

University of Montana

ScholarWorks at University of Montana

Graduate Student Theses, Dissertations, &
Professional Papers

Graduate School

1993

Box-Cox transformations of a travel-cost demand function for stream fishing in Montana

Joseph M. Holliman
The University of Montana

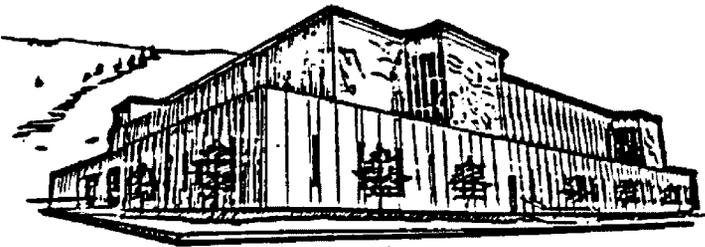
Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Let us know how access to this document benefits you.

Recommended Citation

Holliman, Joseph M., "Box-Cox transformations of a travel-cost demand function for stream fishing in Montana" (1993). *Graduate Student Theses, Dissertations, & Professional Papers*. 8770.
<https://scholarworks.umt.edu/etd/8770>

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.



Maureen and Mike MANSFIELD LIBRARY

Copying allowed as provided under provisions
of the Fair Use Section of the U.S.

COPYRIGHT LAW, 1976.

Any copying for commercial purposes
or financial gain may be undertaken only
with the author's written consent.

University of
Montana

BOX-COX TRANSFORMATIONS OF A TRAVEL-COST DEMAND FUNCTION
FOR STREAM FISHING IN MONTANA

By

Joseph M. Holliman

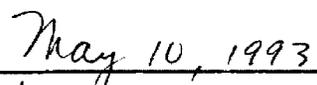
B. A., B. S. University of Montana, 1985

Presented in partial fulfillment of the requirements
for the degree of
Master of Arts
University of Montana
1993

Approved by


Chairman, Board of Examiners


Dean, Graduate School


Date

UMI Number: EP39571

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI EP39571

Published by ProQuest LLC (2013). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

BOX-COX TRANSFORMATIONS OF A TRAVEL-COST DEMAND FUNCTION
FOR STREAM FISHING IN MONTANA (128 pages)

Director: John W. Duffield



The zonal-aggregate, regional travel-cost model (TCM) has become a widely accepted method to estimate the demand and net economic value for access to public lands for participation in outdoor recreation activities. In its traditional form, per capita trips are modeled as a function of the variable travel and time costs associated with traveling to and from a recreation site. The model is also used to evaluate changes in net economic values for public resources in cases when additional resources are made available to the public or when resources are subjected to changes in site attributes. The methods used to estimate net economic values and demand in such cases requires a well specified model resulting in accurate trip prediction across the sample.

Several alternative mathematical forms of estimated travel-cost demand functions have been investigated in attempts to enhance trip prediction and reliability of net economic value estimates. However, no definitive conclusions regarding the appropriate functional form of the model have been reached. Using two prior TCM studies on data collected in 1985 by the Montana Department of Fish, Wildlife and Parks for the Montana cold-water stream fisheries as a bench-mark, this study examines alternative Box-Cox transformations of the basic bivariate demand function.

A comprehensive search of plausible functional forms suggests a model in which the natural log of per capita trips regressed on average round-trip distance raised to the .172 power maximized the log-likelihood function with significant values of all estimated parameters. This model was then discriminated from a previously proposed alternative form of the double-log model in which average round-trip distance was shifted outward at all observations by a constant. The J-test and adjusted- R^2 revealed inconclusive results as to which model was most appropriate. It was concluded that the two models provide alternative means of describing the variation in per capita trip demand. However, a previously suggested theory supporting the double-log model appears to provide greater support for its use than the model in which distance is raised to the .172 power.

TABLE OF CONTENTS

ABSTRACT	ii
LIST OF TABLES	vi
LIST OF ILLUSTRATIONS	vii
ACKNOWLEDGMENTS	ix
CHAPTER	
1. STATEMENT OF THE PROBLEM	1
Organization	3
The Travel Cost Model	3
Simplifying Assumptions	5
Quantity Measures	5
Price	7
Data Requirements	8
Examples of TCM Applications	9
Statistical Determination of the Form of the First-Stage Demand Function	11
Study Methodology	13
Study Outline	17
2. LITERATURE REVIEW OF MODEL SPECIFICATION AND THE BOX-COX TRANSFORMATION	19
General Elements of Model Specification For OLS Estimated Models	20
Approaches to Model Specification	22
The Box-Cox Family of Power Transformations	26
Purpose and Assumptions of the Basic and Extended Box-Cox Regression	28
Methods Used to Estimate a Box-Cox Regressions	30
Methods of Model discrimination	38
The Likelihood Ratio Test	38
Tests for Non-nested models	40
Diagnostic Tests Useful for Model Specification	41
Methods for Testing Normality of Residuals	41
Detection of Heteroscedasticity	41
Methods for Detecting and Measuring the Influence of Outliers	43
Conclusions and Summary	44

3.	DATA SOURCES AND DESCRIPTION	47
	Description of the Sites Comprising the Montana Stream Fisheries	47
	Data Sources Included in the Cold-Water Streams Fisheries Data Base	51
	Data Preparation for First-Stage TCM Demand Function Estimation	51
	Collection of Survey Data	51
	Aggregation of the Supplemented Fisheries Survey	54
4.	ESTIMATION AND ANALYSIS OF DEMAND FUNCTIONS	58
	Bivariate Estimates of Two Previously Estimated Demand Functions	60
	Diagnostic Tests on Models 1 and 2	64
	General Model Specifications for the Box-Cox Regression	72
	Form Analysis of the Bivariate First-Stage TCM Demand Function	75
	Estimated Models	75
	Model Discrimination	78
	Diagnostic Tests on Model 5	80
	Discrimination of Models 2 and 5	83
	Overview of Models 1, 2, and 5	87
	Scatter and Line Plots	87
	Predictive Power	93
	Elasticities	93
5.	CONCLUSIONS	95
	Summary	95
	Limitations and Suggestions For Further Research	98
	Heteroscedasticity	99
	Truncated Distribution of the Residuals in Model 5	99
	Errors in Measurement of the Distance Variable	100
	Omitted Variable Bias	101
	Market Segmentation	102
	LIST OF REFERENCES	104

APPENDIX

A. TECHNICAL DESCRIPTION OF THE TRAVEL COST MODEL . 109

B. TECHNICAL DESCRIPTION OF THE J-TEST 111

C. TECHNICAL DESCRIPTION OF DIAGNOSTIC TESTS 114

 White's Test for Homoscedasticity 114

 DFFITS and DFBETAS Measures 115

 Modeling the Effects of Outliers Using
 Dummy Variables 119

 Analysis of Modeling Trips from Alaska
 as a Dummy Variable in Models 1 and 2 120

 Analysis of Modeling Trips from Alaska
 as a Dummy Variable in Model 5 122

D. SURVEY FORMS 125

LIST OF TABLES

TABLE

1.	Montana Stream Fisheries: Non-Unique Waters	49
2.	Montana Stream Fisheries: Unique Waters	50
3.	Descriptive Statistics for the Variables Specified in the Models Presented in Chapter 4	57
4.	Bivariate Specifications of Models Specified by Duffield et al (1987) and Duffield (1988)	63
5.	Plausible Forms of Equation (2) Other than the Double-log, Semi-log and Linear forms	73
6.	Estimated Bivariate First-Stage TCM Demand Functions	77
7.	Likelihood Ratio Tests of Bivariate Models: with Model 3.a. as the General, Unrestricted Case	79
8.	Results of J-Test for Discrimination Between Models 2 and 5	84
9.	Estimated Own-Price Elasticities of Demand for Models 1, 2, and 5	94
10.	Possible Outcomes of The J-Test	113
11.	Estimates of Models 1 and 2 with Alaska Trips Modeled as an Interaction Parameter Dummy Variable	121
12.	Estimates of Model 5 with Alaska Trips Modeled as an Interaction Parameter Dummy Variable with Round-Trip Distance	124

LIST OF ILLUSTRATIONS

FIGURE

1.	Normal Probability (P-P) Plot of Standardized Residuals for Model 1	67
2.	Normal Probability (P-P) Plot of Standardized Residuals for Model 2	67
3.	Scatterplot of the Residuals and Predicted Natural-Log of Per Capita Trips Produced by Model 1	68
4.	Scatterplot of the Residuals and Predicted Natural-Log of Per Capita Trips Produced by Model 2	68
5.	Scatterplot of the Residuals Produced by Model 1 and the Natural-log of Average Round-trip Distance	69
6.	Scatterplot of the Residuals Produced by Model 2 and the Natural-log of Shifted Average Round-trip Distance	69
7.	Normal Probability (P-P) Plot of the Standardized Residuals for Model 5, Table 6	81
8.	Standardized Scatterplot of predicted Per Capita Trips and Residuals for Model 5, Table 6	81
9.	Standardized Scatterplot of Round-Trip Distance Transformed by the Box-Cox Parameter listed for Model 5, Table 6 with the residuals from the same model	83
10.	Scatterplot of the natural-log of per-capita trips with the natural-log of average round-trip distance	89
11.	Scatterplot of the natural-log of per-capita trips with the natural-log of average round-trip distance plus 64	90

12.	Scatterplot of the natural-log of per-capita trips with average round-trip distance raised to the .172 power	91
-----	--	----

ACKNOWLEDGEMENTS

Primary recognition goes to John W. Duffield for his practical instruction on different aspects of measuring values of outdoor recreation during the process of performing the Montana Bioeconomic studies. His comments toward formulation of this project and feedback on certain issues served most useful in limiting the scope of the project. I also recognize Dr. Duffield's continued support toward completion of this paper.

I also thank the balance of the review committee including Michael H. Kupilik, Douglas R. Dalenberg, and David A. Patterson. Their devotion to the arduous task of reviewing and commenting on key topics and the drafts of this paper have proven most valuable. Recognition also goes to Mark D. Partridge for feedback on econometric issues, Christopher J. Neher for assistance with the data used in this study and information regarding previous analysis of the data base, Becky Hanway for assistance with *WORDPERFECT*, and Denise Peterson for editing.

Finally, I thank my parents, James D. and Mary Helen Holliman and the faculty of the Department of Economics for their support and belief in my abilities to succeed. Particular recognition goes to Dr. Kay C. Unger for her continual inspiration and instruction of economic theory.

CHAPTER 1
STATEMENT OF THE PROBLEM

The purpose of this paper is to examine the functional form of a statistically estimated (first-stage) travel-cost model (TCM) demand function for cold-water stream fishing in Montana during 1985. The TCM is a widely accepted approach to estimating the demand for and net economic value of non-market resources. Applications of the TCM approach to resource valuation problems range from water and air quality to several consumptive use recreation activities.¹ However, the most common applications pertain to specific recreation activities in which natural resource sites are provided by governmental entities as public goods (see, e.g., McConnell (1985)). Since such activities are generally not sold in a market specifically as "recreation packages", no directly observable market prices are available from which economic values can be derived. Thus, TCM exploits the ability to observe actual expenditures associated with outdoor recreation trips which serve as the

¹ Walsh, et. al. (1988) identifies several activities such as camping and swimming (e.g., Sutherland, 1980)), picnicking, hiking, and hunting (e.g., Martin et. al., 1974 (1981)), fishing (e.g., Sorg et. al., 1985), and wilderness use (e.g., Smith and Kopp, 1980).

price proxy for the demand function.

Economic theory provides little guidance on the mathematical form of a demand function (see e.g., Koutsoyiannis (1977) and Russell and Wilkinson (1979)). Accordingly, there is no *a priori* theoretical basis for specifying the form of the first-stage TCM demand function other than recognition of the inverse relationship between price and quantity.² Thus, researchers have identified the appropriate functional form using statistical inference (see, e.g., Smith (1975), Zeimer, Musser, and Hill (1980), and McConnell (1985)). The central topic of this paper is the functional form of the first-stage TCM demand function for fishing on Montana's rivers and streams during 1985. The approach uses statistical inference to specify the form of this function. This study is a reexamination of the original and one of two subsequent analyses of the zonal aggregated data collected by the Montana Department of Fish Wildlife and Parks (DFWP) performed by Duffield, Loomis, and Brooks (1987) and Duffield (1988). Duffield's (1988 and 1992) previous analysis of the stream fisheries data base has focused on the form of the demand function with respect to the price variable. The analysis in this study is limited to the same problem.

² As discussed below, one method of estimating the economic value of access to a recreation site using TCM is based on statistically estimating a first-stage demand function. This function is then used to derive a second-stage function from which consumer surplus is computed.

Organization

The balance of this chapter is organized as follows. First, a general and intuitive description of the zonal travel-cost model is provided.³ This is followed by a review of the literature in which statistical determination of the form of the first-stage demand function has been estimated. This section describes the methods used in this paper to determine the functional form of the bivariate first-stage demand function. Included in this section is a brief overview of the findings of previous TCM demand estimation using the same data base used in this study. Finally, an outline of the remaining chapters of the paper is presented.

The Travel-Cost Model

A wide variety of travel-cost models have been developed and are classified by Ward and Duffield (1992) as conventional travel-cost models, random utility models and hedonic models. However, the analysis in this paper focuses on the single equation zonal aggregate TCM which falls under the general classification of traditional travel-cost models. In this model, recreation demand or participation (per capita visits) for a site or region of sites, aggregated by origin zone, is statistically estimated as a

³ Readers are referred to Appendix A for a more technical description of the TCM.

function of the price for each trip. Average variable travel expenditures incurred to make a trip and foregone time (opportunity) costs associated with each trip for a specific activity serve as a proxy for price. Variables in the demand function other than price typically include income, price of substitutes, socio-economic and/or demographic variables and site attributes. From this function a second-stage demand relation can be derived on the assumption that recreationists would respond to site access prices in the same way as they respond to varying travel costs. The area under the simulated demand function above the costs actually incurred to make a trip is defined as the net economic value surveyed users hold for maintaining access to the recreation sites under study.⁴ This method requires the model to accurately predict trips across the sample. The first-stage demand function can be integrated to determine net economic value (Menz and Wilton (1983) in Duffield *et al.* (1987)). The resulting values represent Marshallian consumer surplus which serve as a reasonable proxy for net amount recreationists would be willing to pay to maintain access to a site⁵ (see, Just *et*

⁴ See Dwyer, Kelly, and Bowes (1977) for an explanation of the mechanics of the travel cost model.

⁵ Although TCM can be used to measure net economic values for environmental concerns other than to maintain access to a recreation site, the focus of this study is to limit the demand study to the access issue. It should also be noted that demand and economic values estimated with the zonal TCM pertain to consumptive use value. Other economic

al. (1982)). Each facet of this process is summarized below.

Simplifying Assumptions. To simplify the zonal aggregate travel cost model, economic and demographic characteristics of recreationists are assumed to be homogeneous across and representative of each origin zone's population. Furthermore, only data for single-destination and single purpose trips are included in the data set. This mitigates problems with allocating costs among destinations and activities. The amount of time recreationists spend at each site is also assumed to be homogeneous for all recreationists in each origin zone. It is also assumed that the opportunity cost of travel time is constant or at least homogeneous for all visitors to the observed site or sites within the study region.

Quantity Measures. Generally quantity has been specified by two different measures: trips taken and user-hours at the site. User-days, a linear transformation of user days, has also been used (McConnell (1975)). However, McConnell (1975) argues, that the marginal cost of user-days is independent of travel costs. Since the utility maximization framework underlying the travel cost method is

values placed on an environmental resource include option, bequest, and existence values. These values pertain to the option to use a resource sometime in the future, the value associated with providing the resources as an endowment to future generations, and the value of knowing that the resource exists, respectively.

based on the change in utility subject to a change in quantity, he argues that the demand for user-days does not fit in this framework. That is, the marginal cost of a user-day is independent of travel costs once a trip is made. He also argues that consumers' surplus computations require knowledge of the relationship of units costs and quantity demanded. This is consistent with the theory of TCM in a household production framework in which the demand function is similar to a derived demand for trips. Specifically, Ward and Duffield (1992) note that the basis for the travel cost model in this framework is the weak complementarity between marketed commodities associated with travel to and from a site, among other inputs, and the non-marketed services provided by the site. Thus, travel is an input into the production function for the provision of recreation services.

Walsh (1985) notes that a precise definition of a user-day, in terms of person hours, can be problematic when different recreationists spend disparate amounts of time participating in the intended purpose of the trip. He suggests that when the length of stay is similar for all recreationists, per capita trips is a suitable measure of quantity.

Per capita trips is the quantity measure adopted in this study. The decision to use this definition is based primarily on maintaining consistency with the two previous

demand studies using the Montana stream fisheries data base, noted above. Based on the above literature review per capita trips also appears to be the most logical measurement of quantity.

Price. Since the market for outdoor recreation trips is non-commercial in nature, price is proxied by out-of-pocket travel and visitation expenditures (short-run marginal costs) and the opportunity cost of time en route to and at the site. Price in this model represents the rate of exchange (or the marginal cost) for producing the recreation trip, on average within the sample used for each study. This proxy for price is advantageous since it allows price variation across origin-destination pairs to be observed cross-sectionally. Furthermore, there is usually more variation in this surrogate price measure than price variation generated in commercial markets (Burt and Brewer (1971)).

At the minimum, travel costs include the variable costs of operating a vehicle (see, e.g., WRC (1983) and Burt and Brewer (1971)). This information can be obtained from Department of Transportation statistics published annually. However, it has been suggested that reported costs or those costs perceived by the surveyed recreationists as the out-of-pocket costs associated with taking the trip more accurately represent true travel costs (Duffield et al. (1987)).

The price proxy also includes the opportunity cost of time in transit to and from the site. While no definitive way of measuring this cost has been developed, two methods are widely accepted. These methods vary between determining the recreationists' willingness to pay to reduce the travel time by one-half (see e.g. Duffield et al. (1987)) to determining the value based on a ratio of the estimated travel cost and combined average family income and travel time (McConnell and Strand (1981)). One method adopted by the Water Resource Council (WRC (1983)) and based on analysis by Cesario (1976) measures the opportunity cost of time as one-third the wage rate for adults. One-fourth of this value would represent the opportunity cost of time for children.

The data used by Duffield et al. (1987) for the travel and time cost measures was gathered in a separate survey supplementing the Montana Statewide Angler Pressure Mail Survey (McFarland (1989)), the primary data source used in this study. Trip-weighted, average round-trip distance is specified as the price proxy in their first-stage demand function. The same practice is used in this study.

Data Requirements. The traditional (zonal) Hottelling-Clawson-Knetch TCM uses survey data aggregated by origin zone for a single site or a collection of sites for

similar activities within a region.⁶ Data are gathered from users whose primary purpose for taking their trip to a particular site was to participate in a single consumptive recreational activity, such as hunting or fishing. Although the TCM has been described as measuring the demand and benefits associated with an entire recreation trip, there is general agreement in the literature that the model should be limited to measuring such demand and benefits for the site and single purpose trips as noted above (Dwyer *et al.* (1977)). Even though a method has been suggested to rationally allocate travel costs to several closely clustered destinations (Haspel and Johnson (1982)), the general consensus is that it is too difficult to consistently allocate the costs for trips taken for several purposes and destinations. Thus, only data for trips made with the sole intent to visit a single site for the purpose of participating in the activity under study (in this case fishing) are included in the analysis.

Examples of TCM applications. TCM may be used to estimate net economic values associated with status quo site environmental qualities⁷ for specific recreation

⁶ Two other methods rely on observations of individual recreationists; the quantity variable in these models is the number of trips per recreationist over a season. (see, e.g., Brown, Sorhus, Chou-Yang, and Richards (1983)).

⁷ "Status quo environmental quality" is defined here as the environmental quality existing prior to the implementation of a new policy or the occurrence of an

activities. TCM may also be used to forecast site usage and changes (gains or losses) in net economic value due to several alternative policy directives affecting site usage. Such directives may affect site access prices (e.g., by introducing or changing entrance or user fees or by taxation) or result in changes to the environmental quality of a site (Ward and Loomis, 1986). Such changes may be due to projects such as damming a river or permanently changing the water level of a reservoir. Changes may also result from an involuntary oil or hazardous material spill. The Water Resources Council (1983) provides guidelines for estimating changes in site usage and economic benefits.⁸ In some cases, it may be reasonable to use previously estimated relationships of valuation and non-price variables such as site attributes.

incident which changes a sites quality. The status quo would then most appropriately refer to the base-line condition similar to the definition provided in section 11.14 of Title 43 CFR. In this case it would be the cross-sectional "snap-shot" of the site quality at the time the data were collected.

⁸ WRC recommends TCM as one of two preferred methods to evaluate changes in economic benefits and to forecast usage when assessing the impacts of water related projects. Additionally, federal regulations governing the methods used to assess economic losses due to oil or hazardous material spills adopt the WRC guidelines by reference (43 CFR Section 11.18 (a) (2)) as one means for determining economic losses associated with such accidents.

Statistical Determination of the Form of the First-Stage Demand Function

Based on the foregoing summary, it can be seen that several criteria should be met to obtain accurate estimates of consumer's surplus. Other than sound and complete data collection, demand functions in general should be fully specified according to demand theory (Koutsoyiannis (1977)). Omission of key variables such as the opportunity cost of time and substitutes in a TCM demand function will generally result in biased net economic value estimates (see, e.g., Strong (1983)). Since the magnitude of consumer's surplus relies heavily on price elasticities of demand, correct functional form plays a key role in the value of consumer's surplus estimate (see, e.g., Zeimer et al. (1980)). However, it is also important that the model be able to accurately predict trips since predicted trips is a factor in estimating consumer's surplus (see, e.g., Ward and Loomis (1986) and Duffield (1988)). This section reviews the literature regarding statistical procedures previously used to specify the form of the first-stage demand function.

A priori specification of the form of the demand function has provided less than optimal estimation of demand and consumer surplus (see, e.g., McConnell (1985)). As an example, McConnell (1975) shows that the functional form cannot be additive since price must be allowed to vary as income varies. This conclusion is based on analysis which

shows the demand slope with respect to income depends on costs and that the symmetry of the cross partial derivatives with respect to all other prices and income is not equal to zero. Therefore, a linear functional form is theoretically incorrect.⁹ The linear form has been used to estimate the first-stage TCM demand function (see, e.g. Bowes and Loomis (1980)). However, Vaughn, Russell, and Hazilla (1982) showed that Bowes and Loomis' data is more appropriately described by a semi-log form once heteroscedasticity is identified and corrected.¹⁰

While the linear form is inappropriate, theory provides little other guidance on the appropriate mathematical form of the first-stage TCM demand function. Much work has been devoted to the determination of the functional form using statistical methods. The following are three examples.

First, based on a study of general recreation usage in the Desolation Wilderness Area in Northern California, Smith (1975) suggests there does not appear to be any empirical justification favoring use of either the linear, semi-log, or the double-log (with log transformations in the

⁹ Quirk (1976) also notes that if prices and income are homogeneous to degree zero a linear functional form is not possible.

¹⁰ All semi-log models in this paper refer to a form in which the dependent variable is transformed by the natural-log and the independent variable takes its linear form.

dependent and independent variables) models. Further, using a discrimination test for non-nested models, namely Pesaran's N statistic (Pesaran (1974)), Smith found that neither the semi-log or double-log forms represent the behavioral patterns of recreation usage described by the data.¹¹

In another study of warm-water fishing in Georgia, Zeimer *et al.* (1980) found that the semi-log versus the double-log form fit their data best (as determined by applying the Box-Cox method of model discrimination). Third, based on the findings of Zeimer *et. al.*, Strong (1983) suggest that functional form is an important specification consideration when applying the TCM approach.

Study Methodology

The goal of this paper is to determine the mathematical form of the first-stage TCM demand function using the Box and Cox (1964) statistical method of discriminating alternative functional forms. As noted, this is done for the bivariate demand function. The determination of functional form is also limited by the bounds of the applicable economic demand theory:

¹¹ Although Smith (1975) notes that the TCM approach is typically used for specific recreation activities and may not be valid for the application in his 1975 paper, he argues that demand for such diverse activities will be indirectly reflected in the derived demand for the sites services. Smith's analysis was based on a single site zonal aggregate travel cost model.

specifically, the inverse relation of price and quantity.

The standard forms used to estimate first-stage outdoor recreation demand functions in the literature have been generally limited to the linear, semi-log (with a natural-log transformation of the dependent variable), double-log, and quadratic forms. Relatively little has been done in determining which of these forms is most appropriate in estimating recreational demand (see, e.g. Smith (1975); Strong (1983); and Zeimer, Musser, and Hill (1980)) or to further investigate forms other than those listed.

By applying the Box and Cox (1964) family of power transformations to the data collected for sport fishing on Montana cold-water streams, this paper examines possible functional forms for the first-stage demand function using Duffield's (1988) findings as a starting point. Transformations on the dependent and price-proxy variable are estimated using the Box-Cox family of power transformations given in equation (1):

$$(1) \quad z^{(\lambda_k)} = \begin{cases} \frac{(z^\lambda - 1)}{\lambda} & ; \lambda \neq 0 \\ \ln z & ; \lambda = 0 \end{cases}$$

Where z is either the dependent or independent variable and λ is the Box-Cox transformation parameter. The general form

of the demand function is presented in equation (2):

$$(2) \quad V_{ij}^{(\lambda_v)} = \beta_0 - \beta_1 D_{ij}^{(\lambda_D)} + e$$

Where V_{ij} is per capita trips taken from origin i to site j , D_{ij} is average round-trip distance from origin i to site j , β_0 and β_1 are estimated parameters, e is the error term, and (λ_k) are the Box-Cox transformations of their respective variables. Much flexibility in the form of the estimated demand function is afforded by the Box-Cox transformation. For instance, if $\lambda_v = \lambda_D = 1$ the function is linear. If, however, $\lambda_v = 0$ and $\lambda_D = 1$ the form becomes the semi-log form (Spitzer (1982a)).¹² The form flexibility of equation (2) is discussed in greater detail in chapter 4.

As noted, this study is a reexamination of two of three prior studies on the stream fisheries data base using different specifications of the first-stage TCM demand function. The results of these studies are summarized here and discussed in more detail in chapter 4. In the first of these studies, Duffield *et al.* (1987) specified a double-log (log-linear) model in which per capita trips were regressed on average round-trip distance, total trout catch by site, and average years of fishing experience of anglers by origin zone. Due to this model's failure to produce homoscedastic

¹² Kmenta (1986) shows that as λ_k approaches zero, the Box-Cox transformation becomes the natural-log of z .

residuals and its poor predictive power, Duffield (1988) examined several alternative forms of the first-stage demand function.¹³ In this study, it was found that a restricted form of the Box-Cox transformation suggested by Zarembka (1974) produced a model which satisfied the homoscedasticity and prediction concerns in the Duffield *et al.* (1987) model. Specifically, this model took a form similar to equation (2), except $D_{ij}^{(\lambda_D)}$ was replaced with $(D_{ij} + C)^{(\lambda_D)}$ where C is a constant added to average round-trip distance.¹⁴ λ_v and λ_D were set equal to zero to produce the double-log form, and C was varied until trip prediction was within .1 percent of observed trips. This result was achieved at $C = 90$ miles for a multivariate model.

Based on Duffield's (1988) findings, λ_v is expected to lie close to zero and λ_D should fall between zero and one. It is not expected that the value of λ_D will be close to or greater than one, given the poor performance of the general polynomial, semi-log, and double-log forms estimated

¹³ Homoscedastic residuals are those in which the variance of each residual is nearly constant across all observations. When the variance of the residuals are not constant across all observations, they are said to be heteroscedastic. The consequences of heteroscedasticity are inefficient parameter estimates and biased variance estimates yielding invalid hypothesis tests of the significance of parameters (see, e.g., Kmenta (1986)).

¹⁴ Another difference between equation (2) and the model estimated by Duffield (1988) for the entire sample was that the several other variables such as catch aggregated by site, a substitute index, and certain socio-economic and demographic variables were included in the function.

by Duffield (1988).

Since there is little theoretical basis for determining the functional form of the TCM demand equation, the form must be found by experimentation. The model judged "best" will be determined through statistical inference. As an aside, and purely for illustration, the predictive power and own price elasticities of the model determined best in this study are compared with bivariate forms of the models estimated by Duffield *et al.* (1987) and Duffield (1988).

Study Outline

The balance of this paper is presented in four chapters. Chapter two provides a literature review of the use of the Box-Cox method of estimation of nonlinear regression parameters. The first section of this chapter compares two approaches used to determine the specification of a model. This is followed by a review of the methods used to estimate the parameters of a Box-Cox regression model. This section closes with a summary of the reasons maximum likelihood estimation was chosen to estimate the Box-Cox regression of the first-stage stream fisheries demand function. The next section summarizes the methods used to discriminate nested and non-nested rival models for purposes of determining model specification by statistical inference. The last section of chapter two summarizes several diagnostic measures used in conjunction with the

model specification approach employed in this paper.

Chapter three summarizes the data used in this study and the methods used to gather this data. Descriptive statistics of the variables specified in each of the functions examined are also provided.

The results of applying the Box-Cox method of model discrimination to the DFWP streams fisheries data is provided in chapter four. Included in this chapter is a summary of models estimated in the two previous studies noted above. Next, five general Box-Cox regression models are estimated such that eight plausible forms could be attained using equation (2) as a basis. These models are then discriminated to determine which form contains the parameters which are most likely to describe the population from which the data are drawn. This model is compared with the bivariate models summarized at the outset of this chapter.

The study is summarized in chapter five. Included in this chapter is a summary of the limitations of the analysis and suggestions for future research.

CHAPTER 2
LITERATURE REVIEW OF MODEL SPECIFICATION AND
THE BOX-COX TRANSFORMATION

This chapter reviews the econometric and statistical literature regarding application of the general Box-Cox family of power transformations to model specification of functional form. First, general elements of model specification within the context of ordinary least squares (OLS) are reviewed. Second, two approaches to model specification are reviewed. Included in this review is a summary of the approach used in this study and a list of the tests used to select the most appropriate form of the Montana stream fisheries demand function. Third, the basic and extended (BCE) Box-Cox transformation and methods used to estimate a Box-Cox regression function are reviewed.¹⁵ Next, a summary of the four methods used to estimate a Box-Cox regression function and a review of available econometric software is provided. This is followed by a

¹⁵ The Box-Cox extended regression equation (attributed to Savin and White (1978) by Seaks and Layson (1983)) includes variations of equation (2) in which Box-Cox transformations are applied to independent variables in addition to the dependent variable. (See also Box and Cox (1964).)

summary of the likelihood ratio test for discriminating nested models and the j-test for discriminating non-nested models. Next, diagnostic tests used in this study are reviewed followed by a summary conclusions reached in this chapter.

General Elements of Model Specification For OLS Estimated Models

Although this study is focused on estimating the functional form of a first-stage TCM demand function using the Box-Cox transformation, the problem is couched within the more general problem of model specification. Kmenta (1986) describes specification for models using (OLS) as the use of an estimation technique which satisfies its general assumptions. Application of OLS estimators assume that the error term (residuals or disturbance term) is normally and independently distributed with a mean of zero and a constant variance ($e \sim N(0, \sigma^2)$); the covariance between any two errors is zero; each of the explanatory variables are measured without error, are nonstochastic, and no "exact" linear relationship exists between any two of these variables; and the number of observations exceeds the number of estimable coefficients in the model. Satisfaction of these assumptions results in best unbiased estimators (BUE) of the regression parameters. A BUE estimator is one which has minimum variance among all linear unbiased estimators.

An unbiased estimator is one in which the expected value of the estimator equals the true value of the parameter.

Kmenta (1986), narrows the definition of specification by noting that specification errors result from failures to include only the variables relevant to the model, to specify the "correct" mathematical form of the model, and correctly specify the way in which the error term enters the model. However, as indicated in the literature, complete model specification must include the above listed basic assumptions of OLS estimators. For instance, a non-constant error variance (heteroscedasticity) may be the result of an poorly specified function form of the regression equation (Kmenta (1986)). Also, data outliers may be the result of a misspecified model (Kennedy (1992)). As noted, Duffield, Loomis, and Brooks (1987) found that the first-stage demand function estimated for the entire stream fisheries sample exhibited heteroscedastic errors. With this *a priori* knowledge, analysis of form specification would be incomplete without tests for and possible correction of any of the basic assumptions of the OLS estimator. However, the focus of this study is limited to specification of the mathematical form of the first-stage demand function. Therefore, analysis of possible tests of and corrections for heteroscedastic errors within the Box-Cox transformation framework are left for further analysis

on the stream fisheries data base.¹⁶

Approaches to Model Specification

Kennedy (1992)¹⁷ classifies state of the art approaches to model specification as 1) "Average Economic Regression (AER)", 2) "Test, Test, Test (TTT)", and 3) "Fragility Analysis"¹⁸. Collectively, the analysis by Duffield *et al.* (1987) and Duffield's (1988) of the form of the Montana stream fisheries demand function would appear best, although not perfectly, classified as an AER approach. In this approach specific forms are determined by, "proceeding from a simple model and 'testing up' to a specific more general model" (Kennedy (1992)). The AER approach is marked by a process of applying diagnostic tests to the residuals of an estimated model known to be correct to determine if the assumptions of an OLS estimator have been met. If the *a priori* model fails to satisfy these assumptions, researchers using the AER approach turn first

¹⁶ See Gaudry and Dagenais (1979), Greene (1990), Lahiri and Egy (1981) for econometric treatment of simultaneous testing and correcting for heteroscedastic errors within the Box-Cox transformation framework. Also see Vaughn, Russell, and Hazilla (1983) for an application of the Lahiri and Egy (1981) method to a travel cost model.

¹⁷ This section relies heavily on Kennedy (1992), Chapter 5.

¹⁸ "Fragility Analysis" incorporates a Bayesian method to determine if estimated parameters of a model fall within an acceptable range. Since this study used classical statistical methods, a review and comparison of this approach to the AER and TTT approaches is omitted.

to more sophisticated estimation techniques to resolve errors in the residuals. Failure of such methods to satisfy OLS assumptions then leads to tests of specification in which high R^2 and significant t-ratios are used to select the best model.

In contrast, researchers using the TTT approach begin with a model more general than that believed correct. The general model is then subjected to several diagnostic tests regarding the assumptions of OLS. If such tests reveal unsatisfactory attainment of OLS assumptions, the analyst concludes that the model is misspecified. The model is then respecified and the testing procedure is repeated. The overall process is repeated until the model is considered "congruent with the evidence" (Kennedy (1992)). That is, the model has a logically plausible predictive power, is theoretically and parametrically consistent, the residuals are completely random (i.e., they exhibit white noise), and the model successfully encompasses all rival specifications.¹⁹ The TTT approach is, therefore, considered a "testing down" approach.

Kennedy (1992) notes that neither the AER nor the TTT approaches are without criticism. Some criticisms

¹⁹ Parametrically consistent refers to a model's ability to predict observations not used to specify the model. Such consistency could be attained using post-sample prediction tests in which a portion of the data is removed from the initial specification process and used to validate the specified model.

directed specifically to the AER approach are 1) the use of econometric analysis to reinforce theoretical relationships; 2) the failure to conclude the existence of specification errors when diagnostic tests show dissatisfaction of OLS assumptions; and 3) maximizing R^2 (i.e., data mining). Critics suggest, for instance, that reliance on adjusted R^2 for model selection appeals to the unique features of the analyzed data, thereby allowing mere chance to become a determining factor in model specification (see Mayer (1975 and 1980) in Kennedy (1992)). Kennedy suggests this latter criticism supports the need for post-sample prediction tests.²⁰

Both the AER and TTT approaches are criticized for their lack of a well-defined structural approach to model specification. Further, both approaches are criticized for the expected occurrence of type I errors, a result due to the multitude of diagnostic tests performed under each approach. Under the TTT approach, this result is also due to the loss of degrees of freedom for general model specifications.²¹

Kennedy (1992) suggests the following principles for model specification: 1) economic theory should serve as the

²⁰ Post-sample prediction tests were omitted from this study.

²¹ Kennedy (1992) suggests reducing the critical region or the probability of a type I error to mitigate such results when the TTT approach is used.

basis for model specification; 2) diagnostic tests revealing residuals not satisfying OLS assumptions should point to specification errors; 3) "testing down" is preferred to "testing up"; 4) tests for misspecification (e.g., tests for omitted variables, functional form, homoscedasticity, outliers, etc.) should be performed simultaneously to mitigate the possibility of erroneously selecting one misspecification over another (e.g., heteroscedasticity versus functional form); 5) a large number of misspecification tests should be employed including post-sample prediction tests; 6) models selected as correct should encompass rival models; and 7) limitations of the selected model should be reported along with the methods used to arrive at the selected the model. Because these principles compare more favorably with the TTT approach than the AER approach, Kennedy appears to lean toward suggesting a TTT approach to model specification.²²

Based on this review, an approach similar to the TTT approach appears to have the greatest merit in model specification analysis. Kennedy's literature review and comments are sufficiently compelling to use a TTT approach to recognize that unsatisfied OLS assumptions may mean the model is misspecified. Further, since this approach begins

²² Both Kennedy (1992) and Harvey (1990) note that the process of model selection is complex and that no generally accepted method has been adopted by econometricians in general.

with the most general form of the model, which, as noted below, is the testable hypothesis of this study, and tests down to a specific form, it is the most applicable approach to model selection for this study. Consistent with the focus of the paper, diagnostic tests for functional form, parameter consistency, outliers, homoscedasticity, and the degree in which the selected model encompasses rival models are reviewed in this chapter for use in this study. However, due to the narrow focus on functional form, simultaneous tests of misspecification such as heteroscedasticity versus functional form are omitted from the analysis in this study.²³

The Box-Cox Family of Power Transformations

As noted in chapter one, demand theory lacks sufficient *a priori* theoretical guidance for complete mathematical form specification of the demand function. Kmenta (1986) suggests that in such circumstances the form becomes a testable hypothesis and lists several methods for testing the form specification of an econometric model. Two of the methods suggested by Kmenta to determine functional form use the Box-Cox family of power transformation described in equations (1) and (2) in chapter one. One use of the transformation is to test a linear form against an

²³ See the comments regarding limitations of the analysis presented herein regarding simultaneous tests for misspecifications in chapter five, *infra*.

alternative in which all λ_v and λ_D , in terms of equation (2), are estimated with equal value (see, e.g., Kmenta (1986) and Zarembka (1968)). In this case, the linear and Box-Cox transformed models are discriminated using a likelihood ratio test (described below). Similarly, a double-log model, in which $\lambda_v = \lambda_D = 0$, may also be tested against a form in which λ_v and λ_D are estimated with equal value.

A second use of the Box-Cox transformation is suggested when the model's general form is questioned. In terms of equation (2) λ_v and λ_D are varied independently. The model's form in this approach is termed "flexible" and is used to test *a priori* specifications against a form determined by the data. Use of this approach allows any number of form restrictions on λ_v and λ_D as testable hypothesis against the flexible form determined by the data.

As noted in chapter one, Duffield (1988) found that several specifications of the form of the demand function failed to accurately describe variation in per capita demand for trips. Furthermore, it was found that a double-log model with a constant added to the price-proxy variable (average round-trip distance) significantly enhanced the predictive power of the model.²⁴ Given the results of the

²⁴ Duffield (1992) improved this approach by replacing round-trip distance with predicted total trip costs based on functions for resident and non-resident anglers that were developed by Duffield, Loomis, and Brooks (1987) to estimate variable travel costs.

specifications examined, Duffield (1988) suggested searching a family of power transformations on the price-proxy variable. Thus, the general form of the first-stage Montana stream fisheries demand function becomes the testable hypothesis in this study. This is consistent with McConnell's (1985) suggestion that functional form should be tested when using the TCM framework.

Purpose and Assumptions of the Basic and Extended Box-Cox regression. The Box-Cox transformation was initially intended for use on the dependent variable to achieve a simple model structure, constant error variance (homoscedasticity), and normal distribution of the errors (Box and Cox (1964)). However, in the initial and subsequent applications it has been suggested that the transformation can also be applied to the independent variables, resulting in the extended Box-Cox model. The extended Box-Cox regression allows flexibility in determining functional form by estimation of the transformations of the variables in the model according to the data. The continuous nature of the Box-Cox transformation allows one to estimate both intrinsically linear and nonlinear forms.

Econometric models are classified as either intrinsically linear or nonlinear. An intrinsically linear function is linear in the estimated parameters but nonlinear

in the variables. Functions specified with polynomials, interaction terms, and as either multiplicative or additive generally qualify as intrinsically linear forms. Such forms are also marked by the error term specified as additive. Alternatively, an intrinsically nonlinear function is nonlinear in the variables and estimated parameters. While such functions may also be multiplicative or additive, they are marked by an error term specified as multiplicative in the model. OLS can be used to estimate all intrinsically linear and some nonlinear forms by transforming the variables into linear forms (although nonlinear forms are generally estimated using maximum likelihood). However, when the error term is specified as multiplicative in nonlinear models, transformations of the variables leads to a nonlinear distribution of the error term and OLS estimation of the function is appropriately termed nonlinear least squares estimation (see, e.g., Kmenta (1986)).

As noted below, OLS could be used to estimate a Box-Cox regression such as that specified in equation (2). However, the Box-Cox transformations are considered *a priori* fixed in this process. In an AER approach to model specification, the form of the model is determined incrementally. However, the Box-Cox method allows flexibility in determining the form of the model in terms of how each variable enters the model, i.e., linearly or nonlinearly. The intuitive appeal of the Box-Cox method is

that greater flexibility is allowed to determine the form of the model by allowing variable linearization in conformation with the data and not a *a priori* expectation. Thus, testing down from the most general to a specific form using a TTT approach is made possible with the extended Box-Cox model.

Methods Used to Estimate a Box-Cox Regression.

Spitzer (1982a and 1982b) lists four approaches to estimating a Box-Cox regression including maximum likelihood and concentrated maximum likelihood, nonlinear least squares, and iterative OLS. This review shows that maximum likelihood or concentrated maximum likelihood estimation provides the most accurate and least costly estimates of the parameters in a Box-Cox regression equation for purposes of hypothesis testing, assuming the software is available. For illustrative purposes, these methods are presented in terms of equation (2). Each of the above listed methods are summarized in turn.

Maximum likelihood estimation (MLE) can be described by comparison with OLS estimation. Regression coefficients estimated using OLS are estimated by minimizing the sum of the squared errors of the regression function from their mean. Further, to attain best unbiased estimators and determine confidence intervals for the estimated parameters, the residuals are assumed to satisfy the classical OLS assumptions summarized at the outset of this chapter. In contrast, parameters estimated using MLE are those which are

more likely to describe the population from which the data are drawn than by any other set of parameters describing a different population (Kennedy (1992)). When the regression variables are not subject to transformations or transformations are assumed fixed, MLE and OLS produce identical estimated regression coefficients.

Kennedy (1992) summarizes MLE in four steps. These steps are summarized assuming the Box-Cox regression function given in equation (2). *First*, the distribution of the error term in the regression equation is specified. One common specification, and that used in this study, is that the error terms are assumed to be independently, identically, and normally distributed ($e \sim N(0, \sigma^2)$). *Second*, the relationship of the error terms are specified in terms of the variables in the regression function. *Third*, given the above specification of the error terms, namely their independence, the likelihood function equals the product of $f(e_i)$ across the sample or the joint probability distribution function of the error terms. Since the natural log of the likelihood function is a monotonic transformation of the likelihood function, the log-likelihood function is

stated as follows:²⁵

$$(3) \quad \ln L = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (V_i^{(\lambda_v)} - \beta_1 - \beta_2 D_i^{(\lambda_D)})^2 \\ + (\lambda_v - 1) \sum_{i=1}^N \ln V_i$$

The last term in equation (3), $(\lambda_v - 1) \sum_{i=1}^N \ln V_i$, is the jacobian determinant of the transformation. This term accounts for possible differences in the probability distributions (probability density functions) of the error term and the dependent variable when the latter is transformed. The log-likelihood function is specified in terms of the unknown, assumed distribution of the error terms. However, known observations of the dependent variable are used to estimate the parameters of the function. Absent the jacobian determinant, the log-likelihood function assumes the distribution of the error terms and the dependent variable are the same. This is true when the dependent variable is not transformed (i.e., $\lambda_v = 1$), thus the jacobian determinant in this case would be zero (see equation (3)). Differences in the probability distribution of the error terms and the dependent variables will occur for all transformations on the dependent

²⁵ See Kennedy (1992), chapter 2 for details of the general development of this function.

variables when $\lambda_v \neq 1$. Thus the jacobian determinant is included in the log-likelihood function to adjust for the differences in the probability distributions of the dependent variable and the error terms (see Kennedy (1992) and Kmenta (1986) for further explanations).

The *fourth* step is estimation of the parameters β_0 , β_1 , λ_v , λ_D , and σ^2 in equation (3). Estimates of these parameters can be found by maximizing equation (3) with respect to each parameter. Necessary and sufficient conditions for a local maximum requires the first and second order conditions (first and second derivatives) of equation (3) to be equal to zero and negative definite, respectively. Theoretically, the negative of the inverse of the expected value of the second order conditions yields the variance-covariance matrix. However, due to the complexities involved in this computation, Spitzer (1982a) maintains the negative of the inverse of the second order conditions produce acceptable results.

A second approach to estimating the parameters of a Box-Cox regression function exploits the fact that the estimate of σ^2 , given in equation (4), reduces equation (3)

to equation (5) (see Greene (1990) and Spitzer (1982a)).²⁶

$$(4) \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^N e_i^2$$

$$(5) \quad \ln L_c = -\frac{N}{2} (\ln(2\pi) + 1) - \frac{N}{2} \ln \hat{\sigma}^2 + (\lambda_v - 1) \sum_{i=1}^N \ln V_i$$

That is, σ^2 is concentrated out of the log-likelihood function. Spitzer (1982a) shows that the first and second order conditions for maximization of equation (5) are identical to those produced by equation (3).

In a third approach to estimating equation (2), initially attributed to Zarembka (1968), data are rescaled by the inverse of the geometric mean of the dependent variable raised to the power of λ_v . Spitzer (1982a) shows that the resulting concentrated log-likelihood function reduces the estimation problem to non-linear least squares. Optimal values of the estimated regression parameters are obtained by minimizing the standard error of the estimate of the scaled model. However, different from the MLE approaches, the negative of the inverse of the second order condition of the NLS concentrated log-likelihood function

²⁶ Based on the literature regarding simultaneous tests of function form and heteroscedasticity, the specification of σ^2 in equation (4) assumes the error terms are homoscedastic (see, e.g., Kmenta (1986) and Greene (1990)).

does not provide the variance-covariance matrix directly due to the rescaling of the data. Thus, the variance-covariance matrix must be estimated by means of converting the estimated scaled coefficients to their original form (see Spitzer (1982a) for the appropriate adjustment). Thus, β_0 , β_1 , λ_v , λ_D , and σ^2 are not a direct result of minimizing σ^2 of the scaled model.

Estimates of the regression parameters may also be made using an iterative OLS/grid-search method. According to Spitzer (1982a) a series of regressions are estimated for β_0 , β_1 , and σ^2 using the scaled model described for the third approach above is one approach. In terms of equation (2), each regression would differ according to each of the values of λ_v and λ_D , assumed for each estimation. The combination of λ_v and λ_D which minimizes the sum of the squared errors for the scaled model, found by scanning a grid of these values, yields the same results as if any of the above three methods were used. However, since the data are scaled and λ_v and λ_D are assumed fixed for the regression function minimizing the sum of the squared errors, the estimated coefficients must be rescaled and, more importantly, the variance-covariance matrix must be adjusted. Spitzer (1984) and Kmenta (1986) note that the standard errors of the estimated coefficients will be biased downward, yielding inflated t-ratios resulting in possible errors in hypothesis testing. Thus, similar to the scaled-

model approach summarized above, the variance-covariance matrix must be adjusted to validate hypothesis testing (see Spitzer (1982a) for details).

There are several advantages to using the MLE approaches to estimating the Box-Cox regression parameters over the alternative methods listed here. Foremost is that there is currently available software to compute a valid variance-covariance matrix directly from the second order conditions of the log-likelihood functions without adjustment.²⁷ Second, Kennedy (1992) notes MLE has several attractive large-sample asymptotic properties. Namely, maximum likelihood estimators are asymptotically unbiased, consistent, and efficient. These features are particularly attractive when compared to OLS in that each may be considered comparable to unbiasedness, efficiency, and BUE, respectively, which are results of satisfaction of the classical assumptions of OLS for large samples. Additionally, in terms of the mechanics of estimation the OLS is limited to internal estimation of only the constant and slope coefficients, whereas MLE produces estimates of these parameters in addition to the variance and Box-Cox transformations (see e.g., Kmenta (1986)). Finally, the MLE approach does not require additional adjustments to variance-covariance matrix as does an NLS or OLS approach.

²⁷ For instance SHAZAM (White (1990)) and LIMDEP (Greene (1990)) are two econometric programs featuring estimation of Box-Cox regressions.

Based on this analysis, software in which a Box-Cox regression is estimated using the full or concentrated log-likelihood approach would provide the best parameter estimates, thereby mitigating erroneous hypothesis testing. However, should an IOLS/grid-search approach be necessary, it would be desirable for the software to incorporate the appropriate adjustments to the variance-covariance matrix as noted by Spitzer (1982a and 1982b). *LIMDEP*, version 6.0 (Greene (1991)) uses algorithms which satisfies both of these constraints. Thus, *LIMDEP* was chosen for the software to estimate the first-stage demand function. Moreover, Spitzer (1982b) notes that computer programs using optimization algorithms which are limited to use of the first derivative of the log-likelihood function provide larger variances than do programs using first and second order conditions. Greene (1991) specifically states that first and second order conditions are used for MLE in the *BOXCOX* procedure in *LIMDEP*.²⁸ ²⁹

²⁸ It appears the IOLS/grid-search method provided in *LIMDEP* makes the appropriate adjustments to the variance-covariance matrix as suggested by Spitzer (1982a). The magnitudes of the estimated t-ratios resulting from the *LIMDEP* IOLS procedure appear similar in magnitude to those resulting from the full MLE procedure (compare models 3.a and 5, table 6, chapter four).

²⁹ Several issues regarding estimation of a Box-Cox regression have been identified. For instance, the distribution of the error terms will be truncated since the data for transformed variables must be positive (Smith 1975) and the equation must contain a constant term (Schlesselman (1971) in Spitzer (1982a)). The last of these concerns is satisfied in this study since the demand function is

Methods of Model discrimination

As noted, the model specification determined best using the TTT approach should encompass all rival specifications. The methods used to discriminate such models depends on whether one model can be nested in the alternative. That is, discrimination tests of nested models may not be applicable for non-nested models (see e.g., Kmenta (1986)). A model is nested in another model if it is a restricted case of the more general model in which it is nested. For instance, if in equation (2) $\lambda_v = \lambda_D = 1$, this form would be considered nested in a form in which λ_v λ_D were allowed to vary independently. This section briefly summarizes two approaches used to discriminate nested and non-nested models in this study: the likelihood ratio test and Davidson and MacKinnon's J-Test.

The Likelihood Ratio Test. The likelihood ratio test is used to discriminate general (unrestricted) Box-Cox regressions with their restricted counterparts. The concept of the test is that the value of the maximized likelihood function for the restricted and unrestricted models will be similar if the restricted model is correct (Kmenta (1986)). Thus, the ratio of the likelihood function for the restricted and unrestricted models would converge to a value of one. More specifically, the ratio of the estimated

specified with a constant term. The remaining concerns are not considered in this study.

regression variances would be nearly equal. Thus, as noted by Kmenta (1986), the null hypothesis in a likelihood ratio test is that the restrictions imposed on the unrestricted model are correct and the alternate hypothesis is that the restrictions are not correct. If the null hypothesis is true, then the values of the maximized log-likelihood functions for the restricted and unrestricted models are similar. This results in a log-likelihood ratio given in equation (6):

$$(6) \quad LR = -2[L(\lambda_R) - L(\lambda_U)] \sim \chi_m^2$$

In equation (6) $L(\lambda_R)$ and $L(\lambda_U)$ equal the values of the log-likelihood function when the Box-Cox transformations are restricted and unrestricted, respectively, and m equals the number of parameter restrictions. In terms of a test between the linear and most general form of equation (2), for example, $L(\lambda_R)$ and $L(\lambda_U)$ would be replaced with values of the maximized log-likelihood function when $\lambda_v = \lambda_D = 1$ and when λ_v and λ_D were allowed to vary either together or independently, respectively. As noted in equation (6), the log-likelihood ratio is asymptotically distributed as a chi-squared (χ^2) distribution with degrees of freedom, m , equal to the number of restrictions in the restricted model.

Two rival models can be discriminated using a log-likelihood ratio test providing the restricted model can be nested as special cases of the unrestricted or less

restricted model (Greene (1990)). However, the likelihood ratio test cannot be used to discriminate two functions in which neither can be nested in the other. For instance, a semi-log cannot be discriminated from a double-log model using a likelihood ratio test. In such instances other tests such as the J-test are applicable, which is summarized below.

Tests for Non-nested Models. Generally, if two rival models are non-nested, traditional methods of model discrimination such as an analysis of variance (nested F-test) or adjusted- R^2 are not applicable (Kmenta (1986)).³⁰ Kennedy (1992) lists several methods for statistically discriminating rival non-nested models which are classified as variants of either Cox's test or Davidson and MacKinnon's J-test.³¹ The J-test consists of model discrimination based on whether the predictive power of one model is enhanced by the predictive power of a rival model. The complete test examines how each pairwise combination of rival models encompasses each other's predictive power. Details of application of the J-test are summarized in appendix B.

³⁰ Model discrimination using R^2 is applicable when the dependent variables are equally defined and transformed.

³¹ For the readers information, variants of Cox's test are used to discriminate models based on their variances.

Diagnostic Tests Useful for Model Specification

One aspect of the TTT approach to model specification is to subject an estimated model to a series of diagnostic tests. The diagnostic tests used in this study summarized below include tests for normality and homoscedasticity assumptions of the residuals and detection and measurement of the influence of data outliers.

Methods for Testing the Normality of Residuals.

Neter *et al.* (1989) suggests the use of normal probability plots as one means of determining whether the residuals are normally distributed. A normal probability plot consists of plotting the residuals (or standardized residuals) against the expected values of the residuals under normality. Large deviations of the residuals from a straight, forty-five degree line representing the expected values of the residuals when conforming precisely to normality suggest deviations from normality.

Detection of Homoscedasticity. Tests for homoscedasticity consist of visual examination of residual plots and formal statistical tests. Neter *et al.* (1989) suggests examination of residuals (or standardized residuals) plotted against the predicted or fitted values of the dependent variable to determine whether the variance of the residuals are constant. If the scatter of the plotted residuals tends to flair either to the right or left, the residuals display a monotonic non-constant variance.

Bulges, either at the left and right of the plot or around the center of the plot may mean that the residuals are non-constant and non-monotonic. If the plotted residuals appear evenly distributed around a value of zero along the residual axis, the variances are considered homoscedastic.

Although detection of and corrections for heteroscedasticity in conjunction with the Box-Cox regression model is beyond the scope of this study, it is important to note whether the Box-Cox model satisfies the assumptions of OLS, including the homoscedasticity of the residuals. Two tests for heteroscedasticity are used in this study.

First, White's test (see Kmenta (1986)) provides a means of detecting heteroscedasticity without knowledge of the form of the non-constant error variance such as required by the Glejser test. Moreover, the test does not require the residuals to be normally distributed. White's test consists of regressing the squared errors of the model being tested on the variables in the model in their modeled form plus the square of each variable and interaction variables for each paired combination of the variables in the model. The null hypothesis of White's test is that the variances of the residuals are constant. Failure to reject this hypothesis also implies that any heteroscedasticity is caused by sampling error. The asymptotic, large-sample statistic is computed by $N(R^2_w)$ which is distributed as a

χ^2 with p degrees of freedom. A complete description of the regression function used in White's test is provided in Appendix C.

Additionally, the Glejser test is used in cases when the degree of colinearity prohibits use of White's test. The Glejser test consists of determining whether a significant correlation exists between the absolute value of the residuals and the variable assumed to be the cause of heteroscedasticity using a regression model. This determination is made by choosing from an array of presumed forms of heteroscedasticity as reflected in different regression equations (see Koutsoyiannis (1977)).

Methods for Detecting and Measuring the Influence of Outliers.³² One approach to detecting data outliers is done by visually examining scatter plots. Neter et al. (1988) identify several useful scatter plots, two of which are summarized here. First, plotting observations of the dependent with an independent variable can reveal observation lying outward of the cluster of observations around the estimated regression line (see also Weisberg (1980)). Second, in a plot of the standardized residuals against an independent variable or predicted values of the dependent variable observations appearing isolated from remaining residuals might be considered outliers.

³² This section is based largely on Neter et al (1989).

Neter *et al.* (1989) suggest use of DFFITS and DFBETAS measure to detect influential observations, both of which make use of leverage values. DFFITS measures the influence a particular case has on the fit of a regression equation. DFBETAS measures the influence a particular observation has on each of the regression coefficients. In large data sets, the absolute value of DFFITS values exceeding the $2(k/n)^{1/2}$ statistic, where k and n are the number of estimated coefficients and n equals the sample size, indicates an influential case affecting the fit of a model. Absolute values of DFBETAS values exceeding $2/(n)^{1/2}$ indicates cases which affect the estimated constant or coefficient of the particular variable under evaluation. Details of the equations making up the DFFITS and DFBETAS measures are summarized in Appendix C.

Once data outliers are identified, specification of the function with dummy variables can be used to assess the collective impact a group of outliers has on certain coefficients including the intercept. Kmenta (1986) shows how this can be done using a dummy variable specified in the function as variable affecting the intercept and as an interaction term with any of the variables in the function thereby affecting the slope of these variables. Details of such an application are summarized in Appendix C.

Conclusions and Summary

This chapter showed that a TTT, versus an AER approach to model specification is preferred for purposes of mitigating type I errors in the process of selecting an appropriate specification of an econometric model. Thus, a TTT approach to statistical determination of the functional form of the first-stage TCM demand function is used in this study. However, due to the information presented in previous analyses of the DFWP stream fisheries data base (namely, Duffield *et al.* (1987) and Duffield (1988) the approach resulting from this study and prior studies would more accurately be labeled as an AER/TTT hybrid approach.

Second, use of a full MLE method was found preferable to the NLS or IOLS/grid-search approaches. This is due largely to the bias in variance estimates resulting in the use of these latter two methods. Third, log-likelihood ratio test for discriminating restricted Box-Cox regression models from more general specifications, non-restricted forms was reviewed. Also reviewed was the Davidson and MacKinnon J-Test for discriminating non-nested models.

Finally, in accordance with the TTT approach to model discrimination, several diagnostic tests useful for determining functional form by statistical inference were reviewed. These tests included, 1) methods to determine whether the residuals satisfy the OLS assumption of

normality, including visual examination of normal probability plots, 2) methods to determine whether the residuals satisfy the OLS assumption of homoscedasticity, including White's and Glejser's tests, and 3) methods for identifying outliers such as plots of the data and of residuals and analytical methods such as DFFITS and DFBETAS measures.

CHAPTER 3

DATA SOURCES AND DESCRIPTION ³³

As noted in chapter one, this study is a reexamination of the original TCM demand analysis performed by Duffield *et al.* (1987) and Duffield (1988) on Montana cold water stream fisheries. This chapter provides a description of this data base (or the cold-water stream fisheries data base). First, a description of the sites included in the study is provided. Second, a list and description of the sources of data used in this study from the cold-water streams fisheries data base is provided. Last, the methods used to gather, organize, and aggregate all primary data included in the data base and used in this study are summarized.

Description of the Sites Comprising the Montana Stream Fishery

The data set used in this paper contains TCM data for 28 tributaries and/or watersheds and 20 "unique waters" in Montana identified by the Montana Department of Fish,

³³ The data used in this study were developed for estimation of demand and net economic values for cold water fishing by Duffield *et al.* (1987). No alterations were made to the untransformed data used in that study.

Wildlife, and Parks (Duffield *et al.* (1987)). The 28 non-unique waters include tributaries to unique waters or watersheds comprised of a river and its tributaries. These streams and tributaries define the regional stream fishery used in this paper.³⁴ Tables 1 and 2 list the 28 non-unique waters and the 20 unique waters, respectively.

³⁴ A regional TCM is generally used to analyze demand for a collection of recreation sites for a specific recreation activity such as hunting or fishing. Thus, site specific models pertain only to one site.

TABLE 1.-- Montana Stream Fisheries: Non-Unique Waters.

River Code	River	Definition
11	Flathead, South Fork	Entire drainage
12	Flathead, Middle Fork	Entire drainage
14	Flathead, North Fork	Entire drainage
15	Flathead	River and tributaries below confluence with south fork and excluding river 89
16	Lower Clark Fork	River and tributaries below confluence with Flathead (Paradise)
17	Kootenai Tributaries	Excludes mainstem river 91
21	Upper Clark Fork Tributaries	Above confluence with blackfoot (Milltown) and excluding mainstem river 86
22	Blackfoot Tributaries	Excludes mainstem river 83
23	Rock Creek Tributaries	Excludes mainstem river 94
24	Bitterroot Tributaries	Excludes mainstem river 82
25	Middle Clark Fork Tributaries	Paradise to Milltown excluding mainstem river 87
31	Upper Yellowstone Tributaries	Springdale to Gardener excluding mainstem river 98
32	Gallatin Tributaries	Excludes mainstem river 90 and 88
33	Upper Missouri (Region 3)	River and tributaries from Threeforks to Canyon Ferry
34	Madison Tributaries	Excludes mainstem river 92
35	Jefferson	Entire drainage
36	Beaverhead Tributaries	Excludes mainstem river 80
37	Big Hole Tributaries	Excludes mainstem river 81
41	Middle Missouri	River and Tributaries below Marias River and above Fort Peck
42	Smith Tributaries	Excludes mainstem river 95
43	Upper Missouri (Region 4)	Canyon Ferry to Marias River excluding mainstem river 93
44	Marias	Entire drainage
52	Middle Yellowstone Tributaries	Springdale to confluence with Bighorn excluding rivers 99, 84, 85, 96, 55, and 56
55	Stillwater Tributaries	Excludes mainstem river 96
56	Boulder Tributaries	Excludes mainstem river 84
61	Lower Missouri	River and tributaries from upper end of Fort Peck Reservoir to North Dakota Border
62	Milk	Entire drainage
71	Lower Yellowstone	River and tributaries below confluence with Big Horn

Source: Duffield et. al. (1987, table 1.)

TABLE 2.-- Montana Stream Fisheries: Unique Waters

River Code	River	Definition
80	Beaverhead	Mainstem
81	Bighole	Mainstem
82	Bitterroot	Mainstem to confluence with East and West Forks
83	Blackfoot	Mainstem
84	Boulder	Mainstem
85	Bighorn	Mainstem
86	Upper Clark Fork	Mainstem above Milltown
87	Middle Clark Fork	Mainstem Milltown to Paradise
88	East Gallatin	Mainstem
89	Upper Flathead	Mainstem above Flathead Lake to confluence of South Fork
90	Gallatin	Mainstem
91	Kootenai	Mainstem
92	Madison	Mainstem
93	Missouri	Mainstem, Holter to Cascade
94	Rock Creek	Mainstem near Missoula
95	Smith	Mainstem
96	Stillwater	Mainstem near Absarokee
97	Swan	Mainstem
98	Upper Yellowstone	Mainstem Springdale to Gardener
99	Middle Yellowstone	Mainstem Springdale to confluence with Bighorn

Source: Duffield et. al. (1987, table 1.)

Data Sources Included in the Cold-Water Stream Fisheries

Data Base

The cold-water streams data base used in this study is comprised of three sources. Chief among these sources are the survey data collected by the Montana Department of Fish Wildlife and Parks (DFWP) through the Montana Statewide Angling Pressure Mail Survey for the 1985 license year (McFarland, 1989, henceforth fisheries survey) and a supplemental telephone survey conducted in 1985. These data sources provided information regarding sites fished, anglers' origins, distance to the site, harvest rates, socio-economic data, and travel and time costs. These two data sources were supplemented with estimated map distances in cases where the reported distance traveled were found to be in error (Duffield (1988)). Distances in these cases were computed from the *Rand McNally Road Atlas: U.S., Canada, Mexico* (1977).

Data Preparation for First-Stage TCM Demand Function

Estimation

Collection of Survey Data. DFWP conducted the fisheries survey during each of the license years of 1982 through 1985 beginning spring 1982 (see, McFarland, p. 2). Questionnaires were mailed monthly to resident and non-resident licensed anglers within one month of the angler's purchase of his or her license. Approximately 1,500 and 100

surveys were sent to resident and non-resident anglers per month, respectively, during 1982 through 1984. Beginning with the 1984 license year the frequency of survey distribution was generally increased to biweekly mailings for the months of March through October, corresponding with high fishing pressure months. This increase in the distribution rate for the same number of monthly surveys was implemented to mitigate memory bias. Non-resident anglers who purchased 2-day licenses were surveyed on an annual basis.

A random drawing of anglers was ensured by using a stratified sampling procedure (see, McFarland (1989), pp. 3 and 144-145). Questionnaires sent during 1985 sought detailed information for each fishing trip taken during the particular sampling period for which the survey was made. This information can be noted from a copy of the surveys sent provided in Appendix D.

The number of monthly fisheries surveys was increased to about 3,000 and 250 monthly mailings for each of resident and non-resident license holders, respectively, in 1985. This increase in frequency was made to accommodate the data needs for the purposes of the Montana Bioeconomics Study. Of the 36,000 surveys sent during 1985, 92% were sent to resident and 8% were sent to nonresident anglers. The overall response rate was 54% or 19,271. (Duffield et al. (1987)).

For the purposes of TCM analysis the data sample collected through the fisheries survey were reduced as follows. First, respondents that had not fished during the month or two-week time period specified on the survey were excluded. Additionally, those who were determined to have been on multi-purpose and/or multi-site trips and those making over-night trips were also purged from the data set. Fishing sites were then coded according to DFWP management designations. Thus, data for the cold-water stream analysis includes those who had fished in the trout stream designated in tables 1 and 2.

The second major data source used in this study is taken from a supplemental survey to the 1985 fisheries survey. The data in this set were gathered by a telephone survey of 2,000 resident and nonresident angler. This survey was administer by DFWP during September and October of 1985. This sample consisted of 1,600 resident and 400 non-resident anglers and produced a response rate of 80% for residents and 52% for nonresidents. The same criteria noted above was used for selecting qualified respondents for TCM analysis. Three types of data were collected in this survey: 1) socio-economic data, such as fishing experience, age, income, and education; 2) data characterizing site use behavior, such as time spent at the site, time spent fishing, and equipment used; 3) and travel cost data such as, expenditures, travel time, and distance. Socio-economic

data from the supplemental survey were appended to the fisheries survey by origin. Site characteristic data were appended by site. Travel costs data were appended by origin/destination pairs. (See Duffield et al. (1987) and McFarland (1985a and b)).

Aggregation of the Supplemented Fisheries Survey

Data. As noted, the demand functions estimated in this study are based on zonal aggregations of fishing trips for a given region. Thus, individual survey responses for trips taken to a given site were aggregated by origin zones. Responses are aggregated by zones of nearly equal distant origins, using the following rules:

1. An origin zone consists of a single county if the county contains the destination site or is contiguous to the county or counties containing the site.
2. Several counties were lumped together to define a zone of intermediate distance observations. This was done to prevent concentric gaps which would otherwise result from zero observations from counties of intermediate distances. Therefore, observed trips from these "super-counties" represents the population of counties where no trips were recorded within the sample.
3. Contiguous counties of states bordering Montana were treated as individual zones.
4. Five nonresident market regions were defined to allow nonresident trips to represent the population of their home state and that of all states they traveled through to get to the site, providing no trips from these states were included in the data base. These market regions were constructed as spokes extending outward from Montana throughout the continental United States. Again, this method of aggregation was used to avoid any

concentric gaps in the surrounding market zones between the site and the non-resident's origin.

Independent variables were summed together and the aggregation process resulted in 839 origin-destination pairs for the cold-water stream fisheries data base. Aggregation methods of origin-destination data by site presented problems with regard to the accuracy of distance traveled. Cases were found in which anglers who traveled to a particular site from the same origin zone may have traveled disparate distances. In fact, the survey data on distance were found unreliable for 50% of the origin-destination pairs. In these cases, map distances were used in lieu of reported distances and population weighted average distances were recomputed (Duffield, 1988).

The sample size for the estimated demand functions presented in chapter four is 741 origin-destination pairs. This size is the combined result of two adjustments made on the stream fisheries data base. First, to remain consistent with previous studies on this data set four tributaries regions (river codes 11, 23, 56, and 61) were excluded from the complete set of DFWP administrative regions (see table 1). The sample was further reduced to origin-destination pairs with complete information on socio-economic and demographic variables as was the case for samples used for previous modeling efforts. Table 3 provides descriptive statistics for the variables specified in the models

presented in Chapter 4.

Table 3.-- Descriptive Statistics for the Variables Specified in the Models Presented in Chapter 4.

Variable	Mean	Standard Deviation	Min	Max	Label
TRIPS	6.79	16.89	1.0	153.00	Zonal Aggregated Trips
POP	4.324 E6	9.966 E6	1.0 E3	7.809 E7	Zonal Aggregated Population
TRIPCAP	2.71 E-4	1.02 E-3	1.58 E-8	1.51 E-2	Per Capita Trips
LTRIPCAP	-11.71	3.04	-17.96	-4.19	Natural Log of Per Capita Trips
CAVEDIST	1119.55	1196.89	4.00	4937.50	Average Round-Trip Distance
LCVEDIST	6.21	1.54	1.39	8.50	Natural Log of Average Round-Trip Distance
LKVEDIST	6.49	1.18	4.22	8.52	Natural Log of Average Round-Trip Distance plus 64
TRIP3A	-11.59	2.98	-17.70	-4.18	Box-Cox of Per Capita Trips at .0016
CDIST3A	11.54	4.16	1.56	18.99	Box-Cox of Average Round-Trip Distance at .169
TRIP3B	-9.34	1.97	-13.02	-3.87	Box-Cox of Per Capita Trips at .038
CDIST3B	7.06	1.91	1.42	10.04	Box-Cox of Average Round-Trip Distance at .038
TRIP4A	-13.05	3.70	-21.00	-4.34	Box-Cox of Per Capita Trips at -.017
TRIP4B	-9.78	2.16	-13.88	-3.94	Box-Cox of Per Capita Trips at .03
CDIST5	11.70	4.25	1.57	19.33	Box-Cox of Average Round-Trip Distance at .172

The sample size for all of the variables listed in this table is 741.

CHAPTER 4

ESTIMATION AND ANALYSIS OF DEMAND FUNCTIONS

This chapter provides a summary of the mathematical form analysis performed on the bivariate first-stage TCM demand function for the stream fisheries data. As noted in Chapter 2, this study takes a model specification approach which combines the AER and TTT approaches. The AER aspect of this analysis consists of the use of two prior studies performed on the stream fisheries data base. The TTT aspect consist of several diagnostic tests on the bivariate forms of the two previously estimated demand function summarized below, estimation and diagnostic testing of the Box-Cox forms of the TCM demand function, and comparison of this model with the bivariate forms of the previously estimated models. This chapter is organized as follows.

The first section reviews the principles applied by Duffield *et al.* (1987) and Duffield (1988) to estimate two different functional forms of the TCM demand function. Bivariate models based on these principles were estimated and are presented in this section. The fit of these models is summarized along with an analysis of adherence to certain

assumptions of OLS. This section serves as a starting point for the analysis of the regional, zonal aggregate TCM demand functions presented in this paper.

Based on the foundation set in section one, a determination of the most statistically sound and practical functional form(s) of the bivariate TCM demand function is presented in section two. This section includes a synopsis of the *a priori* theoretical expectations of the signs and magnitudes of the parameters estimated using various specifications of the original Box-Cox and extended Box-Cox regression approaches. This is followed by a summary of five models estimated using various combinations of the Box-Cox transformations on the dependent and independent variables. These models are then discriminated to determine the form which best satisfies the maximum likelihood principle from this field. An analysis of this model's adherence to certain OLS assumptions is also summarized.

Finally, comparisons between the optimal Box-Cox estimated model and the bivariate forms of the two previously estimated models review in section one are continued in the last section. This includes a comparison of the predictive power of each model and a conclusion regarding which model describes per capita trip demand best. As an aside, the elasticities resulting from each model are provided in this section.

Bivariate Estimates Two Previously Estimated Demand Functions

As noted in chapter 1, two functional forms were used to estimate the first-stage demand function in two previous studies. In the first of these studies, Duffield *et al.* (1987) specified a multivariate, double-log (log-linear) form. This form was specified primarily to satisfy the constraint of economic theory that there be diminishing marginal value for each additional fish caught. This form was also selected to avoid the possibility of predicting negative per capita trips from distant origin zones and to minimize heteroscedasticity of the residuals expected to occur when population varies across origin zones. Both of these latter two problems were expected to occur with the linear form. However, as noted, the model estimated by Duffield *et al.* (1987) failed to produce homoscedastic residuals and overpredicted observed trips by about 60 percent. The overprediction was found to be largely influenced by trips taken from origins within 20 miles of a site. The model also underpredicted trips from intermediate distances and overpredicted trips from longer distances.

Driven primarily by the concerns of heteroscedasticity and prediction, Duffield (1988) examined three general alternative functional forms including polynomial, semi-log, and double-log forms. Although the polynomial specifications, including additions of distance

squared and cubed terms and a hybrid double-log-polynomial form enhanced trip prediction, demand was positively sloped for distances greater than 3,000 miles. The semi-log specification also enhanced trip prediction. However, converse to the double-log model, the semi-log model overpredicted trips at intermediate distances and underpredicted trips at short and long distances. Further, the coefficient of determination for the semi-log model was less than that of the double-log function (.22 for the semi-log versus .78 for the double-log form). A hybrid polynomial semi-log function was also estimated, but did not correct the heteroscedasticity problems noted above. This model also showed a positive slope at about 3,000 miles, an unacceptable theoretical result.

As a result of the above summarized analysis, Duffield (1988) found that a multivariate, double-log form in which average round-trip distance was shifted by constant of 90 enhanced trip prediction to within .1 percent of observed trips and satisfied the assumption of homoscedastic residuals.³⁵

Table 4 summarizes estimated bivariate models based on the forms specified by Duffield *et al.* (1987) (model 1) and Duffield (1988) (model 2). However, different from

³⁵ A constant variance of the residuals in this model was concluded based on visual examination of a plot of the residuals versus the predicted values of the model (Duffield (1988)).

Duffield's (1988) analysis, the shift parameter added to average round-trip distance was found to be 64 for the bivariate model.³⁶

³⁶ The optimal shift parameter of 64 miles was found by searching a range of parameters until adjusted- R^2 was locally maximized.

Table 4.-- Bivariate Specifications of Models Specified by Duffield *et al.*
(1987) (Model 1) and Duffield (1988) (Model 2).

Model	λ_v	λ_D	Log-Likelihood	β_0	β_1	\bar{R}^2	F
1.	0 (R)	0 (R)	7,331.34	-1.03	-1.71	.759	2,335.68
	(t-ratio)			(-4.53)	(-48.32)		
	(p-value)			(<.001)	(<.001)		(<.001)
2.	0 (R)	0 (R)	7,372.34	3.09	-2.28	.784	2,695.48
				(10.68)	(-51.91)		
				(.000)	(<.001)		(<.001)

β_1 is the estimated parameter for average round-trip distance and average round-trip distance plus 64 in each of models 1 and 2 respectively. The critical values for t and F at a 5 percent probability of a type I error 1.6471 and 3.854, respectively. The sample size for both models is 741.

Diagnostic Tests on Models 1 and 2. The following summarizes the fit and statistical significance of the estimated parameters in models 1 and 2. Also OLS assumptions of a normal distribution of the residuals and their variance consistency are examined. Further potential outliers resulting from models 1 and 2 are summarized.

As can be noted in table 4, both the double-log and double-log, shifted distance models have high adjusted-R² values suggesting the majority (75.9 and 78.4 percent, respectively) of the variation in the natural-log of per capita trips are described by the independent variables in each model. Further, with a critical value of 3.854, with degrees of freedom equal to 1 and 739, the hypothesis that β_1 is zero is rejected at a 5 percent probability of a type I error. Thus, each model constitutes a good fit. It is also noted that the signs of β_1 are accepted as correct at a 5 percent probability of a type I error.

Examination of the normal probability plots of the residuals resulting from models 1 and 2 (figures 1 and 2, respectively) suggest these models yield nearly normally distributed residuals. Moreover, model 1 appears to yield greater normality in the residuals. Figures 3 and 4 are standardized scatter plots of the residuals and predicted natural-log of per capita trips produced by each of Models 1

and 2 (table 4).³⁷ Four general conclusions can be made from these plots.³⁸ First, the residuals of the double-log, shifted distance model appear more evenly distributed around the center of the plot than the residuals produced by the double-log model (even though there appear more points in the south-west quadrant than any other quadrant). Thus, the double-log, shifted distance model appears to yield less heteroscedasticity in the residuals than the double-log model. However, the results of White's test for homoscedasticity with a null hypothesis of constant variance and alternate hypothesis of non-constant variance across the residuals suggests that at a 5 percent probability of a type I error the residuals in both models 1 and 2 are not constant.³⁹

Secondly, the inverted-U shape of the residual/prediction plot in figure 3 suggests a misspecified functional form. That is, the double-log functional form does not achieve complete linearity. In contrast, figure 4 displays less of an inverted-U shape suggesting the double-

³⁷ The plots in all of the figures in this chapter were produced using *SPSSX* software which were uniformly resized in *WORDPERFECT 5.1*.

³⁸ This analysis relies heavily on Neter, Wasserman, and Kutner (1989).

³⁹ nR^2 for each of models 1 and 2 are 42.66 and 44.43, respectively. The p-values for each of these χ^2 distributions with 2 degrees of freedom are .54513 E-9 and .22498 E-9, respectively. Further, the 95% critical value is 5.9915.

log, shifted distance model achieves greater linearity.⁴⁰

Third, as noted by Duffield (1988), the double-log model tends to overpredict per capita trips at short and long distances and underpredicts at intermediate distances. This can be readily seen in figure 5 which is a standardized scatter plot of the residuals resulting from model 1 and the natural-log of average round-trip distance.⁴¹ In contrast, the shape of a similar plot for model 2 in figure 6 suggests that model 2 mitigates these prediction errors although this model continues to overpredict at short distances and underpredict at intermediate distances.

⁴⁰ The standard for linearity used in this analysis is that a completely random distribution of the points on the standardized residuals/prediction plot means the function is linear.

⁴¹ Points on figure 5 and 6 falling below zero-valued residuals not counter weighted by points above zero-valued residuals suggest net over-prediction. The opposite is true for under-prediction.

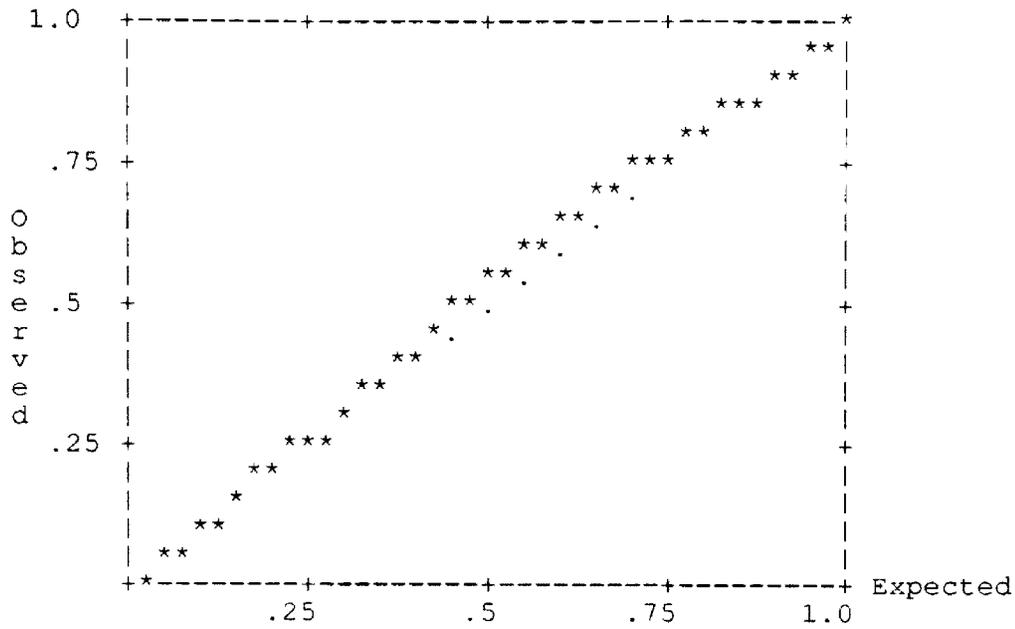


Figure 1.-- Normal Probability (P-P) Plot of Standardized Residuals for Model 1.

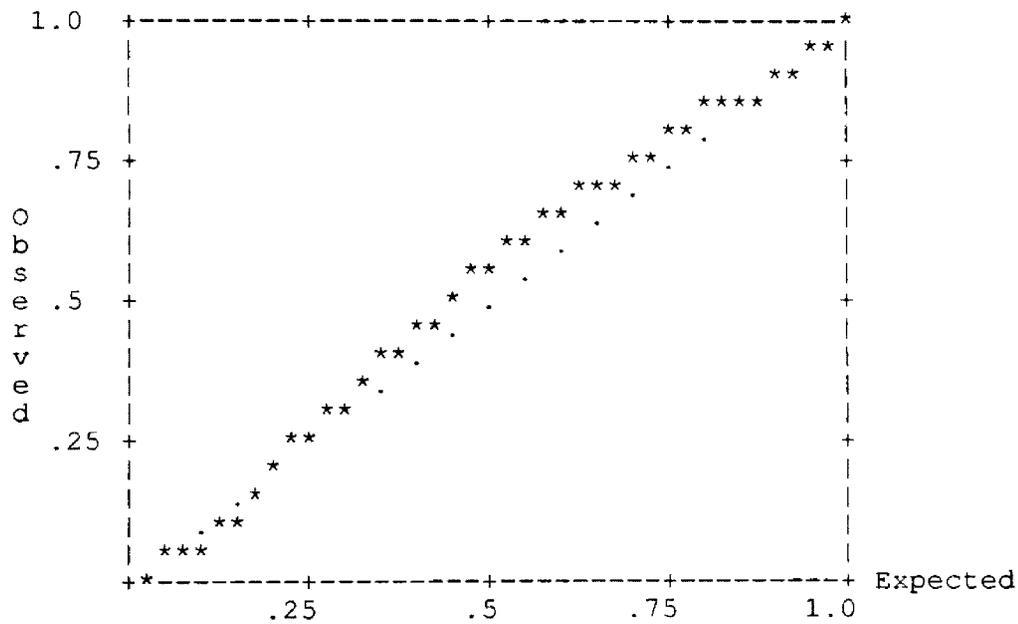


Figure 2.-- Normal Probability (P-P) Plot of Standardized Residuals for Model 2.

Finally, in each of figures 3 and 4 there appear to be several observations lying outward from the general cluster of points in each plot. Thus, a formal analysis of outliers was conducted using DFFITS and DFBETAS measures on models 1 and 2. The DFFITS measure revealed several observations affecting the fit of the models for shorter distance (2 to 30 miles from various sites). Similarly, most of the observations influencing the constant and slope parameters in these models, as measured by DFBETAS, were observations with one-way distances of 2 to 50 miles. Most importantly, however, all three measures (DFFITS and the two DFBETAS measures) identify trips taken from Alaska as observations influencing the fit and constant and slope parameters of the two models. These 20 observations represent all of the trips made from Alaska in the 741 observation aggregated sample and are the only observations which were aggregated using a method different from that summarized in Chapter 3. The exception is due to a physical gap in the defined market between Montana and Alaska, namely Canada. Recall that when trips from states not contiguous with Montana in which there were no trips made to a site from states between the state of origin, these trips carried the population of the intervening states. Thus, since Canada was excluded from the non-residential market, trips from Alaska carry only the population of Alaska, whereas trips from origins of similar distances to the east coast

carry greater populations. Another reason for the Alaska trips to stand out is that distances traveled by recreationists from Alaska were all corrected to a uniform distance.⁴² Due to this anomaly in application of the aggregation rules presented in chapter 3, an effort was made to model trips from Alaska with a dummy variable. The results of this effort are summarized in Appendix C.

Although the methods used to model trips from Alaska improves the fit and satisfaction of normality in models 1 and 2, it is not known with certainty that adding the interactive dummy variable is consistent with the general heteroscedasticity prevalent in models 1 and 2 and 1a and 2a (Appendix C). Thus, correction for the anomaly in the aggregation process for trips originating in Alaska has not been fully explored. Therefore, the adjustment to the bivariate models for trips originating in Alaska is omitted from the primary portion of this study. However, for the readers information, the results of modeling trips from Alaska as noted above in the model proving best among the Box-Cox models estimated (model 5) is presented in Appendix C.

⁴² Distances traveled from Alaska to sites in Montana were among those found to be in error as noted by Duffield (1988a). These distances were corrected using the distance from Anchorage, Alaska to Great Falls, Montana, via the ALCAN highway.

General Model Specifications for the Box-Cox Regressions

In the next step of the analysis, 8 alternative functional forms were specified based on the general specification of equation (2). Table 5 summarizes these plausible general forms and the different restrictions on λ_v

Table 5.-- Plausible Forms of Equation (2) Other than the Double-log, Semi-log and Linear forms.⁴³

<u>Plausible Forms</u>	<u>Restrictions:⁴⁴</u>	<u>λ_v</u>	<u>λ_D</u>
(a) $\frac{V_{ij}^{\lambda_v}-1}{\lambda_v} = \beta_0 - \beta_1 \frac{D_{ij}^{\lambda_D}-1}{\lambda_D} + e$		N/R	N/R
(b) $\ln V = \beta_0 - \beta_1 \frac{D_{ij}^{\lambda_D}-1}{\lambda_D} + e$		0	N/R
(c) $\frac{V_{ij}^{\lambda_v}-1}{\lambda_v} = \beta_0 - \beta_1 \ln D_{ij} + e$		N/R	0
(d) $V_{ij} = \beta_0 - \beta_1 \frac{D_{ij}^{\lambda_D}-1}{\lambda_D} + e$		1	N/R
(e) $\frac{V_{ij}^{\lambda_v}-1}{\lambda_v} = \beta_0 - \beta_1 D_{ij} + e$		N/R	1
(f) $\ln V_{ij} = \beta_0 - \beta_1 \ln D_{ij} + e$		0	0
(g) $\ln V_{ij} = \beta_0 - \beta_1 D_{ij} + e$		0	1
(h) $V_{ij} = \beta_0 - \beta_1 D_{ij} + e$		1	1

⁴³ The error term in all plausible forms estimated was assumed to enter additively. Thus, the functional form is limited to intrinsically linear forms.

⁴⁴ Restrictions of 0 and 1 refer to a natural-log transformation and a linear restriction, respectively. N/R refers to an estimated Box-Cox transformation parameter.

and λ_D required to obtain these forms. As noted in chapter 1, the primary theoretical constraint on equation (2) is that the signs and sizes of the estimated parameters β_0 , β_1 , λ_V , and λ_D result in a function in which per capita trips are inversely related to average round-trip distance traveled (the price proxy) from origin i to site j . With one exception, this constraint is satisfied in all of the forms listed in table 5 when β_1 is less than zero. This exception occurs in plausible form (b) in which β_1 must be positive if λ_D is negative. An additional constraint for the models to be considered demand functions is that the estimated parameters in equation (2) must yield a positive intercept or constant term. As one may note this term will not always be β_0 due to the specification of the Box-Cox transformation. For instance, consider plausible form (a) (table 5) is expressed in its reduced form as shown in equation (7):

$$(7) \quad V = \lambda_V \left[\beta_0 - \frac{\beta_1}{\lambda_D} + 1 \right]^{\frac{1}{\lambda_V}} + \frac{\lambda_V \beta_1 D^{\frac{\lambda_D}{\lambda_V}}}{\lambda_D}$$

If the absolute value of $\beta_1/\lambda_D > \beta_0 + 1$ and λ_V and β_0 are positive, the intercept term in plausible form (a) will be negative. The combinations of possible values and signs of β_0 , β_1 , λ_V , and λ_D contributing to the sign of the intercept term in each of plausible forms (a) through (e) (table 5)

are too numerous to analyze and list here. Thus, the intercept terms in each of the estimated forms presented below are independently assessed for theoretical acceptability.

Form Analysis of the Bivariate First-Stage TCM Demand Function

This section presents the estimation and model discrimination performed on the first-stage TCM demand functions estimated using the plausible forms summarized in table 5. First, each of the models estimated in this section are generally reviewed. This includes a summary of models' adherence to the theoretical constraints, an overview of the fit of models, and review of the significance of the Box-Cox transformation parameters. Next, the results of discriminating these models using the log-likelihood ratio test is summarized. This is followed by a review of the results of applying several diagnostic tests to model 5.

Estimated Models. Table 6 summarizes the results of estimating five forms of the bivariate first-stage TCM demand function listed in table 5.⁴⁵ Based on Duffield's (1988) findings specific estimates of the semi-log and linear forms (f) and (h) (table 5), respectively, were

⁴⁵ Collectively, the approach used to estimate the models in table 6 was designed to allow all of the plausible forms listed in table 5 to result.

omitted from this analysis.

At the outset only models 3.a, 3.b, 4.a, and 4.b were estimated. However, due to the failure of $\hat{\lambda}_v$ in model 3.a to be significantly different from zero at a 5 percent probability of a type I error, model 5, in which λ_v is restricted to a value of zero, was also estimated. Aside from the insignificance of $\hat{\lambda}_v$, the hypotheses that the remaining

Table 6.-- Estimated Box-Cox Bivariate First-Stage TCM Demand Functions.

Model	Maximum Likelihood Estimates					OLS	
	λ_v	λ_D	Log-Likelihood	β_0	β_1	\bar{R}^2	F
3. BCE*							
a. λ_v, λ_D	.0016 (.172)	.169 (6.78)	7,361.84	-4.30 (-11.57)	-.629 (-7.01)	.778	2,601
b. λ_v, λ_D	.038 (5.45)	.038 (5.45)	7,345.91	-2.92 (-11.32)	-.907 (-8.56)	.775	2,550
4. BC**							
a. λ_v	-.017 (-6.68)	1 (R***)	7,129.75	-13.59 (-16.20)	-.0045 (-7.35)	.592	1,076
b. λ_v	.030 (3.35)	0 (R)	7,337.43	-2.15 (-7.50)	-1.22 (10.25)	.768	2,451
5. λ_D (BCE)	0 (R)	.172 (6.87)	7,361.82	-4.33 (11.49)	-.630 (-6.99)	.778	2,599

* BCE stands for use of the Box-Cox extended estimation method.

** BC stands for use of the classical Box-Cox estimation method.

*** "R" signifies a restriction of the Box-Cox transformation parameter at the indicated value.

β_1 are the estimated coefficients for per capita trips transformed by the respective value of λ_1 . The p-value for λ_v in model 3.a is .4317. All other p-values are less than .001. The critical values for t and F at a 5 percent probability of a type I error are 1.6471 and 3.854, respectively. T-ratios were computed from the variance-covariance matrix resulting from the maximum likelihood estimation of each of models 3.a, 3.b, 4.a, and 4.b. Model 5 was estimated using an OLS/grid-search method. Adjusted- R^2 and F-statistics were computed using OLS regression with the appropriate transformations on the dependent and independent variables as listed in columns 2 and 3.

estimated coefficients in table 6, equal zero is rejected at a 5 percent probability of a type I error. Further, the hypothesis that no relationship exists between the independent and dependent variables are rejected at a 5 percent probability of a type I error for all of the models presented in the same table.

The signs on β_D in each of the models listed in table 6 are theoretically correct. With the exception of model 3.a, the signs and magnitudes of all estimated and restricted parameters in the models presented result in positive intercept terms in each model. λ_v in Model 3.a is not significantly different from zero at the 95 percent confidence interval thereby yielding the intercept term in model 3.a equal to zero.

Model Discrimination. Discrimination of the models presented in table 6 was performed using a series of likelihood ratio tests, as summarized in Chapter 2. There are several combinations of models in table 6 that satisfy the restriction that only models that can be nested in their non-restricted counterparts may be discriminated with this counterpart using a likelihood ratio test. However, based on the likelihood principle and the methods used to estimate the models in table 6 (see table 6 and preamble), only four combinations of models need to be tested. These combinations consist of models 3.b, 4.a, 4.b, and 5 as restricted models of model 3.a. Likelihood-ratio tests for

these combinations is summarized in table 7. The null hypothesis in this table is that the restricted models listed in column one are similar to the general model listed in column two. The alternative hypothesis is that the restricted and general models are dissimilar.

Table 7.-- Likelihood Ratio Tests of Bivariate Models with Model 3.a. as the General, Unrestricted Case

Restricted Model	General Model	Degrees of Freedom	Likelihood Ratio (χ^2)	P-Value
5.	3.a.	1	.04	.841
3.b.	3.a.	1	31.86	<.0001
4.b.	3.a.	1	48.82	<.0001
4.a.	3.a.	1	464.18	.000
1.	3.a.	2	61.00	<.0001

It can be concluded from this analysis that only model 5, of the models presented in table 6, is not significantly different from the most general model at a 5 percent probability of a type I error (the critical χ^2 value with one degree of freedom is 3.8415). Thus, according to the maximum likelihood principle, model 5 consists of those parameters which, among the parameters estimated for the remaining models in table 7, best describe the population of

per capita demand for fishing trips during 1985. It can be further concluded that model 5 is statistically superior to the remaining models in table 7.

Since the double-log model, presented in table 4, can be nested in model 3.a, a likelihood ratio test was performed with the same null hypothesis as that established for the remaining tests summarized in table 7. The results of this test, summarized in table 7, suggest that the null hypothesis could be rejected at a 5 percent probability of a type I error (the critical χ^2 value with two degrees of freedom is 5.9915).

Diagnostic Tests on Model 5. Figure 7 is the normal probability (P-P) plot for model 5. As can be seen, model 5 produces fairly normal residuals. In comparison with models 1 and 2, model 5 appears to produce slightly more normally distributed residuals, although they are slightly skewed left. Figure 8 is the standardized residual plot for model 5. In comparison with the similar plots for models 1 and 2 (figures 3 and 4), model 5 appears to result in a greater distribution of the residuals around the center of the plot than do models 1 and 2, thereby suggesting less heteroscedasticity of the residuals produced by model 5. It is noted however, that the results of White's test for homoscedasticity suggests the residuals in model 5 are

heteroscedastic.⁴⁶

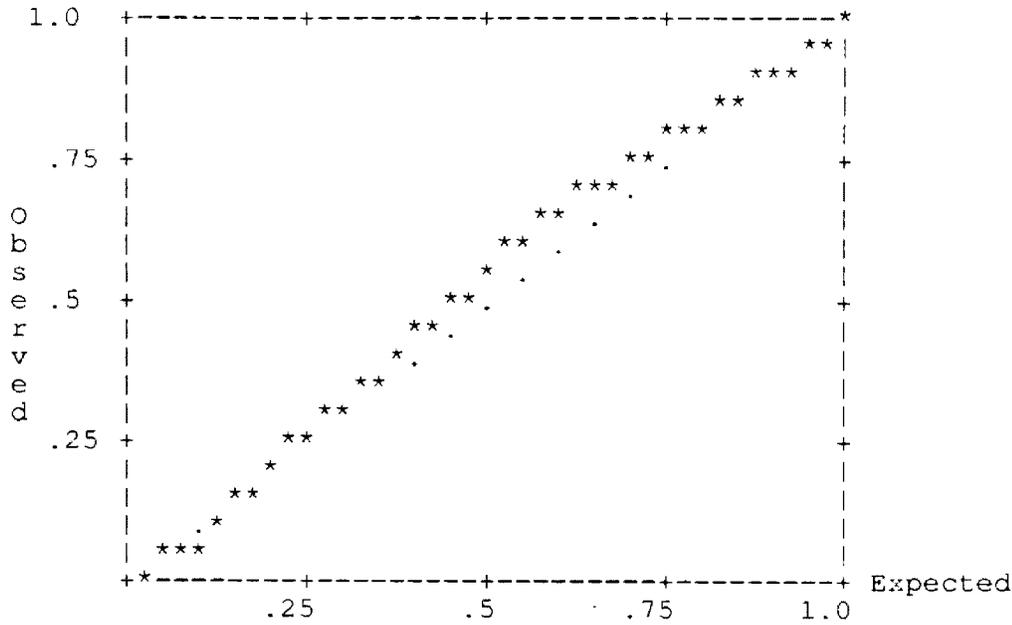


Figure 7.-- Normal Probability (P-P) Plot of the Standardized Residuals for Model 5, Table 6.

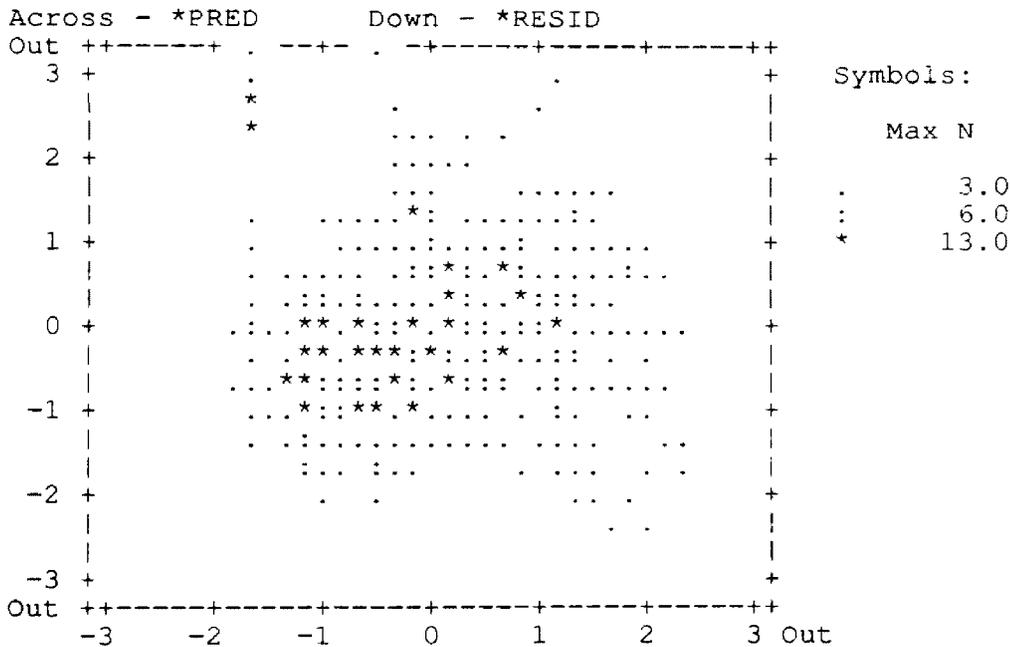


Figure 8. Standardized Scatterplot of predicted Per Capita Trips and Residuals for Model 5, Table 6.

⁴⁶ nR^2 for model 5 is 52.74 with a p-value of this χ^2 distribution with 2 degrees of freedom less than .0001. Further, the 95% critical value is 5.9915.

Figure 9 is a standardized scatterplot of the residuals produced by model 5 and average round-trip distance transformed by the Box-Cox transformation parameter for the same model. Although the inverted-U shape present in similar plots for models 1 and 2 appears less evident for model 5 than it does for model 1, the shape continues to suggest the functional form may not be correctly specified.⁴⁷ Comparison of figures 6 and 9 also suggests further resolution of the over/under trip prediction problem encountered in model 1. This is evident by the wider dispersement of points around the center of figure 9 verses the dispersement of points in figure 6. The predictive performance of models 1, 2, and 5 are further discussed below.

⁴⁷ As noted in chapter 2, the trade-off between heteroscedasticity and functional form is omitted from this study. It may be the case that the inverted-U shape in figure 9 is partially due to the lack of correction for heteroscedasticity.

models considered in this paper. Since these two models are non-nested, a J-test was used for discrimination. The results of the hypothesized models are summarized in table 8.

Table 8. -- Results of J-Test for Discrimination Between Models 2 and 5.

Model (Hypothesis)	β_0	β_1^*	β_2^{**}	\bar{R}^2	F
2. (Null)	4.724	-3.508	-.541	.785	1,352
(t-ratio)	(4.793)	(-4.932)	(-1.728)		
(p-value)	(<.0001)	(<.0001)	(.042)		
5. (Alt)	2.308	.341	1.523	.785	1,352
	(1.703)	(1.728)	(4.932)		
	(.044)	(.042)	(<.0001)		

* β_1 are the estimated coefficients of average round-trip distance plus 64 and average round-trip distance subjected to the Box-Cox transformation with $\lambda=.172$.

** β_2 are the estimated coefficients of the predicted values resulting from estimating models 5 and 2 with OLS, respectively.

The sample size is 741 and $t^*=1.647$ and $F^*=3.007$ at the 5 percent probability of a type I error.

In terms of the method summarized in Appendix B, models 2 and 5 are considered the null and alternate

hypothesis models (for general specifications see equations (B-1) and (B-2)), respectively. The estimates of models 2 and 5 parallel the general specifications of equations (B-4) and (B-3). The hypotheses that the estimated coefficients of β_2 in both models 2 and 5 (table 8) are different from zero are rejected for both hypothesized models at the 5 percent probability of a type I error. The decision criteria summarized in table 10 (Appendix B) suggests that neither model 2 or 5 is acceptable or sufficiently explains the variation in the natural-log of per capita trips when compared to the other. Unfortunately, however, this test provided inconclusive results. That is, there is insufficient evidence at a 5 percent probability of a type I error to reject the null hypotheses that each models' predicted values of the natural-log of per capita trips adds to the fit of the rival model. This result is one of the drawbacks of the J-test. Specifically, the J-test does not provide conclusive evidence that either model 2 or 5 is statistically superior over the other. This is caused by insufficient information regarding the comparison of the conditional distributions of the dependent variables in the two models (see Maddala (1992)).⁴⁸

⁴⁸ Maddala (1992) suggests supplementing the J-test with an analysis of variance in which the models are combined and the coefficients on the differing variables are tested for significance. However, due to strong colinearity between the independent variables in models 2 and 5, such analysis was unsuccessful. Additionally, Maddala cites Mizon and Richard (1986) for a means to overcome the lack of

An alternative means of discriminating models 2 and 5 is to compare their adjusted- R^2 s. These values are .784 and .778 for models 2 and 5, respectively. The slight difference of .006 between these values (*i.e.*, model 2 explains less than 1 percent more of the variation in the dependent variable than model 5) suggests the two models may be considered statistically comparable.

Based on the analysis in this chapter, although models 2 and 5 may be considered statistically comparable in terms of slight differences in adjusted- R^2 , it may also be concluded that model 5 provides an alternative means of adjusting for prediction errors primarily due to measurement error at shorter distances (see Duffield (1988)). A comparison of figures 11 and 10 (see subsection discussing scatter plots below) shows that both models 2 and 5 enhance the linear price/quantity relationship of model 1. However, unlike model 2, model 5 achieves this enhancement without shifting the entire demand function to the right as noted by Duffield (1988). Furthermore, model 5 suggests a degree of nonlinearity not specified in model 2. It is also important to note that both models suggest that the relationship of per-capita trip demand for varying distances from a site is not a one-to-one relationship as suggested for a multivariate specification of model 2 by Duffield (1988).

information on the conditional distributions in the null and alternate hypothesis models.

It is also noted that the theoretical basis for a multivariate specification of model 2 suggested by Duffield has greater merit than a possible theoretical basis for model 5. Specifically, the constant in model 2 may be a measure of the fixed costs associated with making a trip. A theoretical basis for model 5 might be that it describes the degree of marginal diminishing returns for per-capita trips as distance increases more accurately than model 1. However, at least, models 2 and 5 appear to capture characteristics of per-capita trips demand not captured in model 1.

Overview of Models 1, 2, and 5

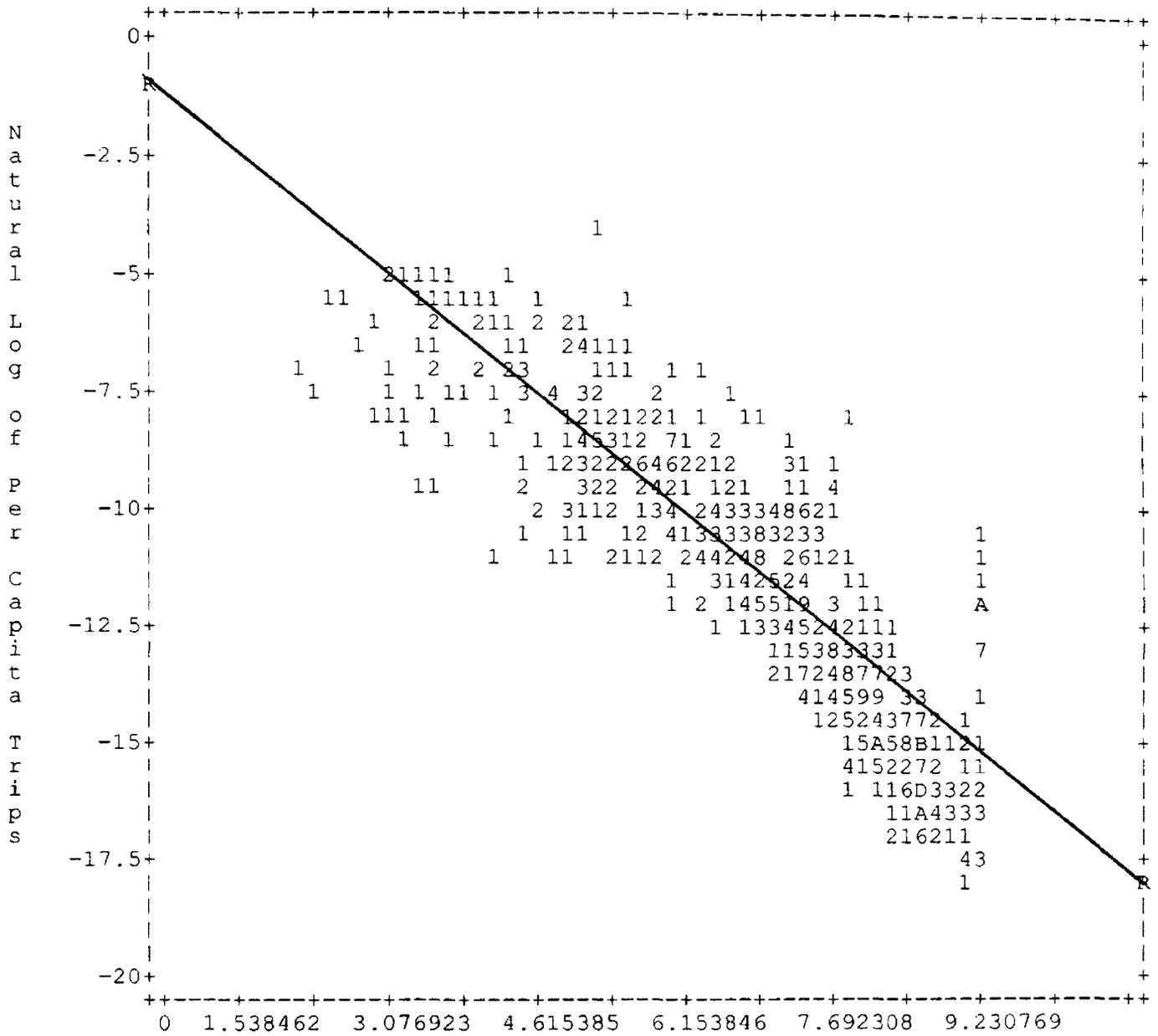
The balance of this chapter provides an overview comparison of the bench-mark models (1 and 2) and model 5. Included in this section is a comparison of models 1, 2, and 5 in terms of scatterplots and line-plots of the three models and their predictive performance across the sample. This section is closed with a summary of the elasticities of each model.⁴⁹

Scatter and Line Plots. Figures 10, 11, and 12 are scatter plots of the natural log of per-capita trips versus

⁴⁹ The model estimation and discrimination analysis presented above is considered sufficient for a conclusion that models 2 and 5 are statistically similar in describing the variation in the natural-log of per-capita trip demand. The plots, predictive performance and elasticities of each model are provided for illustrative purposes only.

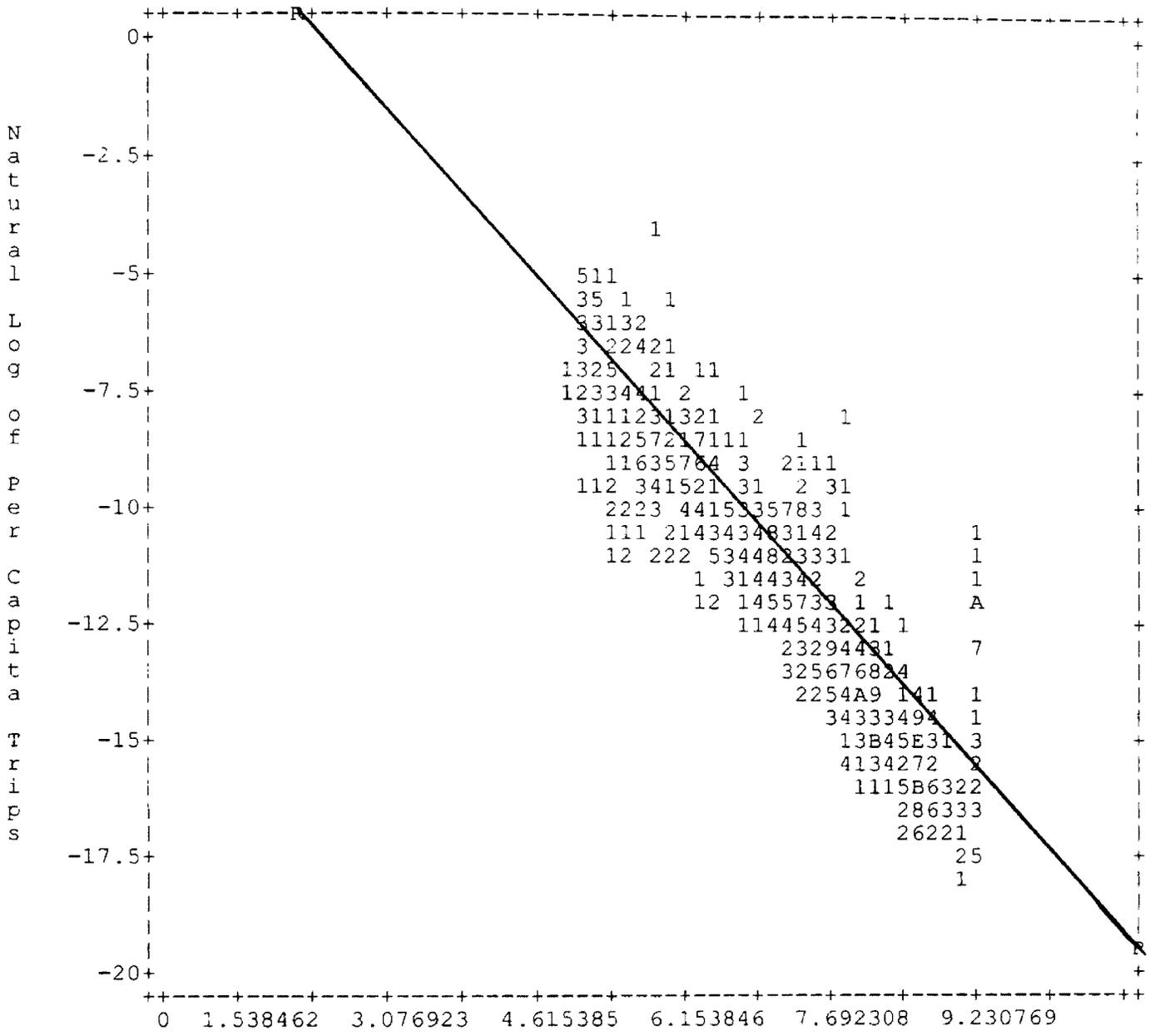
the transformations of average round-trip distance applied in each of models 1, 2, and 5, respectively.⁵⁰

⁵⁰ When comparing figures 10, 11, and 12 it should be noted that the horizontal scale in figure 12 is more compressed than that of figures 10 and 11. This was done to maximize the scatter of observations across the plot and to include the mean of the independent variable in figure 12.



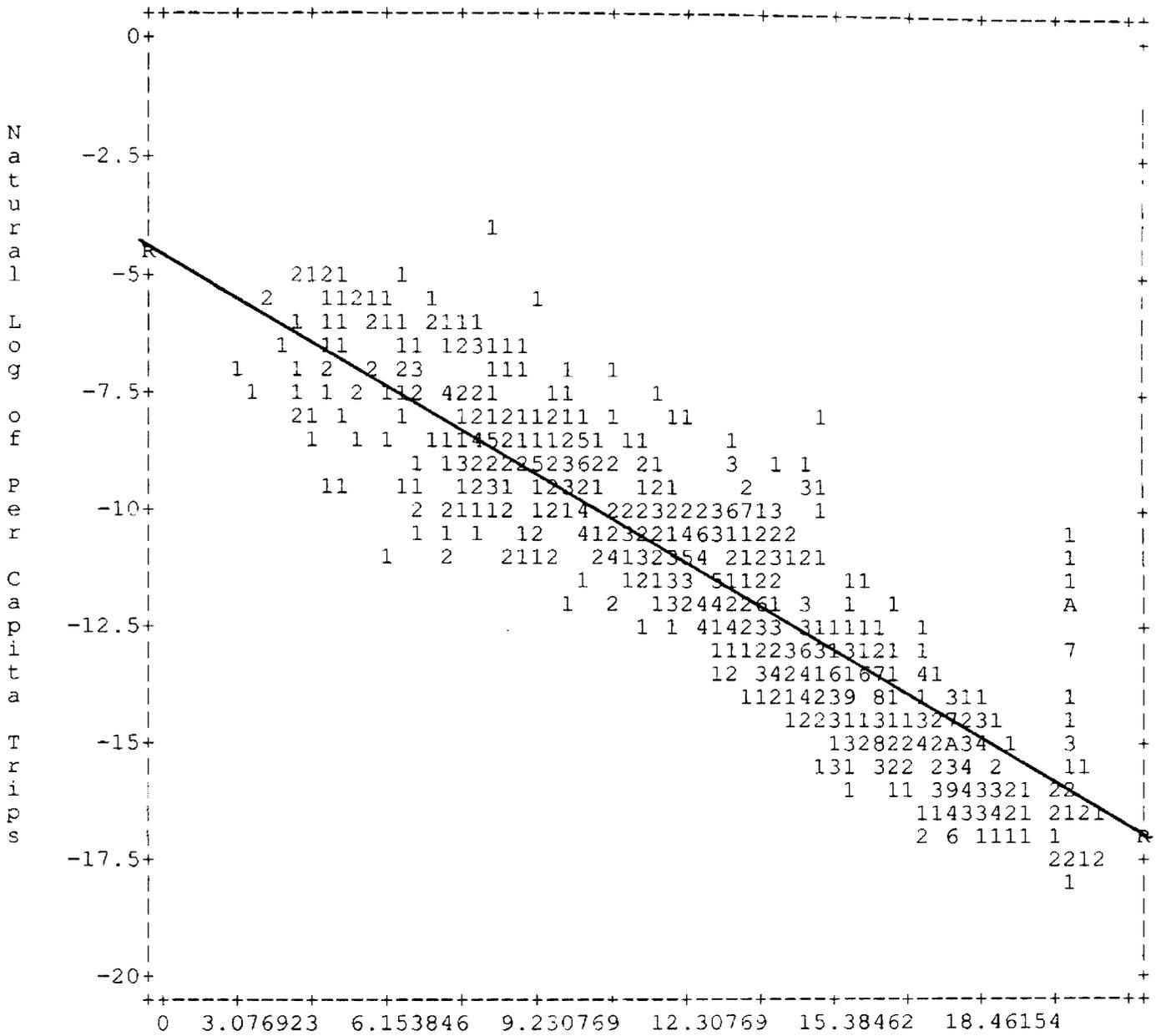
Natural Log of Average Round-Trip Distance

Figure 10.-- Scatterplot of the natural-log of per-capita trips with the natural-log of average round-trip distance. Points with numeric values (1-9) represent the same number of observations. Alphabetic letters, A through Z, represent clusters of 10 through 35 points, respectively.



Natural Log of Average Round-Trip Distance plus 64

Figure 11.-- Scatterplot of the natural-log of per-capita trips with the natural-log of average round-trip distance plus 64. Points with numeric values (1-9) represent the same number of observations. Alphabetic letters, A through Z, represent clusters of 10 through 35 points, respectively.



Average Round-Trip Distance Subject to a Box-Cox Transformation of .172

Figure 12.-- Scatterplot of the natural-log of per-capita trips with average round-trip distance raised to the .172 power. Points with numeric values (1-9) represent the same number of observations. Alphabetic letters, A through Z, represent clusters of 10 through 35 points, respectively.

As one can see in these plots, models 2 and 5 result in three improvements over model 1. These include, a tighter cluster of observations around the regression line, fewer outlying observations at shorter distances (this is confirmed by comparison of the DFFITS and DFBETAS for these two models), and greater linearity in the price-quantity relationship. The first two of these improvements can be considered as improvements of the intended use of the double-log model. Specifically, the double-log model was selected by Duffield *et al.* (1987), in part, as a means of mitigating heteroscedasticity which results from compressing the range of observations by means of the double-log form (see, e.g. Gujarati (1988)).

If one were to plot models 1, 2 and 5 on one graph with per-capita trips on the y-axis, using a continuum of data points across the range of average round-trip distance, models 2 and 5 would appear most similar.⁵¹ Although the plots for all three models are non-linear, model 2 shows the least amount of nonlinearity and model 1 the most.⁵² Further, at round-trip distances less than 48 and 64 miles, the plots of models 2 and 5 are below model 1, respectively. These plots remain above model 1 up to distances of about

⁵¹ This plot was omitted from the text due to illegibility.

⁵² The functions used for these plots were roughly:
 Model 1: $V = .357 D^{-1.71}$
 Model 2: $V = 21.977 (D + 64)^{-2.28}$
 Model 5: $V = \exp(-.667 - 3.663 D^{.172})$

1,200 and 1,500 miles for each of models 2 and 5, respectively. Thus, models 2 and 5 dampen the curvature of model 1 (particularly at shorter distances) providing greater linearity in the price/quantity relationship in the transformed form of these models. This illustrates models' 2 and 5 mitigation of the over/under trip prediction problem with model 1.

Predictive Power. The resulting predictive power of models 1, 2, and 5 provide additional information regarding the performance of these models. As expected, model 1 overpredicts the 5,029 aggregated trips in the sample by 2,820 or 56.1 percent. However, model 2 underpredicts trips by 287 or 5.7 percent and model 5 overpredicts trips by 72 or 1.4 percent.

Elasticities. As an aside and purely for illustrative purposes, table 9 summarizes the elasticities evaluated at the mean of average round-trip distance (1,119.5) for each of models 1, 2, and 5.

Table 9.-- Estimated Own-Price Elasticities of Demand
for Models 1, 2, and 5

Model	Estimated Own-Price Elasticity of Demand
1.	-1.718
2.	-2.281
5.	-2.115

The own price elasticity for model 5 is computed by $\beta_1(D^{\lambda D})$.

CHAPTER 5
CONCLUSIONS

This chapter provides a summary of this study, its limitations and suggestions for further research. Each are addressed in turn.

Summary

The purpose of this study has been to examine the form of the first-stage demand function for the cold-water stream fisheries in Montana during the 1985 season. Chapter one provided an overview of the travel cost methodology. This included a summary of the pertinent economic theory underlying the TCM methodology, simplifying assumptions, definitions of the price and quantity variables, data requirements, and examples of TCM applications. Additionally, the literature regarding determination of the functional form of the TCM demand function supporting the use of statistical inference to determine the correct form was reviewed.

Chapter two outlined the approach to model specification used in this paper. It was concluded that the best approach was to test down from a general to a specific functional form. However, due to the information regarding the form of the first-stage TCM demand function as

investigated by Duffield et al. (1987) and Duffield (1988), an hybrid of a testing down and up (AER/TTT) approaches was concluded as the best approach to use in this study. That is, the study relied on the findings of the two prior studies noted to limit the field of plausible forms of the demand function.

Chapter two also provided an overview of the Box-Cox methodology. Included in this review were four possible approaches to estimating a Box-Cox regression function, of which the full maximum likelihood method was deemed best due to the ability to estimate the variance-covariance matrix encompassing all estimated parameters. Next, two methods of model discrimination were reviewed, including the likelihood ratio test, which was used to discriminate five plausible Box-Cox regression models and the double-log model initially proposed for the data base by Duffield et al. (1987). Davidson and Mackinnon's J-test for discriminating non-nested models was also reviewed. Finally, in accordance with the TTT approach to model specification, several diagnostic tests on the assumptions of OLS used in this study were reviewed.

The results of the econometric analysis performed in this study were reported in chapter four. As a bench-mark for the analysis in this study, two bivariate models were estimated and subjected to several diagnostic tests regarding certain assumptions of OLS. These models followed

the principles used by Duffield et al. (1987) (a double-log model) and Duffield (1988) (a double-log, with shifted distance model) to estimate multivariate counterparts to these models.

Based on the results of prior studies, five Box-Cox regression models were estimated such that eight plausible forms could result. These models were tested for their adherence to generalities regarding *a priori* expectations of the signs and magnitudes of the estimated regression coefficients and Box-Cox transformation parameters. Only the most general form failed to meet these requirements. Using a series of likelihood ratio tests, it was found that a form consisting of the natural log of per-capita trips regressed on average round-trip distance raised to the .172 power (model 5) was the only form statistically similar to the most general form (model 3.a) of the TCM demand function investigated in this study. A likelihood ratio test comparing the most general form with Model 1, based on the findings of Duffield et al. (1987), was also conducted. This model was also dissimilar to the most general form at the 95 percent confidence interval, thus eliminating model 1 from the field of statistically acceptable models.

An attempt to discriminate Model 5 and the bivariate specification of the double-log, shifted distance model with a shift factor of 64 (model 2) using a J-test provided inconclusive results. Thus, these two models were compared

based on their adjusted- R^2 statistics. Because the values of these statistics were found to be close (.784 and .778 for models 2 and 5, respectively), it was concluded that the two models are statistically comparable. Moreover, model 5 was found to provide more information regarding the characteristics of the per-capita demand for stream fishing in Montana during 1985 than the does the double-log model.

Comparisons of the scatter plots of models 1, 2, and 5 revealed that models 2 and 5 result in five improvements over model 1. These include greater linearity in the price/quantity relationship, a tighter cluster of data points around the regression lines, further mitigation of heteroscedasticity (compare figures 3, 4, and 8), mitigation of outliers for shorter distances (possibly related to the mitigation of heteroscedasticity), and enhanced predictive power. As regards the latter of these improvements models 2 and 5 predicted trips within 5.7 and 1.4 percent of observed trips, whereas model 1 overpredicted trips by 56.1 percent.

Limitations and Suggestions for Further Research

This section highlights four limitations to this study, three of which may merit further research. Each of these issues include heteroscedasticity of the residuals, a truncated distribution of the residuals in model 5, errors in measurement of round-trip distance, and omitted variable bias. An additional suggestion for further research includes estimation of the demand function to account for

possible segments in the market by use of a spline function. Each of these issues are summarized in turn.

Heteroscedasticity. As previously noted, no attempts were made to measure or correct heteroscedasticity in any of the models presented in this study. Moreover, while models 2 and 5 appear to mitigate heteroscedasticity, as is evident in the plots of the models' residuals against their predicted values, White's test suggests the residuals are not homoscedastic. Thus, the estimated coefficients in these model are neither best or linear estimates, nor can they be considered efficient (see Kmenta (1986)). A possible solution to this limitation may be determined by reestimating the models presented in this paper by specifying likelihood functions capable of simultaneously estimating the Box-Cox parameters for functional form, the shift parameter (model 2), and the form of heteroscedasticity ⁵³.

Truncated Distribution of the Residuals in Model 5.

Smith (1975) notes that use of the Box-Cox transformation on the dependent variable does not allow for negative values in this variable resulting in a non-normal distribution of the residuals. As a consequence, he notes that estimates of the parameters in a Box-Cox regression model using a likelihood function specified with a normal distribution of the

⁵³ See footnote 19, Chapter 2 for pertinent citations of literature addressing this issue.

residuals results in approximations of these parameters. The likelihood function specified for estimates of the Box-Cox models presented in table 6 assumed the residuals were normally distributed. Therefore, models 3.a, 3.b, 4.a, 4.b, and 5 are considered only approximations of the models that would be estimated given the same restrictions (or latitudes) capsulated in each of these models when the truncated distribution issue is incorporated in the analysis. No further analysis on this issue was conducted.

Errors in Measurement of the Distance Variable. As noted in chapter 3, several of the reported distances were found to be in error and were replaced in the data base with map distances. These distances were measured by assuming trips originated from the population centers of each county or state. Duffield (1988) notes this is a result of the small sample size in the data set. He also notes that aggregation of trips with disparate distances within origin zones to a common site add to the effect of the distance measurement error occurring largely in the range of shorter distances. He suggests that shifting distance by a constant appears to mitigate the effects of the measurement error at shorter distances by shifting the demand function away from these shorter distance.⁵⁴

⁵⁴ Duffield (1992) further refined the shifted distance, double-log model by transforming average-round trip distance for resident and non-resident anglers by the total cost functions estimated for each class in Duffield et al (1987).

Conversely, model 5 adjusts for the nonlinearity not captured by model 1, which can be seen by comparison of figures 10 and 12. However, an explicit examination of an error-in-variables model or other means of correction for errors in measurement within the framework of the Box-Cox regression models were not examined in this study.⁵⁵ Thus, further analysis of this sort may provide a further enhancement to the analysis.

Omitted Variable Bias. Since the scope of this study was limited to determining the functional form of the bivariate TCM demand function, the models estimated in this study are prone to omitted variable bias. Thus, further examination of the functional form of the demand function for stream fishing in Montana might include the addition of non-price variables. These variables might include site attributes, a measure of substitutes, and several socio-economic / demographic variables designed to approximate tastes and preferences. Failure to include these variables in the model would result in biased parameter estimates and thereby biased estimates of consumer

⁵⁵ A general Box-Cox model similar to model 3.a was estimated for model 2. However, it was found that the Box-Cox transformation on the dependent and independent variables both converged to zero, yielding the initial double-log form as that which maximized the likelihood function.

surplus (see, e.g., Dwyer et al. (1977) and Walsh (1986)).⁵⁶ These non-price variables are intended to explain potential shifts in demand. These variables can be used to evaluate the effects on demand and net economic benefits of adding a new site to or deleting an existing site from a region, changing the attributes of one or more sites, or changing a site's access costs.

In this regard, the reader is advised to examine Duffield (1988) in which he lists three additional concerns or limitations to the stream fisheries data set. Other than the measurement error issue, these issues include misspecification of the substitute variable, use of socio-economic/demographic variables limited to the representation of the angler's of an origin zone.

Market Segmentation. Another suggestion for further research regarding functional form would be to estimate per-capita trip demand using a spline function (Greene (1990)). Duffield et al. (1987) found that their double-log model overpredicted trips for short and longer distances and underpredicted trips for medium range distances. Also, analysis of the residuals versus average round-trip distance plots in figures 6 and 9 for models 2 and 5, respectively,

⁵⁶ Such bias consumer surplus is expected to occur if the bivariate demand function presented in this paper is used to compute consumer surplus values. Readers are reminded that the purpose of this paper is to examine the functional form of the bivariate model. Thus, the non-price variables listed in this section are intentionally omitted from the specification presented in chapter four.

reveals the this poor prediction problem was not entirely addressed. Simply stated, a spline function could be used to segment the market by increments in distance, thereby testing whether different distance intervals represent different market segments. In a broader scope, the dummy variables delimiting market segments could be estimated with a log-likelihood function, as suggested by Greene (1990), which also accounts for functional form and heteroscedastic effects.

LIST OF REFERENCES

- Bowes, Michael D. and John B. Loomis. 1980. "A Note on the Use of Travel Cost Models with Unequal Zonal Populations." *Land Economics*. 56 (4). pp. 465-470.
- Box, G. E. P. and D. R. Cox. 1964 "An Analysis of Transformations." *Journal of The Royal Statistical Association: Series B*, 26 pp. 211-243.
- Brown, William G., Colin Sorhus, Bih-lian Chou-Yang, and Jack A. Richards. 1983. "Using Individual Observations to Estimate Recreation Demand Functions: A Caution" *American Journal of Agricultural Economics*. Feb. pp. 154-157.
- Burt, Oscar R. and Durward Brewer. 1971. "Estimation of Net Social Benefits From Outdoor Recreation." *Econometrica* 39 pp. 813-827.
- Cesario, Frank J. 1976. "Value of Time in Recreation Benefit Studies" *Land Economics*. 52 (1) pp. 32-41.
- Duffield, John. 1988. "Alternative Specifications of TCM Demand Functions for Montana Cold Water Stream Fishing." Mimeo
- Duffield, John W. 1992. *Baseline Recreation Value of The Upper Clark Fork River Basin*. Report for Montana Department of Fish, Wildlife and Parks.
- Duffield, John, John Lommis, and Rob Brooks. 1987. *The Net Economic Value of Fishing in Montana*. Helena. Montana Department of Fish, Wildlife and Parks.
- Dwyer, John F., John R. Kelly, and Michael D. Bowes. 1977. *Improved Procedures for Valuation of the Contributions of Recreation to National Economic Development*. University of Illinois, Water Resources Center. pp. 79-132.
- Gaudry, Marc J. I. and Marcel G. Dagenais. 1979. "Heteroscedasticity and the Use of Box-Cox Transformations." *Economics Letters*. 2. pp. 225-229.

- Greene, William H. 1990 *Econometric Analysis*. MacMillan. Chicago.
- _____. 1991. *LIMDEP Version 6.0*. Econometric Software, Inc. Bellport, NY.
- Gujarati, Domodar N. 1988. *Basic Econometrics*. 2 ed. McGraw-Hill. New York.
- Harvey, A. C. 1990. *The Econometric Analysis of Time Series*. 2 ed. MIT Press. Cambridge.
- Haspel, Abraham E. and F. Reed Johnson. 1982. "Multiple Destination Trip Bias in Recreation Benefit Estimation." *Land Economics*. 58 (3). pp. 364-372.
- Henderson, James M. and Richard E. Quandt. 1980. *Microeconomic Theory A Mathematical Approach*. 3 ed. McGraw-Hill. New York.
- Just, Richard E., Darrell L. Hueth, and Andrew Schmitz. 1982. *Applied Welfare Economics and Public Policy*. Prentice-Hall, Inc. New Jersey.
- Kennedy, Peter. 1992. *A Guide to Econometrics*. 3 ed. MIT Press. Cambridge.
- Kmenta, Jan. 1986. *Elements of Econometrics*. 2 ed. MacMillan. New York.
- Koutsoyiannis, A. 1977. *Theory of Econometrics*. 2 ed. Macmillan. London.
- Lahiri, Kajal and Daniel Egy. 1981. "Joint Estimation and Testing for Functional Form and Heteroscedasticity." *Journal of Econometrics*. 15. pp. 299-307.
- Maddala, G. S. 1992. *Introduction to Econometrics*. 2 ed. MacMillan. New York.
- McConnell, Kenneth E. 1975. "Some Problems in Estimating the Demand for Outdoor Recreation." *American Journal of Agricultural Economics*. May. pp. 330 - 334.
- _____. 1985. "The Economics of Outdoor Recreation." *Handbook of Natural Resource and Energy Economics*. Vol. 2. A. V. Knesse and J. L. Sweeney eds. Elsevier Science. pp. 677-722.

- McConnell, Kenneth E. and Ivar Strand. 1981. "Measuring the Cost of Time in Recreation Demand Analysis: An Application to Sportfishing." *American Journal of Agricultural Economics*. (Feb). 153-156.
- McFarland, Robert C. 1985a. Memorandum to Bioeconomic Team Members, Re: Unit Codes. Montana Department of Fish, Wildlife and Parks.
- _____. 1985b. Memorandum to Bioeconomic Team Members, Re: Angler Data. Montana Department of Fish, Wildlife and Parks.
- _____. 1989. *Montana Statewide Angling Pressure Mail Survey 1982-1985*. Montana Department of Fish, Wildlife and Parks.
- Neter, John, William Wasserman, and Michael H. Kutner. 1989. *Applied Linear Regression Models*. 2ed. Irwin, Homewood, IL.
- Pesaran, M. H. 1974. "On The General Problem of Model Selection." *Review of Economic Studies*. 41 (April). 153-171.
- Pfaffenberger, Roger C. and James H. Patterson. 1987. *Statistical Methods for Business and Economics*. 3ed. Irwin. Homewood, IL.
- Quirk, James P. 1976. *Intermediate Microeconomics*. Science Research Associates. Chicago.
- Russell, R. Robert and Maurice Wilkinson. 1979. *Microeconomics A Synthesis of Modern and Neoclassical Theory*. John Wiley and Sons. New York.
- Savin, N. E. and Kenneth J. White 1978. "Estimation and Testing For Functional Form and Auto Correlation." *Journal of Economics*. 8 (1-12). pp.1-12.
- Seaks, Terry G. and Stephen K. Layson. 1983. "Box-Cox Estimation with Standard Econometric Problems." *The Review of Economics and Statistics*. 64 (Feb). pp. 160-164.
- Smith, V. Kerry. 1975. "Travel Cost Demand Models for Wilderness Recreation: A Problem of Non-Nested Hypotheses." *Land Economics*. 51 (2). pp. 104-111.

- Spitzer, John J. 1982a. "A Primer on Box-Cox Estimation." *The Review of Economics and Statistics*. 64 (May). pp. 307-313.
- Spitzer, John J. 1982b. "A Fast and Efficient Algorithm for the Estimation of Parameters in Models With the Box-and-Cox Transformation." *Journal of the American Statistical Association*. 77 (Dec). pp. 760-766.
- Spitzer, John J. 1984. "Variance Estimates in Models With the Box-Cox Transformation: Implications For Estimation and Hypothesis Testing." *The Review of Economics and Statistics*. 66. pp. 645-652.
- SPSS Inc. 1986. *SPSSx User's Guide*. 2ed, McGraw-Hill, New York.
- Strong, Elizabeth J. 1983. "A Note on the Functional Form of Travel Cost Models with Zones of Unequal Populations." *Land Economics*. 59 (3) pp. 342-349.
- U. S. Water Resources Council. 1983. *Economic and Environmental Principles and Guidelines for Water Related Land Resources*.
- U. S. Department of Interior. 1991. Natural Resource Damage Assessments. 43 CFR Part 11.
- Vaughn, William J., Clifford S. Russell, and Michael Hazilla. 1982. "A Note on the Use of Travel Cost Models With Unequal Zonal Populations: Comment." *Land Economics*. 58 (3). pp. 400-407.
- Walsh, Richard G. 1986. *Recreation Economic Decisions: Comparing Benefits and Costs*. Venture Publishing, Inc. State College, PA.
- Walsh, Richard G., Don M. Johnson, and John R. McKean. 1988. "Review of Outdoor Recreation Economic Demand Studies With Nonmarket Benefit Estimates, 1968-1988." Colorado Water Resources Research Institute. Fort Collins.
- Ward, Frank A. and John B. Loomis. 1986. "The Travel Cost Demand Model as an Environmental Policy Assessment Tool: A Review of Literature." *Western Journal of Agricultural Economics*. 11 (2). pp. 164-178.
- Ward, Kevin M. and John W. Duffield. 1992. *Natural Resource Damages: Law and Economics*. John Wiley and Sons, Inc. New York.

- Weisberg, Sanford. 1980. *Applied Linear Regression*. John Wiley and Sons. New York.
- White, Kenneth J., S. Donna Wong, Diana Whistler, and Shirley A. Haun. 1990. *SHAZAM User's Reference Manual Version 6.2*. McGraw-Hill, New York.
- Zarembka, Paul. 1968. "Functional Form in The Demand for Money." *Journal of the American Statistical Association*. 63 (June). pp. 502-511.
- Zarembka, Paul. 1974. "Transformations of Variables in Econometrics." *Frontiers in Econometrics*. Paul Zarembka, ed. Academic Press. New York pp. 81-104.
- Zeimer, Rod F. Wesley N. Musser, and R. Carter Hill. 1980. "Recreation Demand Equation: Functional Form and Consumer Surplus." *American Journal of Agricultural Economics*. Feb. pp. 136-140.

APPENDIX A

TECHNICAL DESCRIPTION OF THE TRAVEL-COST MODEL

As noted in chapter one, the zonal TCM consists of statistically estimating a demand function for use of a recreation site. Although the TCM has been justified in terms of household production, the following provides a justification in terms of utility maximization. McConnell (1985) provides the following general development of the first-stage TCM demand function based on the following quasi-concave ⁵⁷ utility function subject to income and time constraints:

$$(A-1) \quad \underset{x, z}{\text{Max}} \quad u(x, z) \quad \text{s.t.} \quad y = cx + pz, \quad T = h + x(t_1 + t_2)$$

Where:

x = Number of trips to a given site
 z = Hicksian composite commodity
 h = Time spent working
 t_1 = travel time per trip
 t_2 = time spent on site per trip
 T = total available time
 y = exogenous income
 c = out-of-pocket costs per trip
 p = Price of the composite commodity
 $y = y_0 + wh$
 w = the wage rate

This model assumes recreationists can trade between

⁵⁷ Henderson and Quandt (1980) show the utility function must be strictly quasi-concave in terms of the utility and price space.

work and leisure at a constant rate which means t_1 , t_2 , and h are all measured in the same units. When the income and time constraints are combined utility maximization becomes:

$$(A-2) \quad \text{Max}_{x, z} u(x, z) + [y^* - c^*x - pz]$$

Where

$$(A-3) \quad y^* = y + wT$$

$$(A-4) \quad c^* = w(t_1 + t_2) + c$$

With first order condition of:

$$(A-5) \quad \frac{\delta U}{\delta x} = \lambda c^* = \lambda (c + w(t_1 + t_2))$$

which results in demand given by:

$$(A-6) \quad x = f(c^*, p, y^*)$$

which reduces to:

$$(A-7) \quad x = f(c^*, y^*)$$

when p is assumed constant in the cross section, the time frame typically used for TCM demand function estimation. Thus, demand is a function of full income and full costs. Full costs generally constitute out-of-pocket costs and the opportunity cost of travel time. According to general demand theory these variable would be inversely related to the quantity of recreation trips demanded.

APPENDIX B
 OVERVIEW OF THE APPLICATION OF
 THE DAVIDSON AND MACKINNON J-TEST⁵⁸

The Davidson and MacKinnon J-test is a joint test of the predictive power of two rival, non-nested models. The following summarizes the methodology for a general case which may include Box-Cox transformations in the regression variables. The null and alternate hypothesis models are specified as in equations (B-1) and (B-2) in which the error term of each is assumed normally and independently distributed with zero mean and constant variance:

$$(B-1) \quad H_0: Y = \beta_0 + \sum_{k=1}^K \beta_k X_k + e_i$$

$$(B-2) \quad H_a: Y = \gamma_0 + \sum_{k=1}^K \gamma_k Z_k + e_i$$

The test is based on an artificial nesting of the two models and is performed in two stages. In the first stage, the null hypothesis model (equation (B-1)) is artificially nested in the alternate hypothesis model (equation (B-2)).

⁵⁸ This summary relies heavily on Maddala (1992).

This is done by specifying the predicted values of equation (B-1) resulting from OLS estimation as an additional variable in equation (B-2) to form equation (B-3).

$$(B-3) \quad H_a: Y = \gamma_0 + \sum_{k=1}^K \gamma_k Z_k + \alpha y_0 + e_i$$

y_0 represents the predicted values of the null hypothesis function and α is the estimated parameter. A secondary hypothesis test is then established in which $H_0: \alpha = 0$ and $H_a: \alpha \neq 0$. If this test shows the null hypothesis is acceptable, it is concluded that the null hypothesis function (equation (B-1)) does not add any further description of the variation in Y then that provided by the independent variables in equation (B-2). However, this information is insufficient to determine whether the null hypothesis model encompasses the alternative hypothesis model. Thus, the second stage of the test consists of artificially nesting the alternative hypothesis model (equation (B-2)) in the null hypothesis mode (equation (B-1)). Similar to the first stage of the test, this is done by specifying the predicted values of equation (B-2) resulting from OLS estimation as an additional variable in equation (B-1) to form equation (B-4).

$$(B-4) \quad H_0: Y = \beta_0 + \sum_{k=1}^K \beta_k X_k + \delta y_a + e_i$$

y_a represents the predicted values of the alternate hypothesis function and δ is the estimated parameter. The secondary hypothesis test is then $H_0: \delta = 0$ and $H_a: \delta \neq 0$. If this test shows the null hypothesis is acceptable, it is concluded that the alternate hypothesis function (equation (B-2)) does not add any further description of the variation in Y then that provided by the independent variables in equation (B-1).

Maddala (1992) summarizes the possible outcomes of the J-test in a table similar to that provided in table 10 below.

Table 10.-- Possible Outcomes of The J-test.

	Hypothesis: $\alpha = 0$	
Hypothesis: $\delta = 0$	Not Rejected	Rejected
Not Rejected	Both $H_0:$ and $H_a:$ are acceptable	$H_a:$ is acceptable $H_0:$ is not acceptable
Rejected	$H_0:$ is acceptable $H_a:$ is not acceptable	Neither $H_0:$ nor $H_a:$ are acceptable

Source: Maddala (1992)

APPENDIX C

TECHNICAL DESCRIPTION OF DIAGNOSTIC TESTS
AND MODELING OF OUTLIERS

This appendix provides technical descriptions of several of the diagnostic tests summarized in chapter 2. Technical descriptions of White's test for homoscedasticity, DFFITS and DFBETAS measures for detecting outliers, and a method of modeling the effects of outliers using dummy variables are summarized in turn. This is followed by a summary of preliminary modeling efforts designed to address the anomaly in the data aggregation process presented by trips originating in Alaska noted in Chapter 4. Alternative specifications of models 1, 2 (table 4), and 5 (table 6) are presented in turn.

White's Test for Homoscedasticity

White's test for homoscedasticity consists of regressing the squared errors on the variables in the regression model being tested using the general form in equation (C-1).

$$(C-1) \quad e_i^2 = \delta_0 + \sum_{p=1}^P \delta_p Z_{ip} + u_i$$

In this equation e_i are the residuals, δ are estimated regression coefficients, Z_p , and p is determined by the number of independent variables in the regression function being tested using the following example taken from Kmenta (1986). When there is one independent variable (X), $p = 2$ and $Z_{i1} = X_{i1}$ and $Z_{i2} = X_i^2$. When there are two independent variables, $p = 3$ and $Z_{i1} = X_{i1}$, $Z_{i2} = X_{i2}$, $Z_{i3} = X_{i1}X_{i2}$, $Z_{i4} = X_{i1}^2$, and $Z_{i5} = X_{i2}^2$. Regressors in equation (C-1) are determined likewise for regressions with more independent variables.

The null hypothesis of White's test is that the variances of the residuals are constant. Failure to reject this hypothesis also says that any heteroscedasticity is caused by sampling error. The asymptotic, large-sample statistic is computed by $n(R_w^2)$ which is distributed as a χ^2 with p degrees of freedom.

DFFITS and DFBETAS Measures

The general approach to detecting and measuring the impact of outliers on a regression function used in this study is based on the leverage a single observation or group of observations has on the mean and variance of the

remaining observations. These methods can be divided into those in which observations on the dependent and independent variables are examined as outliers influencing the fit of the regression and those which identify cases influencing the fit of the regression equation or the values of estimated coefficients. Both methods utilize leverage values (h_{ii}) which consist of the diagonal elements of the hat matrix stemming from the data for the independent variables (see, e.g. Neter et al. (1989), Kmenta (1986) or Greene (1990) for development of the hat matrix). An explanation of the second of these general methods classified here and use of leverage values in each are summarized below.

Neter et al. (1989) suggest use of DFFITS and DFBETAS measure to detect influential observations, both of which make use of leverage values. DFFITS measures the influence a particular case has on the fit of a regression equation. This measure may be computed for each case by equation (C-2):

$$(C-2) \quad DFFITS_i = \frac{Y_i - Y_{i(i)}}{\sqrt{MSE_i h_{ii}}}$$

In this equation Y_i is the fitted value of Y (the dependent variable) for the i th case, $Y_{i(i)}$ is the predicted value of Y for the i th case when it is omitted from the estimated regression equation, MSE_i is the mean squared error when the

ith case is omitted, and h_{ii} is the leverage value of the ith case. Thus, $DFFITs_i$ is a standardized measure of the deviation of the regression fit when all data are used and when the ith case is removed. In large data sets, the absolute value of DFFITS values exceeding the $2(k/n)^{1/2}$ indicate an influential case. k and n are the number of estimated coefficients in the regression equation and the sample size, respectively.

DFBETAS measures the influence a case has on the slope parameter of a particular variable or on the constant term. DFBETAS values are computed using equation (C-3):

$$(C-3) \quad DFBETAS_{k(i)} = \frac{\beta_k - \beta_{k(i)}}{\sqrt{MSE_i c_{kk}}}$$

In this equation β_k is the estimated coefficient when all N cases are included for the i th case, $\beta_{k(i)}$ is the estimated regression coefficient when the i th case is omitted from the estimated regression equation, MSE_i is the mean squared error when the i th case is omitted, and c_{kk} is the k th diagonal element of the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix. $DFBETAS_i$ is a standardized measure of the difference between the estimated regression coefficient when the i th case is included and when it is omitted. In large data sets Neter et al. (1989) recommend that absolute values of DFBETAS values exceeding $2/(n)^{1/2}$ should be considered influential.

Modeling the Effects of Outliers Using Dummy Variables

Once data outliers are identified, specification of the function with dummy variables can be used to assess the collective impact a group of outliers has on certain coefficients including the intercept. Kmenta (1986) shows how this can be done using a dummy variable specified in the function as variable affecting the intercept and as an interaction term with any of the variables in the function thereby affecting the slope of these variables. For instance, preliminary analysis of the first-stage demand function estimated in this study shows that aggregation of per capita visits from Alaska do not follow the rules used to aggregate the balance of the origin-destination pairs in the stream fisheries data set (see Chapter Four). To account for this difference equation (2) could be specified as

$$(C-4) \quad V_{ij} = \beta_0 - \beta_1 D_{ij} + \beta_2 AK + \beta_3 AK(D_{ij}) + e$$

where AK is a dummy variable in which $AK = 1$ for observations of trips originating in Alaska and $AK = 0$ for all trips from all other origins. This specification then allows analysis of how trips from Alaska influence the constant term (β_0) and price coefficient (β_1). Specifically, if $AK = 1$ then equation (C-4) would reduce to equation (C-5) in which the impact of trips taken from Alaska would influence the constant and slope coefficients

of equation (2) in its linear form.

$$(C-5) \quad V_{ij} = (\beta_0 + \beta_2) + (\beta_3 - \beta_1)D_{ij} + e$$

However, if $AK = 0$, equation (C-4) would reduce to equation (C-6) in which there would be no impact on either the constant or slope coefficients due to trips from Alaska.

$$(C-6) \quad V_{ij} = \beta_0 - \beta_1 D_{ij} + e$$

The impact of trips from Alaska on the constant and slope coefficients can be determined from hypothesis tests of the statistical significance of β_2 and β_3 , respectively.

Expansion of the analysis to incorporate the Box-Cox transformation on V_{ij} and D_{ij} as specified in equation (2) would result in equation (C-7) where the Box-Cox transformation on average round-trip distance is the same for both occurrences of this variable in equation (C-7). Thus, inclusion of AK in this way allows adjustment of the Box-Cox transformation (λ_D) according to the influence of trips from Alaska.

$$(C-7) \quad V_{ij}^{(\lambda_V)} = \beta_0 - \beta_1 D_{ij}^{(\lambda_D)} + \beta_2 AK + \beta_3 AK(D_{ij}^{(\lambda_D)}) + e$$

Analysis of Modeling Trips from Alaska as a Dummy Variable
in Models 1 and 2

To understand the affect of the Alaska trips on each of models 1 and 2, an attempt to include a dummy variable for trips from Alaska affecting the model both as a shift and slope parameter (i.e., as an interaction term between Alaska trips and the log of the distance variables) in one equation for each model failed due to high colinearity between these two variables. Thus, two models, one including a dummy variable for Alaska trips affecting the intercept and another the slope parameters were estimated for each of the double-log and double-log, shifted distance models, holding the functional form and shifted distance constant. The analysis of variance of these new models with their original counterparts (table 4) showed the additional variables to be significant additions. Further, the adjusted-R² for each model was enhanced. The most significant result of these new models is that the two different dummy variable specifications impacted the slope and intercept parameters equally. However, since the Alaska trips represent relatively extreme distances in the data set (versus a more uniform dispersement across the range of distances) it was decided to model Alaska trips as a dummy variable affecting the slope parameter using an interaction variable. The resulting equations are presented in table 11.

Table 11.-- Estimates of Models 1 and 2 with Alaska Modeled as an Interaction Parameter Dummy Variable.

Model	Log-Likelihood	β_0	β_1	β_2	\bar{R}^2	F
1.a.	7,386.29	- .587	-1.805	.418	.792	1,411
(t-ratios)		(-2.72)	(-53.10)	(10.86)		
(p-values)		(.003)	(<.001)	(<.001)		(<.001)
2.a.	7,446.27	4.01	-2.44	.508	.832	1,840
		(15.24)	(-60.64)	(14.59)		
		(<.001)	(.000)	(<.001)		(<.001)

β_0 , β_1 , and β_2 are the estimated parameters for the intercept, average round-trip distance (plus 64 in model 2.a) and the dummy variable for Alaska trips times the natural log of distance (plus 64 in model 2.a), respectively. The dummy variable for Alaska trips was set equal to one for trips made from Alaska and zero otherwise. The critical values for t and F for both models are 1.6471 and 3.0079, respectively, at a 5 percent probability of a type I error. Finally, the sample size for both models is 741.

Although not provided in this paper, the normal probability plots for models 1a and 2a show less negative skewness than models 1 and 2 (*i.e.* observed points are less bowed to the left than those in figures 1 and 2). Furthermore, the residual versus predicted value plot for models 1a and 2a show greater dispersement of points around the center of each graph and model 2a appears less heteroscedastic in the residuals. However, Glejser tests with a linear form reveal a negative and significant correlation of the distance terms in each of models 1a and 2a suggesting the residuals continue to be heteroscedastic in these models.⁵⁹ Finally, model 1a continues to show the same over and underprediction pattern observed for model 1. Yet, the overprediction for longer distances in model 2a appears mitigated when compared to model 2.

Analysis of Modeling Trips from Alaska as a Dummy Variable in Model 5

Outlier analysis was conducted for model 5 using the DFFITS and DFBETAS measures. Collectively and similar to models 1 and 2, the DFFITS measure and the two DFBETAS measures on the constant and slope parameters revealed influential observations for one-way distances of 2 to 50

⁵⁹ The Glejser test for heteroscedasticity was chosen for models 1a and 2a due to a high degree colinearity among the squared and interaction terms including the Alaska dummy variables required for White's test.

miles. Additionally, all three measures again revealed trips from Alaska as influential observations. Thus, model 5 was reestimated with a dummy variable for trips from Alaska specified as an interaction term with average round-trip distance transformed by the Box-Cox transformation as shown in table 6. This model is designated as model 5.a. This model is presented in table 12. The analysis of variance between models 5 and 5.a shows that inclusion of the interaction Alaska-trip/ Box-Cox transformed average round-trip distance variable enhanced the fit of model 5.

Table 12.-- Estimation of Model 5 with Alaska Trips as an Interaction Dummy Variable with Round-Trip Distance.

Model	Log-Likelihood	β_0	β_1	β_2	\bar{R}^2	F
5.a.	7,540.69	-3.900	-.677	.235	.830	1,807
(t-ratios)		(-28.29)	(-60.1)	(15.02)		
(p-values)		(<.001)	(<.001)	(<.001)		(<.001)

β_0 , β_1 , and β_2 are the estimated parameters for the intercept, average round-trip distance transformed by the Box-Cox parameter estimated for model 5 (table 6) and the interactive dummy variable for Alaska trips. The dummy variable for Alaska trips was set equal to one for trips made from Alaska and zero otherwise. The critical values for t and F in this model are 1.6471 and 3.854, respectively, at a 5 percent probability of a type I error. The sample size is 741.

APPENDIX D
SURVEY FORMS



Montana Department of
Fish, Wildlife & Parks



FISHERIES SURVEY

DID YOU FISH IN MONTANA DURING THE MONTH OF MAY? YES NO IF YES, HOW MANY DAYS DID YOU FISH IN MAY? _____

IF YOU ARE A PIONEER (62 OR OLDER) OR A YOUTH (12 TO 14), DO YOU PLAN TO USE YOUR CONSERVATION LICENSE TO FISH? HUNT? BOTH?

PLEASE REFER TO THE MAPS TO HELP US IDENTIFY THE WATERS YOU FISHED. IF YOU NEED MORE SPACE, PLEASE USE A SEPARATE PIECE OF PAPER AND RETURN IT WITH THIS SURVEY. THANK YOU FOR YOUR TIME AND COOPERATION.

DATE FISHED IN MAY	NAME OF LAKE OR STREAM FISHED	SECTION NUMBER IF INDICATED ON MAP	NEAREST TOWN AND/ OR POINT OF ACCESS OR LANDMARK	TOTAL HOURS FISHED PER DAY	TOTAL NUMBER OF FISH CAUGHT PER DAY		TOTAL NUMBER OF FISH KEPT PER DAY		WAS THE PURPOSE OF YOUR TRIP TO FISH? (Y or N)	DID YOU STAY OVER-NIGHT? (Y or N)	ROUND TRIP DISTANCE TRAVELED ..
					TROUT AND SALMON	OTHER SPORT FISH*	TROUT AND SALMON	OTHER SPORT FISH*			
5/ /86											
5/ /86											
5/ /86											
5/ /86											
5/ /86											
5/ /86											
5/ /86											
5/ /86											
5/ /86											
5/ /86											
5/ /86											

ENTER EACH DAY AND EACH WATER FISHED ON A SEPARATE LINE. LIST ALL FISHING IN MONTANA, NOT JUST WATERS INDICATED ON MAPS.

* SUCH AS: THE NUMBER OF WHITEFISH, PERCH, BASS, ECT. ** IF YOU STAYED OVERNIGHT, PLEASE MAKE A SEPARATE ENTRY FOR EACH FISHING TRIP. THIS INFORMATION WILL BE HELD IN STRICT CONFIDENCE AND USED FOR MANAGEMENT PURPOSES ONLY.

10) HOW MANY PEOPLE WERE IN THE VEHICLE? 92 93 PEOPLE

20) HOW MANY OF THESE WERE FISHING AT HOME? 94 95 ANGLERS

21) HOW MANY OF THESE PEOPLE FISHING WERE CHILDREN UNDER 12? 96 97 UNDER 12

The next 2 questions ask about the time you spent travelling to _____?

22) SUPPOSE YOU COULD SHORTEN YOUR TRAVEL BY ONE-HALF, SO INSTEAD OF _____ (insert time from #12) IT TOOK ONLY _____ (1/2 #12) TO GET FROM YOUR HOME TO _____. WHAT IS THE MAXIMUM AMOUNT YOU WOULD BE WILLING TO PAY TO REACH IN HALF THE TIME YOU ACTUALLY TOOK? 98 99 100 101

If amount in #22 is 0, ask #23. If not, skip to #24.

23) WHY WOULDN'T YOU PAY MORE TO REDUCE THE TIME YOU SPENT TRAVELLING? REASONS: 102 103 CODE

24) DID YOU ENJOY THE TIME YOU SPENT TRAVELLING TO _____? YES=1 NO=2 104 OTHER CODE: 105 106

The next few questions will help us to determine how much people spend on Montana fishing trips. Could you tell me how much you spent, if anything, on each of the following categories during the trip?

25) a) LODGING, SUCH AS HOTELS, OR CAMPGROUND FEES? 107 108 109 110 1.00

b) SPENT IN ROUTE? 111 112 113 1.00

c) SPENT NEAR FISHING SITE? 114 115 116 1.00

d) FOOD AND BEVERAGES BOUGHT IN RESTAURANTS? 117 118 119 120 1.00

e) SPENT IN ROUTE? 121 122 123 1.00

f) SPENT NEAR FISHING SITE? 124 125 126 1.00

g) FOOD AND BEVERAGES BOUGHT IN STORES? 127 128 129 130 1.00

h) SPENT AT HOME? 131 132 133 1.00

i) SPENT IN ROUTE? 134 135 136 1.00

j) SPENT NEAR FISHING SITE? 137 138 139 1.00

0) TACKLE AND BAIT BOUGHT FOR THIS TRIP? 140 141 142 143 1.00

1) SPENT AT HOME? 144 145 146 1.00

2) SPENT IN ROUTE? 147 148 149 1.00

3) SPENT NEAR FISHING SITE 150 151 152 1.00

e) FEES FOR GUIDES OR OUTFITTERS? 153 154 155 156 1.00

f) ANY OTHER ON SITE PURCHASES SUCH AS BOAT GAS OR RENTAL, DOCKING FEES, FILM, AUTO REPAIRS, ETC. 157 158 159 160 161 1.00

Finally, I have a few more questions about you and your household that don't relate specifically to your trip to _____, but will help us to compare the information we get from different types of anglers.

26) HOW MANY YEARS HAVE YOU BEEN FISHING? 162 163 YEARS

27) HOW MANY DAYS PER YEAR DO YOU FISH, ON THE AVERAGE? 164 165 166 DAYS PER YEAR

28) HOW DOES FISHING COMPARE TO OTHER TYPES OF OUTDOOR RECREATION YOU DO? CODE 167

1) FAVORITE 3) ONE OF MANY ACTIVITIES

2) NONE OF YOUR FAVORITE OTHER: _____

29) COULD YOU ESTIMATE THE TOTAL AMOUNT OF MONEY YOU HAVE INVESTED IN FISHING RELATED EQUIPMENT SUCH AS FISHING BOATS, RODS, TACKLE BOXES, WADERS, ETC? 168 169 170 171 172 1.00

These last questions are for statistical purposes, and will be kept in strict confidence.

30) AS I READ THE FOLLOWING AGE CATEGORIES, PLEASE INDICATE WHICH ONE YOU ARE IN. 173

1) 12 - 15 4) 30 - 39 6) 50 - 61

2) 16 - 20 5) 40 - 49 7) 62 or older

3) 21 - 29

31) HOW MANY PEOPLE LIVE IN YOUR HOUSEHOLD? 174 175 PEOPLE

32) HOW MANY YEARS OF EDUCATION HAVE YOU HAD? 176 177 YEARS

33) AS I READ THE FOLLOWING INCOME CATEGORIES, PLEASE INDICATE THE CATEGORY YOUR TOTAL HOUSEHOLD INCOME BEFORE TAXES IN 1964 IS IN. INCOME 178 179

1) 0 - 5,000 5) 20,000-25,000 9) 40,000-50,000

2) 5,000-10,000 6) 25,000-30,000 10) 50,000-75,000

3) 10,000-15,000 7) 30,000-40,000 10) More than 75,000

4) 15,000-20,000

10) HOW MANY PEOPLE WERE IN THE VEHICLE? 92 93 PEOPLE

20) HOW MANY OF THESE WERE FISHING AT HOME? 94 95 ANGLERS

21) HOW MANY OF THESE PEOPLE FISHING WERE CHILDREN UNDER 12? 96 97 UNDER 12

The next 2 questions ask about the time you spent travelling to _____?

22) SUPPOSE YOU COULD SHORTEN YOUR TRAVEL BY ONE-HALF, SO INSTEAD OF _____ (insert time from #12) IT TOOK ONLY _____ (1/2 #12) TO GET FROM YOUR HOME TO _____. WHAT IS THE MAXIMUM AMOUNT YOU WOULD BE WILLING TO PAY TO REACH IN HALF THE TIME YOU ACTUALLY TOOK? 98 99 100 101

If amount in #22 is 0, ask #23. If not, skip to #24.

23) WHY WOULDN'T YOU PAY MORE TO REDUCE THE TIME YOU SPENT TRAVELLING? REASONS: 102 103 CODE

24) DID YOU ENJOY THE TIME YOU SPENT TRAVELLING TO _____? YES=1 NO=2 104 OTHER CODE: 105 106

The next few questions will help us to determine how much people spend on Montana fishing trips. Could you tell me how much you spent, if anything, on each of the following categories during the trip?

25) a) LODGING, SUCH AS HOTELS, OR CAMPGROUND FEES? 107 108 109 110 1.00

b) SPENT IN ROUTE? 111 112 113 1.00

c) SPENT NEAR FISHING SITE? 114 115 116 1.00

d) FOOD AND BEVERAGES BOUGHT IN RESTAURANTS? 117 118 119 120 1.00

e) SPENT IN ROUTE? 121 122 123 1.00

f) SPENT NEAR FISHING SITE? 124 125 126 1.00

g) FOOD AND BEVERAGES BOUGHT IN STORES? 127 128 129 130 1.00

h) SPENT AT HOME? 131 132 133 1.00

i) SPENT IN ROUTE? 134 135 136 1.00

j) SPENT NEAR FISHING SITE? 137 138 139 1.00