATMENT PLANTS

Equations		R	R^2	
$\ln Y'=8.53953-0.53893\ln X_1+0.26243\ln X_2$	(**)	0.726	0.527	(83)
$\ln Y''=8.53987+0.46104\ln X_1-0.73754\ln X_2$	(**)	0.733	0.537	(84)
$\ln Y'_E=9.95789-0.84273\ln X_1+0.57640\ln X_2+0.31306\ln X_4$		0.768	0.589	(85)
$\ln Y''_E=9.95797+0.15727\ln X_1-0.42359\ln X_2+0.31306\ln X_4$		0.704	0.495	(86)
$\frac{1}{\ln Y'}=-0.18753+0.05216\ln X_1-0.02837\ln X_2$	(*.)	0.693	0.480	(87)
$\frac{1}{\ln Y''}=0.1054-0.00293\ln X_1+0.0046\ln X_2$	(*.)	0.73	0.532	(88)
$\frac{1}{\ln Y'_E}=-0.37085+0.09327\ln X_1-0.06865\ln X_2-0.04381\ln X_4$		0.708	0.501	(89)
$\frac{1}{\ln Y''_E}=0.09725-0.00095\ln X_1+0.00258\ln X_2-0.00229\ln X_4$		0.692	0.478	(90)
$\ln Y'=3.91785+0.0000X_1-0.08894X_2$		0.419	0.175	(91)
$\ln Y''=13.15394+0.00001X_1-0.15659X_2$	(*.)	0.445	0.198	(92)

TABLE 8--Continued

Equations		R	R^2	
$\ln Y_E'=3.23913+0.0000X_1-0.03988X_2+0.00614X_4$		0.364	0.132	(93)
$\ln Y_E''=12.74705+0.00001X_1-0.13485X_2+0.0036X_4$		0.456	0.207	(94)
$\frac{1}{Y'}=0.02494+0.00000X_1+0.00149X_2$		0.413	0.170	(95)
$\frac{1}{Y''}=0.0000-0.000X_1+0.0000X_2$	(*.)	0.396	0.156	(96)
$\frac{1}{Y_E'}=0.07357+0.000X_1-0.00087X_2-0.00049X_4$		0.379	0.143	(97)
$\frac{1}{Y_E''}=0.00001-0.0000X_1+0.0000X_2-0.0000X_4$		0.400	0.160	(98)
$Y'=61.075+0.00025X_1-5.04907X_2$		0.230	0.053	(99)
$Y''=638408.43750+3.98415X_1-71568.43750X_2$		0.192	0.036	(100)
$Y_E'=55.1477+0.00015X_1-3.274X_2-0.02718X_4$		0.264	0.069	(101)
$Y_E''=643328.187+3.0577X_1-55252.20X_2-1173.664X_4$		0.351	0.123	(102)

(**) = Satisfies sequential F-test criterion and indicates no evidence of residual correlation.

(*.) = Satisfies sequential F-test criterion but test to check the correlation of the residuals was not made.

TABLE 9

EQUATIONS CONSIDERED WITH DUMMY VARIABLES FOR ESTIMATING THE UNIT CONSTRUCTION COST OF THE SECONDARY MUNICIPAL WASTE TREATMENT PLANTS

Equations		R	R^2	
$\ln Y' = 8.32342 - 0.53216 \ln X_1 + 0.21725 \ln X_2 + 0.01535 D_1 + 0.12031 D_2$	(**)	0.728	0.530	(103)
$\ln Y'' = 8.32390 + 0.46780 \ln X_1 - 0.78271 \ln X_2 + 0.01535 D_1 + 0.12031 D_2$	(**)	0.768	0.589	(104)

Dummy Variables		Plants
D_1	D_2	
0	0	Standard-rate trickling filter plants
1	0	High-rate trickling filter plants
0	1	Activated sludge treatment plants.

(**) = Satisfies sequential F-test criterion and indicates no evidence of residual correlation.

TABLE 10

EQUATIONS CONSIDERED WITH DUMMY VARIABLES FOR ESTIMATING THE UNIT CONSTRUCTION COST OF THE PRIMARY AND SECONDARY TREATMENT PLANTS TREATING INDUSTRIAL WASTES

Equations		R	R^2	
$\ln Y''_P = 12.93509 - 0.09734 \ln X_2 - 2.09333 D_1 - 0.22875 D_2$	(**)	0.806	0.649	(105)
$\ln Y''_S = 12.83150 - 0.56827 \ln X_2 + 0.00311 \ln X_4 + 0.09035 D_1 + 0.42049 D_2$		0.814	0.662	(106)
$\ln Y''_S = 11.99492 - 054896 \ln X_2 + 0.20091 \ln X_3 + 0.00369 \ln X_5 - 0.10790 D_1 - 0.10706 D_2$		0.810	0.656	(107)
$\ln Y''_S = 11.99740 - 0.54917 \ln X_2 + 0.20309 \ln X_3 - 0.10770 D_1 - 0.10804 D_2$	(**)	0.820	0.672	(108)

Dummy Variables		Type of Wastes
0	0	Petroleum
1	0	Pulp and paper
0	1	Chemical

(**) = Satisfies sequential F-test criterion and indicates no evidence of residual correlation.

TABLE 11

VALID PREDICTION RANGES

Type of Treatment Plant	Valid Prediction Range (Design PE)
Primary Treatment Plant.	300-244,000
Waste Stabilization Ponds.	250-55,000
Standard-Rate Trickling Filter . .	215- 80,000
High-Rate Trickling Filter	600-125,000
Activated Sludge Plant	300-500,000

Standard-Rate Trickling Filters

A sample size of 67 plants was included in the study. Relationships of Y' and Y" with X_1, X_2, and X_3, and X_1, X_2 and X_4 were derived. An examination of the correlation matrix indicated a high correlation between X_1 and X_3, and therefore X_3 was eliminated. A sequential F-test resulted in the rejection of all the equations with X_4 as one of the explanatory variables. Those equations relating Y" to the variables X_1 and X_2 had generally a higher coefficient of correlation than the rest of the equations. Equations using cost per PE (Y') besides having lower R values did not, in general, satisfy the sequential F-test. Therefore, Equation 44 can be considered as the best equation for estimating the unit construction cost in terms of dollars per MGD. Table 6 presents this set of equations.

High-Rate Trickling Filters

Data on 123 plants were obtained in the category of the high-rate trickling filters. As in the case of standard-rate filters relationships of Y' and Y"; and X_1, X_2, and X_3, and X_1, X_2, and X_4 were evaluated. The explanatory variable X_3 was eliminated for the reason given above while discussing the standard-rate trickling filter equations. A sequential F-test justified the acceptance of each variable into the regression equations, Equations 63, 64, 67, 68, 71, 75, and 79. All other equations were rejected which, in essence, contained X_4 as one of the explanatory variables. Out of those equations which were accepted by the sequential F-test, Equations 63 and 64 were selected, one from the set of equations using Y', and another from the set using Y". To be sure, these equations have the highest correlation coefficient R among their respective sets. All the equations considered in the study have been presented in theTable 7.

Activated Sludge Treatment Plants

A sample size of 115 was included in this study. For the reasons given earlier while discussing the other types of plants explanatory variable X_3 was eliminated. Applying the sequential F-test, Equations 83, 84, 87, 88, 92, and 96 contained the acceptable variables. The rest of the equations were rejected by this criterion because these equations had the variable X_4 which was not acceptable. Out of those equations which were accepted as a result of the sequential F-test, Equations 83 and 84 using response variables cost per PE (Y'), and cost per MGD (Y") have the highest R value in their respective groups and therefore were chosen for estimating the unit cost of construction. Table 8 presents the results of the analysis.

Use of Dummy Variables

The data on standard-rate and high-rate trickling filter plants, and activated sludge treatment plants were combined and regression analysis was made with a sample size of 304. Since the logarithmic transformation of the variables gave the best "fit" dummy variables were used with this form of the linear equation.

The cost per MGD equation has slightly higher R value than the cost per PE equation. Both equations however, have the same explanatory variables. Either of the equations can be used for estimating the unit construction cost depending on whether one wants to estimate it in terms of organic or hydraulic loading. Table 9 presents both equations. The dummy variables included were D_1 and D_2. These variables were assigned the values 0 or 1 according to the treatment process to give the related intercept. Thus the computed intercept for either the high-rate trickling filters or the activated sludge treatment plants can be added to the basic equation to get the unit construction cost of either treatment plant. The basic equation itself would estimate the unit cost of construction of the standard-rate trickling filters. The combinations of the values assigned to the dummy variables are given in Table 9.

Industrial Waste Treatment Plants

A nationwide list of industrial waste treatment plants was not available and thus no random sampling procedure was followed. Very limited data did not permit computation of a regression equation which could predict the unit construction cost with a given precision.

Limited data was available (22) on plants treating petroleum, chemical and pulp and paper wastes. The data from these plants was combined to get a larger sample size. A total sample size of 25 primary treatment plants and 26 secondary plants was available. The analysis was performed using dummy variables. This procedure was adopted for both the primary and secondary type of plants.

On the basis of the argument put forth in Chapter III, only the relationship between Y" and X_2 was developed for estimating the cost of primary treatment plants. The dummy variables D_1 and D_2 were introduced to evaluate the intercepts. This Equation 105 is given in Table 10. For secondary treatment plants, three relationships were derived, one relating Y", and X_2, X_4, D_1 and D_2; the second, relating Y", and X_2, X_3, X_5, D_1 and D_2; and the third one relating, Y", and X_2, X_3, D_1 and D_2. The sequential F-test for the acceptance of each variable resulted in the rejection of Equations 106 and 107. Thus Equation 108 can be used for estimating the unit construction cost of a secondary treatment plant treating petroleum refinery wastes, chemical industry wastes, or pulp and paper wastes. Very little or no data were available on plants treating other types of wastes and therefore it was not possible to analyze and derive any equation for plants treating wastes other than the ones mentioned above.

General Discussion of Regression Equations

Of the various forms of linear equations derived (Tables 4 through 10), in general, the logarithmic form resulted in higher R values and satisfied the sequential F-test criterion with the exception of Equation 39 for stabilization ponds. Table 5 presents the various forms of linear equations derived for estimating the unit construction cost of the stabilization ponds. Equation 39 which estimates the cost per PE has a higher R value than the remaining equations It appears that since there are fewer ponds of capacities above 10,000 PE, the sample spread was not as large as in the case of other types of treatment plants.

Among the equations selected for estimating the unit construction costs Equation 44 for estimating the unit cost of the standard-rate trickling filter plants has the highest R value. This is possible because variations in the design practices of such plants are not as wide as in case of some other types of plants.

In almost all cases (with the exception of the stabilization ponds) the cost per MGD equations gave higher R values than cost per PE equations. One can think of two possible reasons for the greater correlation coefficient with the design flow: One, the hydraulic loading intensities are rough measures of process loading intensities because influent contains the organic matter to be removed (incorporated into PE, the organic loading measure, is the flow); two, the PE for this study was computed based on 0.17 lb per capita per day, but this may vary. In some cases it may go as high as 0.2 lb per capita per day. Such variations in per capita organic loading can result in a relatively low correlation coefficient. On the other hand, designed flows were taken directly from the completed questionnaires. However, from the equations selected either of the Equations (estimating unit cost in terms of organic load or flow) will give fairly satisfactory estimates.

The explanatory variable X_3 was highly correlated with another explanatory variable X_1 for all types of treatment plants. However, the latter variable had a better correlation with the response variable and so it was decided to eliminate X_3 and include X_1 as one of the explanatory variables. An attempt was made to include efficiency (X_4) as one of the explanatory variables for stabilization ponds and all the secondary type of treatment plants. In all cases it turned out to be insignificant. This has been discussed earlier in this chapter.

Those equations which satisfied the sequential F-test criterion and had the highest R values were tested for the independence of the residuals. This test has been discussed in Chapter V and the results are given in Appendix C. No evidence of correlation among the residuals was found except for the stabilization pond equations. It is difficult to pinpoint any reason at this stage. The ponds are very susceptible to local climatological conditions and therefore design practices vary widely from area to area. It is possible that variations in the design practices (loading may vary from 40 lbs of BOD to 100 lbs of BOD per acre per day) may be causing the error correlation. Johnston (36) points out that such correlation of error terms causes inefficient predictions, that is, predictions with needlessly large sampling variances. More detailed information concerning the design practices and costs will be needed before anything definite can be said about this situation.

The equations which are considered suitable for estimating the unit construction costs are given in Tables 12 through 16. The equations for estimating the unit construction costs of industrial waste treatment plants were derived from a very limited data and should not be considered very reliable. However, these equations will be helpful in giving some idea of the cost involved in constructing an industrial waste treatment plant. As more data becomes available these equations can be improved and the confidence interval can be established.

In the study costs were not included for such facilities as outfall sewers, pumping stations not contiguous to the plant, and administrative, engineering and legal services. The construction costs of a project represent approximately 80 per cent of the total costs (8). The costs which were excluded from this investigation can be incorporated in probable total project costs by increasing the construction estimates by a factor of 20 per cent. The study

TABLE 12

EQUATIONS FOR ESTIMATING CONSTRUCTION COST PER PE DESIGNED

	Type of Treatment Plant	
	Stabilization Ponds	High-Rate Trickling Filters
Regression Equations	$\frac{1}{Y'}=b_0+b_1X_1 b_2X_2$	$\ln Y'=b_0+b_1\ln X_1+b_2\ln X_2$
Standard Error of Estimate	$\frac{1}{S_{Y'}}=\sqrt{ReMS}[\frac{1}{n}+c_{11}x_1^2+2c_{12}x_1x_2]^{1/2}$	$S_{\ln Y'}=\sqrt{ReMS}[\frac{1}{n}+c_{11}x_1^2+c_{22}x_2^2+2c_{12}x_1x_2]^{1/2}$
Regression Coefficients		
b_0	0.0511	9.39
b_1	0.0001	-0.6443
b_2	-0.0640	0.3557
b_3	-	-
b_4	-	-
df	154	120
n	157	123
$\sqrt{ReMS}$	0.081	0.438

TABLE 12--Continued

	Type of Treatment Plant	
	Stabilization Ponds	High-Rate Trickling Filters
Deviations		
x_1	X_1-3798.29	$\ln X_1$-8.92
x_2	X_2-0.3438	$\ln X_2$-(-0.2440)
d_1	-	-
d_2	-	-
c_{ij}		
c_{11}	0.0000	0.0297
c_{12}	0.0000	0.0244
c_{22}	0.0146	0.0257
c_{13}	-	-
c_{23}	-	-
c_{33}	-	-
c_{14}	-	-
c_{24}	-	-
c_{34}	-	-
c_{44}	-	-

TABLE 13

EQUATIONS FOR ESTIMATING CONSTRUCTION COST PER PE DESIGNED

	Type of Treatment Plant	
	Activated Sludge Plants	All the Secondary Treatment Plants
Regression Equations	$\ln Y' = b_o + b_1 \ln X_1 + b_2 \ln X_2$	$\ln Y' = b_o + b_1 \ln X_1 + b_2 \ln X_2 + 0.015 D_1 + 0.120 D_2$
Standard Error of Estimate	$S_{\ln Y'} = \sqrt{ReMS}[\frac{1}{n} + c_{11}x_1^2 + c_{22}x_2^2 + 2c_{12}x_1x_2]^{1/2}$	$S_{\ln Y'} = \sqrt{ReMS}[\frac{1}{n} + c_{11}x_1^2 + c_{22}x_2^2 + c_{33}D_3^2 + c_{44}D_4^2 + 2c_{12}x_1x_2 + 2c_{13}x_1D_3 + 2c_{14}x_1D_4 + 2c_{23}x_2D_3 + 2c_{24}x_2D_4 + 2c_{34}D_3D_4]^{1/2}$
Regression Coefficients		
b_0	8.53	8.32
b_1	-0.5389	0.467
b_2	0.2624	-0.782
b_3	-	0.0150
b_4	-	0.1200
df	112	300
n	115	304
$\sqrt{ReMS}$	0.468	0.446

TABLE 13--Continued

	Type of Treatment Plant	
	Activated Sludge Plants	All the Secondary Treatment Plants
Deviations		
x_1	$\ln X_1 - 8.34$	$\ln X_1 - 8.60$
x_2	$\ln X_2 - (-0.8675)$	$\ln X_2 - (-0.6041)$
d_1	-	$D_1 - 0.4013$
d_2	-	$D_2 - 0.3782$
c_{ij}		
c_{11}	0.0460	0.0160
c_{12}	0.0452	0.0150
c_{22}	0.0478	0.0160
c_{13}		0.0002
c_{23}		0.0005
c_{33}		0.0140
c_{14}		0.0002
c_{24}		0.0005
c_{34}		0.0130
c_{44}		0.0140

TABLE 14

EQUATIONS FOR ESTIMATING CONSTRUCTION COST PER MGD DESIGNED

	Type of Treatment Plant	
	Primary Treatment Plants	Stabilization Ponds
Regression Equations	$\ln Y'' = b_0 + b_2 \ln X_2$	$\frac{1}{\ln Y''} = b_0 + b_1 \ln X_1 + b_2 \ln X_2$
Standard Error of Estimate	$S_{\ln Y''} = \sqrt{ReMS}[\frac{1}{n} + \frac{(x_2)^2}{252.2}]^{1/2}$	$S_{\frac{1}{\ln Y''}} = \sqrt{ReMS}[\frac{1}{n} + c_{11}x_1^2 + c_{22}x_2^2 + 2c_{12}x_1x_2]^{1/2}$
Regression Coefficients		
b_0	12.42	0.1291
b_1	-	-0.0044
b_2	-0.3852	0.0073
b_3	-	-
b_4	-	-
df	100	154
n	102	157
$\sqrt{ReMS}$	0.488	0.006

TABLE 14--Continued

	Type of Treatment Plant	
	Primary Treatment Plant	Stabilization Ponds
Deviations		
x_1	-	$\ln X_1 - 7.10$
x_2	$\ln X_2 - 0.4722$	$\ln X_2 - (-2.16)$
d_1	-	-
d_2	-	-
c_{ij}		
c_{11}	-	0.0248
c_{12}	-	0.0229
c_{22}	-	0.0250
c_{13}	-	-
c_{23}	-	-
c_{33}	-	-
c_{14}	-	-
c_{24}	-	-
c_{34}	-	-
c_{44}	-	-

TABLE 15

EQUATIONS FOR ESTIMATING CONSTRUCTION COST PER MGD DESIGNED

	Type of Treatment Plant	
	Standard-Rate Trickling Filters	High-Rate Trickling Filters
Regression Equations	$\ln Y'' = b_0 + b_1 \ln X_1 + b_2 \ln X_2$	$\ln Y'' = b_0 + b_1 \ln X_1 + b_2 \ln X_2$
Standard Error of Estimate	$S_{\ln Y''} = \sqrt{ReMS}[\frac{1}{n} + c_{11}x_1^2 + c_{22}x_2^2 + 2c_{12}x_1x_2]^{1/2}$	$S_{\ln Y''} = \sqrt{ReMS}[\frac{1}{n} + c_{11}x_1^2 + c_{22}x_2^2 + 2c_{12}x_1x_2]^{1/2}$
Regression Coefficients		
b_0	7.90	9.39
b_1	0.5007	0.3557
b_2	-0.9568	-0.6443
b_3	-	-
b_4	-	-
df	64	120
n	67	123
$\sqrt{ReMS}$	0.383	0.437

TABLE 15--Continued

	Type of Treatment Plant	
	Standard-Rate Trickling Filters	High-Rate Trickling Filters
Deviations		
x_1	$\ln X_1 - 8.36$	$\ln X_1 - 8.92$
x_2	$\ln X_2 - (-0.8249)$	$\ln X_2 - (-0.2440)$
d_1	-	-
d_2	-	-
c_{ij}		
c_{11}	0.0848	0.0297
c_{12}	0.0740	0.0244
c_{22}	0.0751	0.0257
c_{13}	-	-
c_{23}	-	-
c_{33}	-	-
c_{14}	-	-
c_{24}	-	-
c_{34}	-	-
c_{44}	-	-

TABLE 16

EQUATIONS FOR ESTIMATING CONSTRUCTION COST PER MGD DESIGNED

	Type of Treatment Plant	
	Activated Sludge Plants	All the Secondary Treatment Plants
Regression Equations	$\ln Y'' = b_o + b_1 \ln X_1 + b_2 \ln X_2$	$\ln Y'' = b_o + b_1 \ln X_1 + b_2 \ln X_2 + 0.015D_1 + 0.120D_2$
Standard Error of Estimate	$S_{\ln Y''} = \sqrt{ReMS}[\frac{1}{n} + c_{11}x_1^2 + c_{22}x_2^2 + 2c_{12}x_1x_2]^{1/2}$	$S_{\ln Y''} = \sqrt{ReMS}[\frac{1}{n} + c_{11}x_1^2 + c_{22}x_2^2 + c_{33}D_3^2 + c_{44}D_4^2 + 2c_{12}x_1x_2 + 2c_{13}x_1D_3 + 2c_{14}x_1D_4 + 2c_{23}x_2D_3 + 2c_{24}x_2D_4 + 2c_{34}D_3D_4]^{1/2}$
Regression Coefficients		
b_0	8.53	8.32
b_1	0.4610	0.467
b_2	-0.7375	-0.782
b_3	-	0.0150
b_4	-	0.1200

TABLE 16--Continued

	Type of Treatment Plant	
	Activated Sludge Plants	All the Secondary Treatment Plants
df	112	300
n	115	304
$\sqrt{ReMS}$	0.468	0.446
Deviations		
x_1	$\ln X_1 - 8.34$	$\ln X_1 - 8.60$
x_2	$\ln X_2 - (-0.8675)$	$\ln X_2 - (-0.6041)$
d_1	-	$D_1 - 0.4013$
d_2	-	$D_2 - 0.3782$
c_{ij}		
c_{11}	0.0460	0.0160
c_{12}	0.0452	0.0150
c_{22}	0.0478	0.0160
c_{13}	-	0.0002
c_{23}	-	0.0005
c_{33}	-	0.0140
c_{14}	-	0.0002
c_{24}	-	0.0005
c_{34}	-	0.0130
c_{44}	-	0.0140

does not include land costs because of their wide variations.

A few examples have been solved to illustrate the use of the regression equations. Tables 12 through 16 present the necessary parameters for estimating the unit costs by using the derived equations. In these Tables d_f is degrees of freedom corresponding to the residual mean square, n is the number of observations and ReMS is residual mean square. The 95 per cent confidence limits for an estimated cost value is given by,

$$CL = \ln Y \pm t S_{\ln Y} ,$$

where

CL = 95 per cent confidence limits,

lnY = estimated expected value for a given set of X's, (ln is the base e logarithm),

t = "student's t" value for confidence coefficient of 0.95 and degrees of freedom corresponding to the residual mean square, and $S_{\ln Y}$ = standard error of estimate for lnY. The dimensions of $S_{\ln Y}$ are logarithmic (base e) units. The antilog (base e) of lnY gives the expected construction cost in dollars per PE or MGD (depending on the equation used) and the antilog of the upper and lower values of the computed confidence limits gives the 95 per cent confidence limits.

Sample Calculations

A few examples are cited to illustrate the application of the equations. Assume that a city located within the area of influence of Kansas City desires to estimate the construction cost of a proposed municipal waste treatment plant. The design PE and flow of the influent are 12,000 and 1.34 MGD, respectively. The stream conditions are such that an effluent BOD greater than 30 MG/L is not allowed. Obviously a secondary treatment plant which can remove about 84 per cent of the influent BOD is needed. In order to be able to compare the construction costs of different types of conventional treatment plants and also to illustrate the use of the estimating equations it was decided to estimate the costs of construction of a primary treatment plant and conventional secondary plants namely; standard-rate and high-rate trickling filters, and an activated sludge plant. The equations given below, from Tables 4 and 9 are appropriate for estimating the expected costs.

Primary Treatment Plant

$$\ln Y'' = 12.42 - 0.3852 \ln X_2 \qquad (24)$$

$$X_2 = 1.34$$

$$= 12.42 - 0.3852 \ln 1.34$$

$$= 12.31$$

Antilog of 12.31 (an estimate of the expected cost of primary treatment plant = \$222,444)

From Table 14

$$x_2 = (\ln 1.34 - 0.4722) = 0.032$$

$$S_{\ln}Y'' = 0.488 \left[\frac{1}{102} + \frac{0.0010}{252.2}\right]^{1/2}$$

$$= (0.488)(0.0110)^{1/2}$$

$$= 0.049$$

$$= t.S_{\ln}Y''$$

$$= (2.0)(0.049)$$

$$= 0.098$$

95% confidence limits $= \ln Y' \pm tS_{\ln}Y'$

$$= 12.31 \pm 0.098$$

Antilog of these (95% confidence interval in 1957-59 dollars per MGD designed) = $200,887 to $244,856. The probability is 0.95 that the true mean construction cost per MGD for the primary treatment plant with a design flow of 1.34 MGD lies in the interval of $200,887 to $244,856 and the best estimate of the expected value is $222,444. To obtain the total expected cost for say August 1962, compute (WPC-STP Index for Kansas City = 103.49):

$$222,444\ (1.34)\ \frac{103.49}{100} = \$308,478$$

Lower limit of 95% confidence interval =

200,887 (1.34)(1.0349) = $278,583 (approximate)

Upper limit of 95% confidence interval =

244,856 (1.34)(1.0349) = $339,558 (approximate).

Costs for such facilities as outfall sewers, and administrative, engineering and legal services but not the land costs are incorporated in probable total project costs by increasing the construction estimates by a factor of 20 per cent. In this case, 20 per cent of the total expected cost of construction = \$61,695. Thus the cost of the treatment plant (including the above items) = \$308,478 + \$61,695 = \$370,173 (approximate).

Similarly construction costs of high-rate trickling filters, standard-rate filters and activated sludge treatment plants are computed. In order to illustrate the use of dummy variables the construction cost is computed by using the following equation:

$$\ln Y'' = 8.323+0.467\ln X_1-0.782\ln X_2+0.015\ D_1+0.120\ D_2 \quad (104)$$

$$X_1 = 12{,}000 \text{ PE}$$

$$X_2 = 1.34 \text{ MGD}$$

$$= 8.323+0.467\ln 12{,}000 - 0.782 \ln 1.34 + 0.015D_1 + 0.120D_2 = 12.488 + 0.015D_1 + 0.120D_2 .$$

To account for the different secondary processes the following values from Table 9 are substituted:

$D_1 = 0$ and $D_2 = 0$ for standard-rate trickling filters.

$D_1 = 1$ and $D_2 = 0$ for high-rate trickling filters.

$D_1 = 0$ and $D_2 = 1$ for activated sludge treatment plants.

Antilog of 12.488 (an estimate of the expected cost of the standard-rate trickling filter per MGD designed in 1957-59 dollars) = \$265,793 (approximate) antilog of 12.503 (an estimate of the expected cost of the high-rate trickling filter per MGD designed in 1957-59 dollars) = \$269,815 (approximate).

Antilog of 12.608 (an estimate of the expected cost of the activated sludge treatment plant per MGD designed in 1957-59 dollars) = \$299,050 (approximate)   From Table 16

$$x_1 = [\ln 12{,}000 - 8.603] = 0.789$$

$$x_2 = [\ln 1.34 - (-0.604)] = 0.896$$

$$d_1 = [1 - 0.401] = 0.599$$

$$d_2 = [1 - 0.378] = 0.622$$

$$S_{\ln}Y' = 0.446\ [\tfrac{1}{304} + 0.016(0.789)^2 + 0.016(0.896)^2 + 0.014(0.599)^2 + 0.014(0.622)^2 + 2(0.015)(0.789)(0.896) + 2(0.00020)(0.789)(0.599) + 2(0.00020)(0.789)(0.621) + 2(0.0005)(0.896)(0.599) + 2(0.0005)(0.896)(0.622) + 2(0.013)(0.599)(0.622)]^{1/2}$$

$$= 0.117$$

$$tS_{\ln}Y' = (2.0)(0.117) = 0.234.$$

95% confidence limits = lnY' ± tSlnY'

$$= 12.488 \pm 0.234$$

= 12.254 to 12.722 for standard-rate trickling filters

$$= 12.503 \pm 0.234$$

= 12.269 to 12.737 for high-rate trickling filters

$$= 12.608 \pm 0.234$$

= 12.374 to 12.842 for activated sludge treatment plants

Antilog of these give 95 per cent confidence intervals in 1957-59 dollars per MGD designed. These confidence intervals are given below:

$210,241 to $335,833 per MGD for the standard-rate trickling filters.

$213,478 to $340,830 per MGD for the high-rate trickling filters.

$236,832 to $378,582 per MGD for the activated sludge treatment plants.

To estimate the total expected cost of, say, a high-rate trickling filter in Kansas City to be designed for a PE of 12,000 and a flow of 1.34 MGD for August 1962 (WPC-STP Index for Kansas City = 1.034) multiply the expected unit cost by the design flow and the cost index (Appendix B):

$$269{,}815(1.34)\frac{103.49}{100} = \$374{,}170 \text{ (approximate)}$$

95 per cent confidence interval for this cost is:
$296,044 to $472,651 (approximately).

Following similar procedure the equation below estimates the construction cost for PE designed:

$$\ln Y' = 8.323 - 0.5321 \ln 12{,}000 + 0.2172 \ln 1.34 + 0.0153\, D_1 + 0.1203\, D_2 \qquad (103)$$

Substituting $D_1 = 1$ and $D_2 = 0$

$$= 3.40$$

Antilog of 3.40 (an estimate of the expected cost of the high-rate trickling filter per PE designed in 1957-59 dollars) = \$30.00 (approximate).

Total expected cost in Kansas City for the same design loadings (12,000 PE and 1.34 MGD):

$$30 \times \frac{103.49}{100} \times 12{,}000 = \$372{,}564 \text{ (approximate)}$$

$$95\% \text{ confidence limits} = \ln Y' \pm t\, S_{\ln Y'}$$

$$= 3.403 \pm 0.234$$

Antilog of this gives the confidence interval for the cost per PE designed in 1957-59 dollars. The range of the total estimated construction cost of just the plant itself (in August 1962 dollars) for a design PE of 12,000 is:

\$295,567 to \$471,542 (approximate)

It may be of interest to note that the estimated construction cost based on detailed estimates made by the FWPCA (7) for a model high-rate trickling filter plant in Kansas City was \$462,034.29 based on the August 1962 WPC-STP Cost Index. It may be pointed out here that one of the reasons for selecting an organic loading of 12,000 PE and a hydraulic loading of 1.34 MGD was to afford a comparison between the actual detailed cost estimates and those computed from the regression equations. The total cost figure computed from the detailed estimates lies within the predicted cost range. Also a comparison of the two estimated total plant construction costs, one based on the unit flow and second based on unit PE, are fairly close.

As done in previous examples, increase the estimated construction costs by a factor of 20 per cent to include the cost of outfall sewers, pumping stations, administrative, engineering and legal services.

Stabilization Ponds

As a further example suppose that a small town of 1,200 PE in the state of Oklahoma needs to treat its wastes to a satisfactory degree. Due to uncertainty of the future development and non-availability of adequate funds it was decided to construct a waste stabilization pond to treat the wastes. Assuming an average flow of 0.12 MGD, the cost was computed as follows:

$$\frac{1}{\ln Y''} = 0.1291 - 0.0044 \ln X_1 + 0.0073 \ln X_2 \qquad (34)$$

$$X_1 = 1{,}200 \text{ PE}$$

$$X_2 = 1.34 \text{ MGD}$$

$$= 0.1291 - 0.0044 \ln 1{,}200 + 0.0073 \ln 0.12$$

$$= 0.0824$$

or log Y" = 12.13

Antilog of 12.13 = (an estimate of the expected costs per MGD designed in 1957-59 dollars) = \$186,501 (approximate).

From Table 14

$$x_1 = [\ln 1{,}200 - 7.10] = - 0.0194$$

$$x_2 = [\ln 0.12 - (-2.16)] = 0.0410$$

$$\frac{1}{S_{\ln Y''}} = 0.006\ [\frac{1}{157} + 0.0248(0.0194)^2 + 0.0250(0.0410)^2 + $$
$$+ 2(0.0229)(-0.0194)(0.0410)]^{1/2}$$
$$= 0.006(0.0063)^{1/2} = 0.00048$$

$$t \frac{1}{S_{\ln Y''}} = (2.0)(0.00048) = 0.00096$$

$$95\% \text{ confidence limits} = \frac{1}{\ln Y''} \pm \frac{t}{S_{\ln Y''}}$$
$$= 0.0824 \pm 0.00096$$

Inverse and antilog gives (95% confidence interval in 1957-59 dollars per MGD designed) \$161,524 to \$215,036. Total expected cost (WPC-STP Index for August 1962 = 97.23).

$$= 186{,}510(0.12)\ \frac{97.23}{100} = \$21{,}760 \text{ (approximate)}$$

Lower limit of 95% confidence interval $161{,}524(0.12)\ \frac{97.23}{100}$

$$= \$18{,}846 \text{ (approximate)}$$

Upper limit of 95% confidence interval $215{,}036(0.12)\ \frac{97.23}{100}$

$$= \$25{,}089 \text{ (approximate)}$$

Since the major portion of the State of Oklahoma is considered under the area of influence of Dallas, Texas, the cost index used here is the one which was effective in August 1962 in Dallas.

Following a similar procedure, construction cost per PE designed is estimated as illustrated below:

$$\frac{1}{Y'} = 0.0511 + 0.00001(1{,}200) - 0.0640(0.12) \qquad (39)$$

$$X_1 = 1{,}200 \qquad\qquad = 0.0554$$

$$X_2 = 0.12$$

Inverse of 0.0554 (an estimate of the expected cost per PE designed in 1957-59 dollars) = \$18.04.

$$\frac{1}{S_{Y'}} = 0.081[\frac{1}{157} + 0.0000(-2598.29)^2 + 0.0146(-0.2238)^2]^{1/2}$$

$$= 0.081(0.00710)^{1/2} = 0.0068$$

$$t\frac{1}{S_{Y'}} = (2.0)(0.0068) = 0.0316$$

$$95\% \text{ confidence limits} = \frac{1}{Y'} \pm t\,\frac{1}{S_{Y'}}$$

$$= 0.0554 \pm 0.0136$$

Inverse of these (95% confidence interval in 1957-59 dollars per PE designed) = \$14.49 to \$23.91.

To obtain the total cost (WPC-STP Index = 97.23):

Expected value = $18.04(1{,}200)\,\frac{97.23}{100}$ = \$21,053 (approximate).

Lower limit of 95% confidence interval = $14.49(1{,}200)\,\frac{97.23}{100}$

= \$16,906 (approximate).

Upper limit of 95% confidence interval $= 23.91(1{,}200)\ \frac{97.23}{100}$

$= \$27{,}897.$

The contract cost reported for a stabilization pond of similar capacity constructed in 1962 to treat the wastes on a comparable community in the State of Oklahoma was \$21,500 (37).

Costs for outfall sewers, and administrative, engineering and legal services are incorporated in probable total project costs by increasing the construction estimates by a factor of 20 per cent.

CHAPTER VII

SUMMARY AND CONCLUSIONS

Summary

This Study

In order to enable engineers, planners, public officials and economists charged with planning and development of water resources to calculate the preliminary estimates of the construction cost of a waste treatment facility for a city or a region equations are presented which would give reliable estimates. The use of equations will provide only the preliminary estimates prior to any detailed engineering studies and is not meant to replace such studies for any given project.

It was thought that the efficiency of the treatment plant variable would contribute a significant amount of the regression, since one would expect that more efficient plants (removing greater percentage of the influent BOD) would cost relatively more. It is possible that division of the secondary treatment plants into three categories: standard-rate and high-rate trickling filter plants and activated sludge treatment plants; and very limited data on the efficiency may have masked the effect of efficiency on the cost.

The economies of scale affect the unit construction cost of different types of secondary treatment plants by a factor which changes very little from type to type. In other words, the unit cost does change with the type of treatment plant, for example an activated sludge treatment plant's cost might differ from high-rate trickling filter plant, but change in the unit cost due to scale factor is not significantly different between the two plants. This fact has been demonstrated by the use of the dummy variables in Chapter VI.

There is some risk in extrapolating beyond the range of values for the independent variables observed in the study. The same regression function may not apply to values outside the range and the estimates may be either too large or too small. Approximate valid range is given in Table 11.

Suggestions for Future Work

To insure better unit cost estimates the regression equations must be revaluated at regular intervals of five to ten years. Such revaluation may be necessitated due to changes in design practices and new treatment methods. The equations for estimating the construction costs of the industrial waste treatment plants need to be substantiated with adequate data.

Due to lack of cost data for plants constructed before 1957, effect of time on future costs as a result of technological advancements could not be evaluated. The WPC-STP Construction Cost Index (7) reflects to some extent the current practices in the sewage treatment plant industry and new developments or uses of process equipment. However, if one could obtain data on old plants and also newly constructed treatment plants, it may be possible to predict the construction cost of a future treatment plant with a given efficiency based on the study of past trends. Needless to say, such an attempt should be based on adequate data analyzed by modern statistical procedures.

In order to get more insight into the cost-BOD removal relationship detailed cost estimates could be made of a plant with a certain capacity starting from preliminary treatment units to secondary treatment units including sedimentation tanks. Such estimates will have to be made for various detention times and different recirculation factors, both of which will provide different BOD removal efficiencies and which, in turn, can be related to different costs.

Conclusions

The author has developed a technique for estimating the unit construction costs of the waste treatment plants. It takes into account regional differences, volume and strength. The explanatory variables used are PE and flow in

MGD. The former variable is a measure of organic loading and the latter hydraulic loading. Equations providing unit costs in terms of PE and flow have been presented. In a deterministic sense, volume is most important in municipal fixed cost estimates, in that the processes consist of mainly tanks and pipes, are directly related to volume. On the other hand, highly concentrated wastes, though still responsive to volume, do involve aeration gear responsive to strength.

In the previous studies, Velz (10), Diachishin (12), and Logan, et al. (17) related the cost to flow. These studies did not take into account the organic loading even though secondary treatment plants are designed to remove the organic matter. The ENR-C Index was used to adjust all costs. The authors did not take into consideration regional differences in the total costs. Cost estimating equations were also derived by Rowan et al. (16) for different types of treatment processes. The authors related costs per capita to design population. Since municipal sewage generally contains some industrial waste, design PE rather than design population is a better explanatory variable. Also, the flow which is one of the important explanatory variables was not considered. Regional differences were taken into consideration by using 20 ENR regions.

ENR-C Index rather than WPC-STP Index was used in adjusting the costs. It has been pointed out in Chapter V that WPC-STP Index is representative of cost changes peculiar to municipal sewage treatment plant construction.

Park (19) considered both hydraulic and organic loadings for estimating the construction costs. Certain flat assumptions concerning BOD and flow were made and have been discussed in Chapter II. The data was not analyzed by multiple linear regression techniques.

Assenzo (20) attempted to derive multiple regression equations for each of the 9 USPHS administrative regions. He considered BOD, PE, and type of treatment plant as the explanatory variables and cost per PE produced, cost per PE treated, and cost per capita as the response variables.

Like other workers, Assenzo considered ENR-C Index. Also a single estimating equation for all types of treatment plants was developed. It is believed that separate equations for primary and secondary plants, and stabilization ponds provide better estimates of unit costs. Table 17 presents a comparison of some of the studies.

The developed equations can be used to predict future, short term cost, by using projected flows and strengths; for longer terms, technological changes must be imputed, and there was not sufficient data available to analyze it statistically.

Table 18 gives the projected ratios of PE to population and sewage to water use. The ratios have been developed from studies made by Reid (38), and Wells and

TABLE 17

A COMPARISON OF STUDIES BY VARIOUS AUTHORS

Study	Velz	Diachishin	Rowan *et al*.
Consideration of regional differences in construction costs	None	None	20 areas ENR-C Index 1913=100
Explanatory variables used to account for the size of plant	Flow	Flow	Design population
Techniques used to take different processes into consideration	Single equation, primary 35% BOD removal, chemical coagulation 65% BOD removal, trickling filter 85% BOD removal, activated sludge 90% BOD removal.	Separate equations for primary and secondary treatments	Separate equation for each type of process
Equation format	$Y=aX^b$	--	$Y=ax^b$

TABLE 17--Continued

Logan *et al.*	Assenzo	Shah
None for field studies	9 USPHS regions based on ENR-C Index = 100	20 regions WPC-STP Index=100
Flow	PE, design population, type of plant, BOD of effluent	PE, flow
Separate equation for each type of process	Single equation, variable people per effective area accounts for different processes	Separate equations for each type of process. Dummy variables used to account for different types of secondary treatment plants
$Y=ax^b$	$\ln Y=b_o+\sum b_i \ln X_i$	$\ln Y=b_o+\sum b_i \ln X_i$ for primary and secondary plants $\ln Y=b_o+\sum b_i \ln X_i+D_1+D_2$ for secondary municipal, and primary and secondary industrial wastes $\frac{1}{\ln Y}=b_o+\sum b_i \ln X_i$ $\frac{1}{Y}=b_o+\sum b_i X_i$ for stabilization ponds

TABLE 18

PROJECTED RATIOS

Year	PE to Population	Sewage to Water Use
1900	1.0	--
1954	1.36	--
1960	1.42	0.62
1980	1.50	0.59
2000	1.76	0.57
2020	1.87	0.55

Gloyna (39). For any city one can usually develop the population and water use projection from the past records. The above table can thus be used to estimate the PE and sewage volume for say year 2000.

Figure 3 presents a comparison between the costs of secondary treatment plants estimated by Velz (10), Diachishin (12), and the author. Selected points were directly taken from the graphs presented by Velz and Diachishin. The costs were then adjusted to 1957-59 dollars. The Velz's curve is based on cost information obtained for treatment plants constructed before 1948 and Diachishin's curve is based on cost information obtained for plants constructed between 1947 to 1957. Author's curve is based on the data obtained for plants constructed between 1957-67. The graph indicates that the construction costs of older plants are higher than newer plants when compared at the same dollar value. The most recent data gathered by the author indicates higher capital costs of the plants designed for a flow less than 1 MGD and lower costs for plants designed above 1 MGD when compared with the cost information for the older plants.

Better equipment for many of the unit processes, new and improved methods developed as a result of research, and economies of scale have helped to reduce the capital costs in recent times. Whereas improved technology has lowered the capital costs of higher capacity plants it has

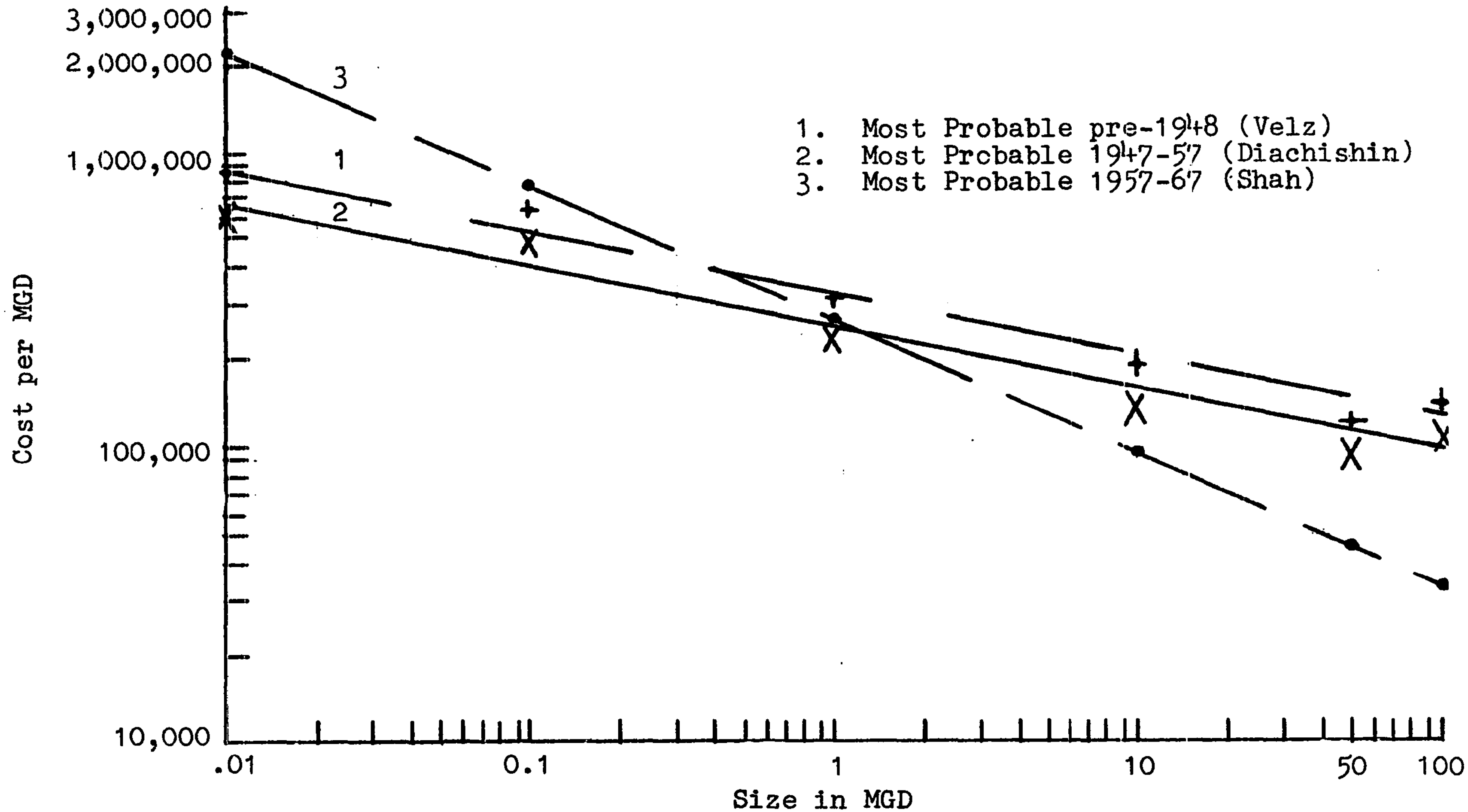

Figure 3. Construction cost of secondary sewage treatment plants. All data adjusted to WPC-STP Index (1957-59 = 100).

not lowered the costs of lower capacity plants. In fact increased requirements of a fixed nature such as the need for laboratories etc. have increased costs.

Generally, the smaller size plants are constructed so that they are easy to operate, although the costs for such plants are higher. For example, standard-rate trickling filters are expensive to build for capacities lower than 1 MGD. The activated sludge treatment plants and high-rate trickling filter plants are relatively economical to construct (particularly for higher capacities) but need skilled operators. Small size communities can not afford high operation costs. This is one of the main reasons why the technological developments in the field of waste treatment have not effected the capital costs of the small size plants. Needless to say that although costs per design capacity are increasing by the years modern technology tends to lower the costs based on the constant dollar value. Thus, for the same dollar value one can expect at least some or better efficiency from the waste treatment plants constructed in the future.

For completeness, tertiary treatment costs are included. Today, there are simply not enough plants to derive equations statistically. Curves in Figure 4 compare the costs of tertiary treatment (activated carbon + phosphate removal) with various secondary processes for common capacities. Cost figures were obtained from a report prepared

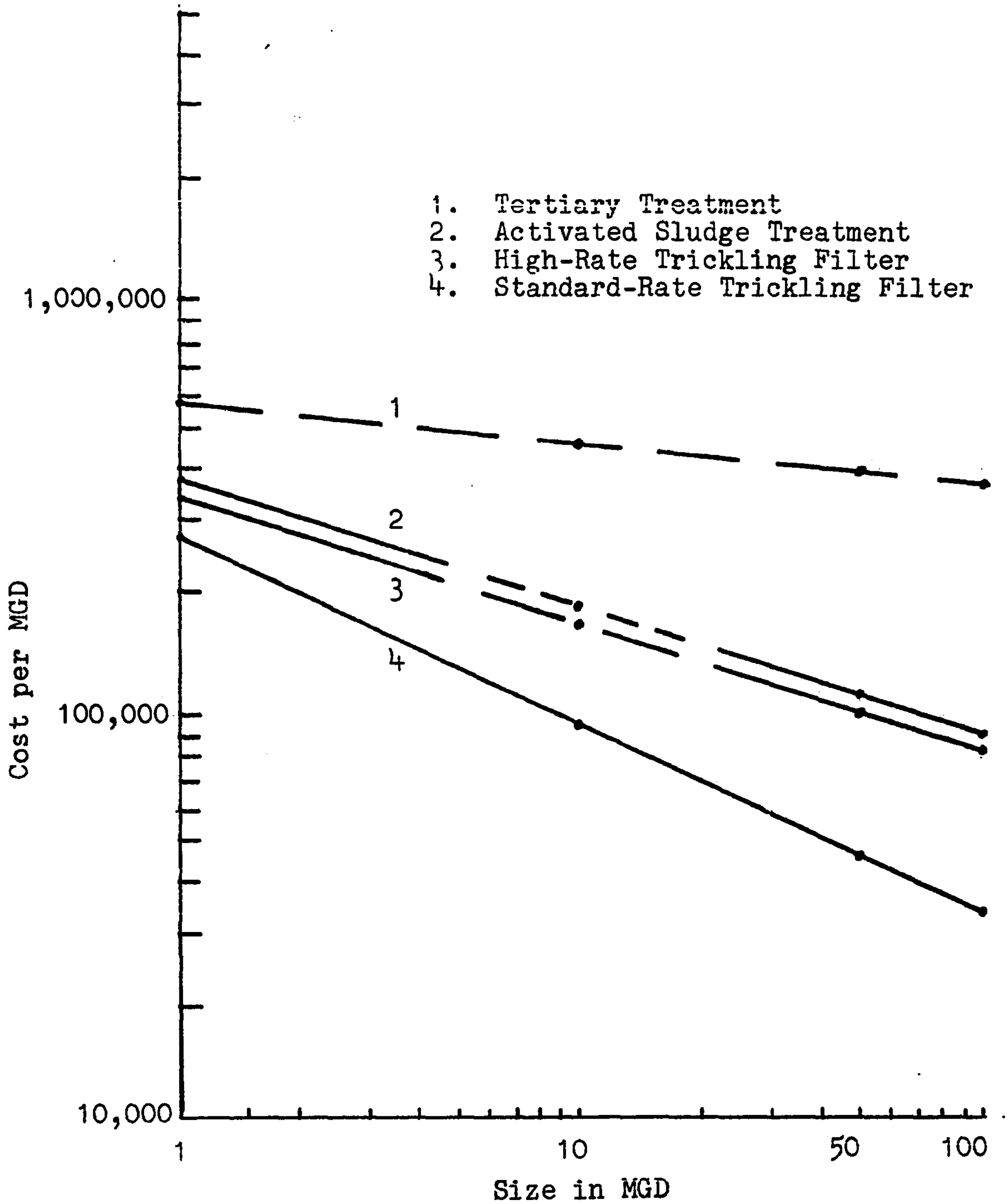

Figure 4. Construction cost of secondary and tertiary treatment plants. All data adjusted to WPC-STP Index (1957-59 = 100).

by Chou-Shong Chow, et al. (40). The secondary treatment cost figures are estimates of expected costs computed from the equations developed in this study. One can thus get some idea of the additional costs involved in the tertiary treatment.

Thus the use of the regression equations presented in this study will give reliable estimates of the cost of sewage treatment plant to those concerned with water resources planning. Use of the estimates thus obtained in present, or new, mathematical models for the operation of a water resources region will greatly aid in deciding the proper balance between dilution of wastes and degree of treatment wastes and will also aid in the decision of overall operation. That is, in any given region it may be possible that provision of storage may be more economical than treating the wastes beyond a certain degree, say for removing over 90 per cent of influent BOD. For certain other regions the contrary may be true. In either case, this study does provide information, in the form of cost estimation, to help the planner of a water resources region in making more rational decisions in the operation of a region. Tables 12 through 16 present the regression equations and the necessary parameters for computing the expected unit construction cost estimates and the confidence limits. Appendix E presents estimates of mean unit construction costs of a few selected conditions.

REFERENCES

1. Select Committee on National Water Resources, United States Senate, Water Quality Management. Committee Print No. 24, Govt. Printing Office, 1960.

2. Bureau of the Census, Current Population Reports. (Series P-25, No. 187).

3. Bureau of the Census, "Population Projections and Economic Assumptions." Committee Print No. 5, Water Resources Activities in the United States, U.S. Senate Select Committee on National Water Resources, March 1960.

4. United States Public Health Service, "Future Waste Requirements for Municipal Use." Committee Print No. 7, Water Resources Activities in the United States, United States Senate Select Committee on National Water Resources, January 1960.

5. Reid, George W., "Water Requirements for Pollution Abatement." Committee Print No. 29, Water Resources Activities in the United States, United States Senate Committee on National Water Resources, July 1960.

6. United States Public Health Service, "Pollution Abatement." Committee Print No. 9, Water Resources Activities in the United States, United States Senate Select Committee on National Water Resources, Jan. 1960.

7. Federal Water Pollution on Control Administration, U.S. Department of the Interior, "Sewer and Sewage Treatment Plant Construction Cost Index." Series CWT-1, Division of Construction and Grants, Washington, D.C., Dec. 1967.

8. United States Public Health Service, "Modern Sewage Treatment Plants--How Much Do They Cost?". PHS Publication No. 1229. U.S. Department of Health, Education, and Welfare.

9. Schroepfer, G. J., "Economics of Sewage Treatment." Transactions, American Society of Civil Engineers, Vol. 104, No. 2036 (1939), p. 1210.

10. Velz, C. J., "How Much Should Sewage Treatment Cost?" Engineering News Record, Vol. 141, No. 16 (Oct. 1948), p. 84.

11. Morgan, P. E. and Baumann, E. R., "Recent Sewage Treatment Plant Costs in Iowa." Sewage and Industrial Wastes, Vol. 27, No. 3 (March 1955), p. 257.

12. Diachishin, A. N., "New Guide to Sewage Plant Costs." Engineering News Record, Vol. 159, No. 15 (Oct. 1957), p. 316.

13. Mau, G. E., "Sewage Treatment Plant Costs in Kansas." Sewage and Industrial Wastes, Vol. 30, No. 12 (Dec. 1958), p. 1443.

14. Howells, D. H. and Dubois, D. P., "Design Practices and Costs for Small Secondary Sewage Treatment Plants in the Upper Midwest." Sewage and Industrial Wastes, Vol. 30, No. 11 (Nov. 1958), p. 1327.

15. Thoman, J. R. and Jenkins, K. H., "How to Estimate Sewage Plants Quickly." Engineering News Record, Vol. 161, No. 26 (Dec. 1958), p. 64.

16. Rowan, P. P., Jenkins, K. H. and Butler, D. W., "Sewage Treatment Construction Costs." Journal Water Pollution Control Federation, Vol. 32, No. 6 (June 1960), pp. 594-604.

17. Logan, J. A., Hatfield, W. D., Russell, G. S. and Lynn, W. R., "An Analysis of the Economics of Waste-Water Treatment." Journal Water Pollution Control Federation, Vol. 34, No. 9 (Sept. 1962), pp. 860-882.

18. Wollman, N., "Cost of Treatment." University of New Mexico, New Mexico, 1963.

19. Park, W., "Make Quick Sewage Plant Construction Estimates." Engineering News Record, Vol. 168, No. 12 (March 1962), p. 132.

20. Assenzo, Joseph R., "Use of Multiple Regression Techniques for Estimating Municipal Sewage Treatment Costs." Ph.D. Dissertation, The University of Oklahoma, 1963.

21. Smith, Robert, "Preliminary Design and Simulation of Conventional Wastewater Renovation Systems Using the Digital Computer." Water Pollution Control Research Series Publication No. WP-20-9, U.S. Department of Interior, FWPCA (March 1968).

22. Eckenfelder, W. W., Jr., "Effluent Quality and Treatment Economics for Industrial Wastewaters." U.S. Department of Interior, FWPCA (Oct. 1967).

23. Johnston, J., Econometric Methods. McGraw Hill Co., Inc., New York, 1960, p. 207.

24. Draper, N. R., and Smith, H., Applied Regression Analysis. John Wiley and Sons, Inc., New York, 1966, pp. 171-172.

25. Johnston, J., Econometric Methods. McGraw-Hill Co., Inc., New York, 1960, pp. 221-228.

26. Draper, N. R., and Smith, H., Applied Regression Analysis. John Wiley & Sons, Inc., New York, 1960, pp. 134-142.

27. Cochran, William G., Sampling Techniques. John Wiley and Sons, Inc., New York, 1963, pp. 5-8.

28. Yamane, Taro, Elementary Sampling Theory. Prentice-Hall, Inc., New Jersey, 1967, p. 3.

29. Federal Water Pollution Control Administration, U.S. Dept. of Interior, Private Communication, Dec. 24, 1968.

30. Yamane, Taro, Elementary Sampling Theory. Prentice-Hall, Inc., New Jersey, 1967, pp. 136-141.

31. Neville, A. M., and Kennedy, J. B., Basic Statistical Methods. International Text Book Co., Pa., 1964, p. 261.

32. Johnston, J., Econometric Methods. McGraw-Hill Co., New York, 1960, pp. 106-126.

33. Draper, N. R., and Smith, H., Applied Regression Analysis. John Wiley & Sons, New York, 1960, pp. 63-64.

34. Draper, N. R., and Smith H., Applied Regression Analysis. John Wiley & Sons, New York, 1960, pp. 86-97.

35. Johnston, J., Econometric Methods. McGraw-Hill Co., Inc., New York, 1960, p. 192.

36. Johnston, J., Econometric Methods. McGraw-Hill Co., Inc., New York, 1960, p. 179.

37. Oklahoma State Department of Health, Private Communication, March 22, 1969.

38. Reid, George W., Unpublished Personal Notes, June 1969.

39. Wells, D. M., and Gloyna, E. F., "Management of Return Flows in Texas," U.S. Public Health Service, Water Supply and Pollution Control Division, May 31, 1966.

40. Chou-Shong Chow, Eckenfelder, W. W., Jr., Malina, J. F., Jr., and Ford, D. L., "A Summary Report on Costs of Municipal Wastewater Treatment and Estimation of Industrial Wastewater Characteristics and Costs of Treatment." The University of Texas, May 1967.

APPENDIXES

APPENDIX A

INSTRUCTIONS FOR FILLING IN THE QUESTIONNAIRE

1. Please return the completed questionnaire by March 22, 1969.

2. LOCATION: The town, city, or Sanitary district, etc.

3. YEAR OF CONTRACT AWARD: Indicate the year of contract award, if available. Otherwise, give the year construction was completed and indicate by (*) in latter case.

4. TYPE OF PLANTS:

   (a) Lo: Stabilization Ponds, designed mainly for aerobic treatment of raw sewage. No distinction is made as to the number of cells provided or flow pattern.

   (b) Prim: Primary treatment: Employs gravity settling and separate sludge digestion (exclude Imhoff tanks).

   (c) A.S. Activated sludge plants: Projects which employ primary settling, aeration by either diffused air or mechanical means, and final settling.

   (d) H.R.T.F. High rate trickling filter.

   (e) S.R.T.F. Slow rate trickling filter.

   Both (d) and (e) include primary treatment, sludge digestion, and final clarification.

5. DESIGNED FOR: Give population, 5 day BOD in milligrams per liter (MG/L) flow in million gallons daily (MGD.), for which the treatment plant has been designed for.

6. CURRENT DATA: Give current population served, flow, and influent and effluent BOD from the plant records (average figures).

Check if a copy of the report is desired.

				Designed for			Current Data				
Location	Year of Contract Award	Cost	Type of Plant	Population	BOD_5 MG/L	Flow MGD.	Population Served	Flow MGD.	Influend BOD_5 MG/L	Effluent BOD_5 MG/L	Remarks

APPENDIX B

TABLE 19

MUNICIPAL WASTE TREATMENT PLANT COST INDEXES

	1930	1956	1957	1958	1959	1960
Atlanta	29.37	86.48	92.73	96.05	96.53	97.71
Baltimore	34.02	89.30	96.52	100.08	102.46	103.29
Birmingham	32.37	85.08	90.68	93.75	95.82	98.08
Boston	34.33	94.76	100.41	103.36	105.84	106.88
Chicago	37.26	95.92	102.10	104.50	107.39	107.51
Cincinnati	36.00	92.17	98.74	102.15	102.59	104.73
Cleveland	39.06	99.18	105.30	107.04	107.13	108.71
Dallas	35.24	84.53	91.11	93.64	95.74	96.64
Denver	37.51	86.15	91.32	94.10	98.13	99.43
Detroit	31.86	97.66	102.34	105.65	108.99	109.88
Kansas City	32.45	90.08	94.54	97.70	100.43	101.72
Los Angeles	30.88	92.45	98.45	102.61	105.66	107.86
Minneapolis	30.97	94.27	99.84	104.50	107.71	108.80
New Orleans	31.16	85.17	90.15	93.43	95.20	96.09
New York	41.30	99.05	106.31	110.44	114.78	116.19
Philadelphia	33.93	95.13	100.39	105.76	103.85	104.52
Pittsburgh	35.18	91.49	100.95	104.06	107.30	108.20
St. Louis	36.56	92.02	97.93	102.07	104.92	105.54
San Francisco	37.17	92.41	100.01	103.47	104.68	108.08
Seattle	31.92	94.12	100.88	105.57	107.86	109.29
National averages (1957-59 = 100)	34.43	91.87	98.04	101.50	103.65	104.96
National averages (1930 = 100)	100.00	266.79	284.71	294.75	301.00	304.89

Source: Abridged from Sewer and Sewage Treatment Plant Construction Cost Index, Federal Water Pollution Control Administration, Division of Construction Grants, CWT-1, December, 1967.

TABLE 19--Continued

Years						
1961	1962	1963	1964	1965	1966	1967
98.68	100.31	100.92	103.63	103.23	105.88	110.52
103.57	105.19	104.98	104.54	104.95	113.15	115.32
97.25	96.86	98.24	99.70	100.62	104.17	104.08
107.86	108.27	111.11	112.48	112.91	116.97	120.00
108.94	109.41	110.83	113.12	115.63	120.31	123.43
107.04	107.54	108.28	110.86	112.36	115.50	119.08
109.71	110.62	110.43	113.28	116.17	120.25	125.26
96.39	97.23	98.56	100.12	100.92	105.19	106.10
99.43	101.16	102.21	104.90	106.07	109.39	111.63
110.73	111.70	112.79	114.57	119.56	122.56	128.62
102.77	103.49	104.35	105.87	107.19	111.69	112.77
108.53	110.66	111.72	114.10	116.64	124.69	128.03
109.93	110.10	111.67	113.63	115.62	118.22	122.17
97.41	98.88	99.76	100.29	102.67	106.82	109.59
118.39	119.48	121.46	127.03	131.41	133.94	139.02
105.62	107.82	109.14	111.08	112.77	116.89	119.07
107.81	110.84	112.39	114.13	115.84	119.41	121.35
106.28	108.29	109.92	112.58	114.70	120.65	125.84
108.65	109.51	116.50	118.28	123.18	128.24	133.24
111.63	112.49	115.08	116.58	118.99	124.50	130.51
105.83	106.99	108.52	110.54	112.57	116.92	120.28
307.33	310.70	315.14	321.01	326.91	339.54	349.29

APPENDIX C

TABLE 20

RESULTS OF THE DURBIN-WATSON TEST FOR THE CORRELATION OF THE RESIDUALS FOR THE SELECTED EQUATIONS

Type of Treatment	d	d_L	d_u
Municipal Waste Treatment Plants			
Activated Sludge	1.90	1.66	1.75
High-Rate Trickling Filters	1.71	1.68	1.77
Standard-Rate Trickling Filters	1.98	1.54	1.66
Stabilization Ponds	1.06*	1.74	1.83
Primary Plants	1.68	1.63	1.72
Industrial Waste Treatment Plants			
Secondary Plants	1.70	1.08	1.53
Primary Plants	1.95	0.95	1.54

*Evidence of correlation.

TABLE 21

RESULTS OF THE DURBIN-WATSON TEST FOR THE CORRELATION OF THE RESIDUALS OF THE SELECTED EQUATIONS USING DUMMY VARIABLES

Type of Treatment	d	d_L	d_u
Municipal Waste Treatment Plants			
Secondary Plants	1.78	values are not available	
Industrial Waste Treatment Plants			
Secondary Plants	1.56	1.06	1.76
Primary Plants	1.77	1.04	1.77

APPENDIX D

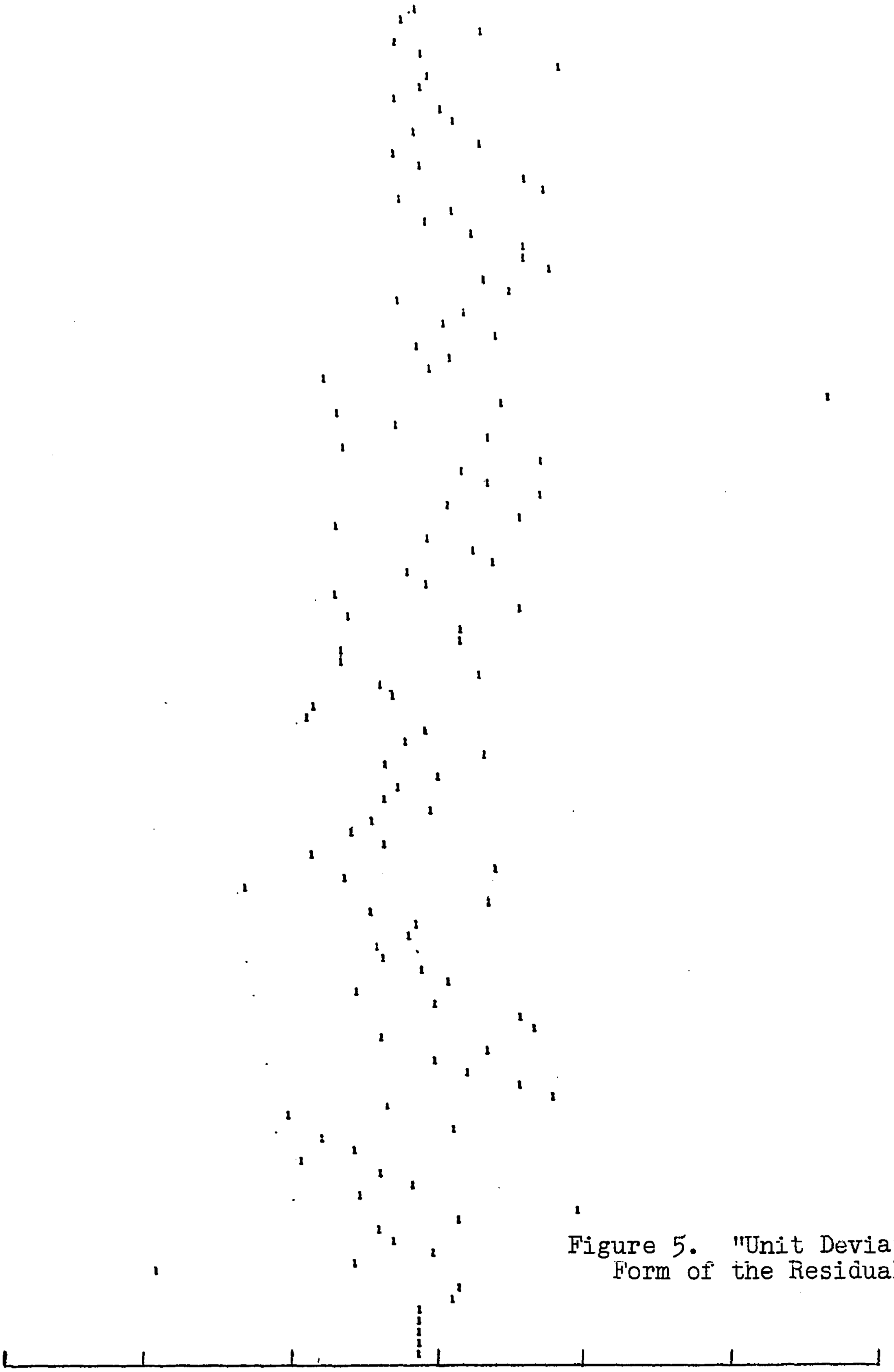

Figure 5. "Unit Deviate" Form of the Residuals.

APPENDIX E

TABLE 22

ESTIMATED MEAN CONSTRUCTION COSTS PER DESIGN PE FOR SELECTED CAPACITIES

Type of Plant	PE Designed	Flow in MGD Designed	Estimate of Mean Cost
Standard-Rate Trickling Filter	100	0.01	$ 222.91
	1,000	0.1	77.98
	10,000	1.0	27.28
	100,000	10.0	9.54
	500,000	50.0	4.58
	1,000,000	100.0	3.33
High-Rate Trickling Filter	100	0.01	120.14
	1,000	0.1	61.81
	10,000	1.0	31.80
	100,000	10.0	16.36
	500,000	50.0	10.28
	1,000,000	100.0	8.41
Activated Sludge	100	0.01	127.62
	1,000	0.1	67.52
	10,000	1.0	35.72
	100,000	10.0	18.89
	500,000	50.0	12.11
	1,000,000	100.0	9.99
All Secondary Treatment Plants Combined*	100	0.01	146.93
	1,000	0.1	71.27
	10,000	1.0	34.57
	100,000	10.0	16.76
	500,000	50.0	10.11
	1,000,000	100.0	8.13

*Using dummy variables.

TABLE 23

COST PER MGD

Type of Treatment Plant	PE Designed	Flow in MGD Designed	Estimate of Mean Cost
Primary	100	0.01	$1,300,000
	1,000	0.1	580,165
	10,000	1.0	258,916
Stabilization Ponds	100	0.01	568,389
	1,000	0.1	220,351
	10,000	1.0	85,425
	100,000	10.0	33,117
	500,000	50.0	17,076
	1,000,000	100.0	12,839
Standard-Rate Trickling Filter	100	0.01	2,229,110
	1,000	0.1	779,888
	10,000	1.0	272,855
	100,000	10.0	95,462
	500,000	50.0	45,818
	1,000,000	100.0	33,399
High-Rate Trickling Filter	100	0.01	1,201,480
	1,000	0.1	618,184
	10,000	1.0	318,067
	100,000	10.0	163,650
	500,000	50.0	102,847
	1,000,000	100.0	84,201
Activated Sludge	100	0.01	1,276,370
	1,000	0.1	675,269
	10,000	1.0	357,254
	100,000	10.0	189,006
	500,000	50.0	121,119
	1,000,000	100.0	99,994