ECONOMIC THEORY AND ESTIMATION OF THE DEMAND FOR CONSUMER DURABLE GOODS AND THEIR UTILIZATION: APPLIANCE CHOICE AND THE DEMAND FOR ELECTRICITY

by

Jeffrey Alan Dubin

A.B. University of California, Berkeley

(1978)

Submitted to the Department of Economics
In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

May 1982

Signature of Author Signature Redacted
Department of Economics
May 17, 1982

Certified by Signature Redacted
Daniel L. McFadden
Thesis Supervisor

Accepted by Signature Redacted
Richard Eckaus
Graduate Registration Office

ECONOMIC THEORY AND ESTIMATION OF THE DEMAND FOR CONSUMER DURABLE GOODS AND THEIR UTILIZATION: APPLIANCE CHOICE AND THE DEMAND FOR ELECTRICITY

by

Jeffrey Alan Dubin

Submitted to the Department of Economics on May 17, 1982
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy at the Massachusetts Institute of Technology

ABSTRACT

This thesis develops the theory of durable choice and utilization. The basic assumption is that the demand for energy is a derived demand arising through the production of household services. Durable choice is associated with the choice of a particular technology for providing the household service. Econometric systems are derived which capture both the discrete choice nature of appliance selection and the determination of continuous conditional demand.

Conditional moments in the generalized extreme value family are derived to extend discrete continuous econometric systems in which discrete choice is assumed logistic. An efficiency comparison of various two-stage consistent estimation techniques applied to a single equation of a dummy endogenous simultaneous equation system is undertaken and asymptotic distributions are derived for each estimation method.

Using the National Interim Energy Consumption Survey (NIECS) from 1978 we estimate a nested logit model of room air-conditioning, central air-conditioning, space-heating, and water heating. The estimated probability choice model is used to forecast the impacts of proposed building standards for newly constructed single family detached residences. Monthly billing data matched to NIECS is analyzed permitting seasonal estimation of the demand for electricity and natural gas by households.

The theory of price specification for demand subject to a declining rate structure is reviewed and tested. Finally, consistent estimation procedures are used in the presence of possible correlation between dummy variables indicating appliance ownership and the equation error. The hypothesis of simultaneity in the demand system is tested.

Signature redacted

Thesis Supervisor: Dr. Daniel L. McFadden
Title: Professor of Economics

JEFFREY ALAN DUBIN

BIOGRAPHY

Personal Data

Date of Birth: January 24, 1957 (Berkeley, California)
Married to Jacqueline Mary Dubin

Education

A.B. (Economics) University of California, Berkeley, 1978, with highest honors and with great distinction.

Academic Appointments

Reader (Economics), University of California Berkeley, 1977.

Research Assistant (Economics), Massachusetts Institute of Technology, with Professor Daniel L. McFadden, 1979-1981.

Teaching Assistant (Economics), Massachusetts Institute of Technology, Graduate Applied Econometrics (appointment for Spring 1982).

Memberships in Scholarly and Professional Organizations

Phi Beta Kappa

American Economic Association

The Econometric Society

University of California, Lifetime Alumnus

Professional Service

Referee to Review of Economic Studies and Bell Journal of Economics

Fellowships, Scholarships, Honors, and Awards

Edward Frank Kraft Prize 1974-1975

Honor Roll 1974-1978 University of California, Berkeley

Phi Beta Kappa 1978

Department Citation in Economics, University of California, Berkeley, 1978

Massachusetts Institute of Technology Fellowship 1978-1980

ACKNOWLEDGEMENTS

There are several people who have made this thesis possible. First I wish to thank my thesis committee Daniel McFadden, Ernst Berndt, and Franklin Fisher. Ernie and Frank read and commented on early drafts of this work and provided general guidance during this last year. My mentor and friend Dan McFadden provided very important advice and support and much of his valuable time. It has been a great pleasure to work as an apprentice under Dan during the last three years and I owe him a principal intellectual debt in teaching me the art and science of econometrics.

I wish to acknowledge further the support and understanding of my parents over the many years of my education. This thesis must be dedicated to them. I want to mention my oldest and dearest friend, Harry Kraus, who often encouraged me to complete this work and who introduced me to my wife Jackie. Jackie suffered with me the joys and trials of completing an empirical thesis. I am indebted to her for her help with typing and assembling of the manuscript.

I acknowledge the support of NSF 7920052-SOC., NSF 80-16043-DAR through Dan McFadden, the typing staff of the MIT Energy Laboratory, and the Department of Economics at MIT for fellowship support during the first two years in graduate school.

Jeffrey Alan Dubin
May 17, 1982

TABLE OF CONTENTS

Pages

INTRODUCTION AND SUMMARY

Economic Theory and Estimation of the Demand for Consumer Durable Goods and their Utilization: Appliance Choice and the Demand for Electricity

I. Overview

In the years 1947 to 1972 the United States experienced an almost seven-fold increase in the use of electricity. The early 1970's brought the interwined problems of depleting oil resources, increased dependence on oil imports, and a heightened need for a consensus in national energy policy. However, increasing concern over the safety of nuclear power mitigated the trend toward pervasive electrification and the nation's all-electric future.

The need to quantify the responsiveness of electricity utilization to various energy policies rose rapidly in the energy turbulent 1970's. This need was felt all the way down to home owners who became concerned with efficiency and costs of alternative heating and cooling systems. Of course home owners who had witnessed an increase in their energy budget from 26% in 1972 to 33% in 1980 knew all too well that the composition of their appliance stock greatly influenced their usage.[1]

Energy researchers also noted the importance of durable stocks in the energy demand process.[2] Yet, only in very recent attempts have econometric simulation models allowed policy scenarios simultaneously to affect appliance holdings and resultant usage. In one direction are aggregate studies which fit average appliance saturations to the time trend of income, prices, and other socio-economic variables. This approach is best exemplified in the modeling efforts of Hirst and Carney (1978). Other aggregate based studies are extensively reviewed in Hartman (1978, 1979).[3]

In contrast to the aggregate studies, several attempts to model jointly the demand for appliances and the demand for fuels by appliance have been completed using cross-sectional micro level survey data.[4] The use of disaggregated data is desirable as it avoids the confounding effects of either misspecification due to aggregation bias or misspecification due to approximations in rate data.

Either approach has a common objective in modeling household energy consumption patterns from which to evaluate conservation and load management policies. For example, can we evaluate the welfare and distributional impacts of proposed government policies to decontrol the price of natural gas? How rapidly do consumers repond to rising energy prices? What are the differences between the energy consumption of owner-occupiers and tenants? What are the implications for public information programs that provide energy efficiency labeling and building and appliance standards? Does the marketplace offer sufficient incentives to pursue appropriate levels of conservation; what actions should government take, if any, to encourage conservation? Can we quantify the long and short-run responses to policy actions and describe the time path of conservation?

To answer these questions in a logical fashion requires us to conceptualize the residential energy consumption process.

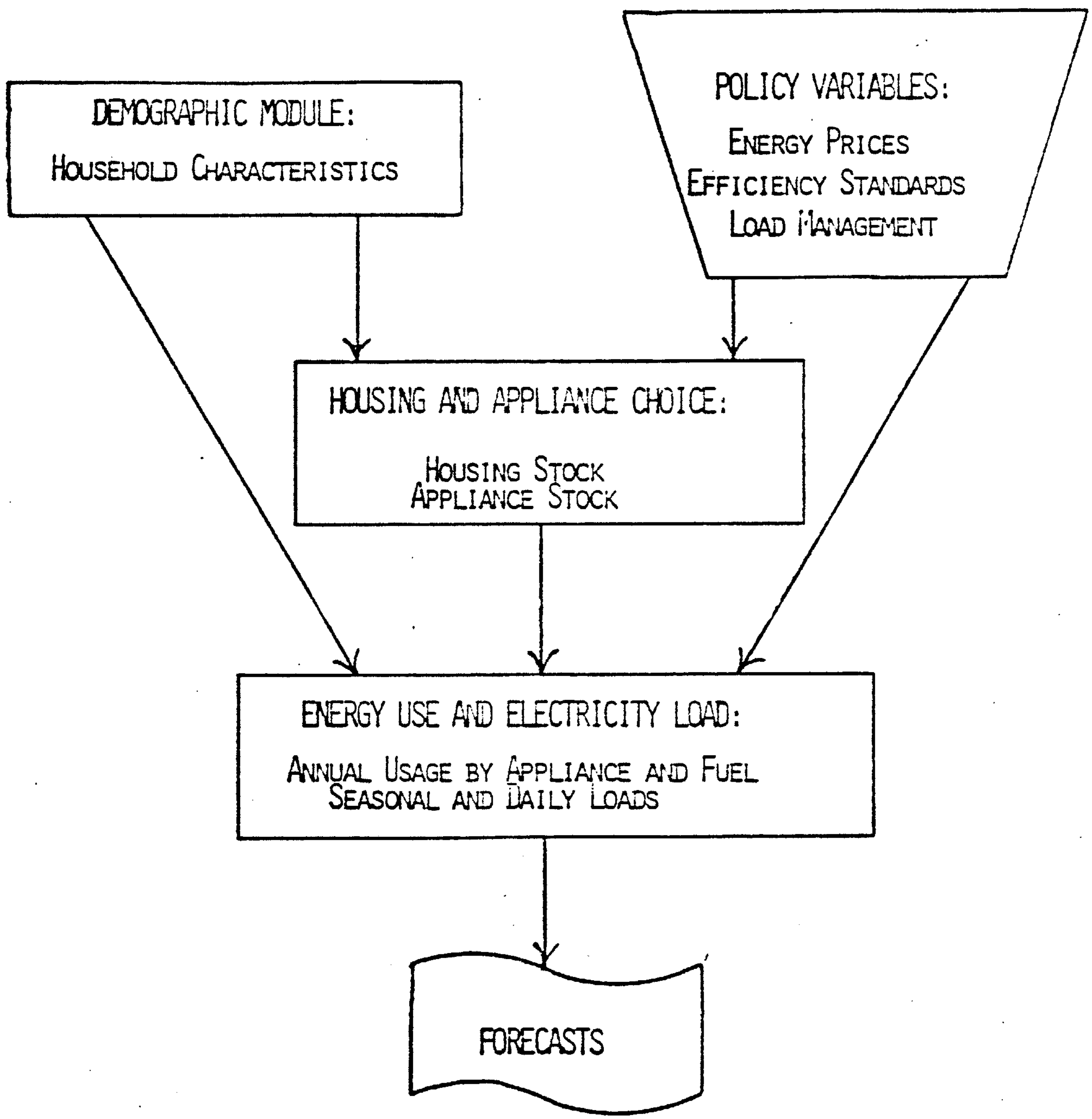
II. The Residential Energy Consumption Process

Figure 1 illustrates the residential energy consumption process. Household demographics, household income, fuel prices, equipment prices, and climate are inputs to a residential choice process which determines appliance and dwelling characteristics. Included in appliance characteristics are fuel types, capacities, efficiencies, and holdings. Included in dwelling characteristics are structure type, size, and thermal integrity. Given the appliance and housing stock, households react to policy and market variables such as energy prices, efficiency standards, etc. to determine energy usage by appliance and by fuel type. Each policy question may be traced in its effects through the diagram in Figure 1. For example, consider a proposed change in the building code which would require all new dwellings to meet a baseline thermal integrity standard through wall and ceiling insulation. The increased thermal integrity in the housing stock would alter the structure of operating and capital costs of available heating and cooling systems available for purchase. Changes in expected operating and capital costs would produce a predictable shift in the saturations of alternative heating and cooling systems. Furthermore, the demand for fuels by appliance would be different to reflect the increased thermal integrity of the dwelling and the resultant changes in the marginal costs of providing these services. For details concerning the implementation of a large scale energy forecasting model the reader should consult Goett (1979) and Cambridge Systematics/WEST (1979).

For the purposes of forecasting, the residential energy consumption process is assumed to be recursive. In the first stage a housing

Figure 1

The Residential Energy Consumption Process



Source: Cambridge Systematics (1979)

decision is made. Conditional on the housing decision, appliance portfolios are chosen by the household, and finally, energy demand is determined conditional on the choice of appliance stock. For the purposes of estimation, however, it must be recognized that the demand for durables and their use are related decisions by the consumer. Econometric specifications which ignore this fact lead to biased and inconsistent estimates of price and income elasticities. It is to these issues that we now turn.

III. Economic and Statistical Issues in Modelling the Choice of Durables and Their Utilization

Economic analysis of the demand for consumer durables suggests that such demand arises from the flow of services provided by durables ownership. The utility associated with a consumer durable is then best characterized as indirect. Durables may vary in capacity, efficiency, versatility, and of course will vary correspondingly in price. Although durables differ, the consumer will ultimately utilize the durable at an intensity level that provides the "necessary" service. Corresponding to this usage will be the cost of the derived demand for the fuel that the durable consumes. The optimization problem posed is thus quite complex. In the spirit of the theory the consumer unit must weigh the alternatives of each appliance against expectations of future use, future energy prices, and current financing decisions.

The specification of econometric demand systems for fuel usage presupposes that consumers can detect prevailing marginal fuel rates in the presence of automatic appliances, billing cycle variations, and limited information on appliance operating characteristics. More fundamentally, there is the assumption that the shares of appliance portfolios in recent construction provide information on consumer preferences, independently of portfolio decisions made by contractors.

Dubin and McFadden (1979) explore these issues and apply several tests to determine the exogeneity of appliance dummy variables typically included in demand for electricity equations. Their approach derives an indirect utility function which is consistent with the specification of a partial demand equation. The indirect utility function is used to predict portfolio choice while the demand equation predicts conditional electricity usage.[5] The demand system consists of simultaneous equations with dummy endogenous variables (Heckman (1978, 1979)) and may be thought of as a switching regression with a structure analyzed by Lee (1981), Goldfeld and Quandt (1972, 1973, and 1976), Maddala and Nelson (1974 and 1975), and Fair and Jaffee (1972).

Employing a logistic discrete choice model of all electric versus all natural gas space and water heat systems combined with conditional demand for electricity, Dubin and McFadden (1979) reject the hypothesis that unobserved factors influencing portfolio choice are independent of the unobserved factors influencing intensity of use.

The purpose of this thesis is to analyze the residential demand for electricity and natural gas conditional on the choice of space heat, water heat, central and room air conditioning choice utilizing the National Interim Energy Consumption Survey (NIECS) 1978 survey of 4081 households. The model developed in this thesis is intended to have the flexibility to be included into a large micro-simulation forecasting system (such as the Residential End-Use Energy Policy

System (REEPS)).[6] The thesis further extends the theoretical development of durable choice and utilization and seeks to examine the hypothesis of simultaneity between appliance choice and electricity and natural gas demand. The thesis is organized into four chapters and three appendices.

IV. Organization of Thesis

In Chapter One we develop the theory of durable choice and utilization. The basic assumption is that the demand for energy is a derived demand arising through the production of household services. The technology which provides the household service is the appliance durable. Durable choice is then associated with the choice of a particular technology from a set of alternative technologies. Using results from household production theory, we derive econometric systems which capture both the discrete choice nature of appliance selection and the determination of continuous conditional demand.

Chapter two reviews the theory of price specification and considers the comparative static analysis of demand subject to a declining block rate schedule. We further investigate the statistical endogeneity of prices whose construction requires utilization of the observed consumption level, and determine price specification within a sample of 744 households surveyed in 1975 by the Washington Center for Metropolitan Studies (WCMS). We finally consider the construction of marginal prices using the WCMS data and monthly billing data from NIECS.

Chapter Three describes the estimation of a discrete choice model for room air-conditioning, central air-conditioning, space heating, and water heating. The form of the appliance choice model results from the assumption that the unobserved components of utility have a generalized extreme value distribution. A particular form of this distribution is considered which implies that the choice of room air conditioning given the choice of central air conditioing is independent of the choice of space heat system given the choice of central air conditioning. Water heat fuel choice is assumed to depend only on the choice of space heat system.

Chapter Four presents the estimation of the demand for electricity and natural gas. Consistent estimation procedures are used in the presence of possible correlation between the dummy variables indicating appliance holdings and the equation error term. We perform tests for simultaneity using the methods of Hausman (1978). Estimation is based on monthly billing data matched to each household in the NIECS survey. The monthly billing data provides an excellent time profile of usage which permits the determination of individual seasonal effects.

The main text of the thesis is followed by three technical appendices. The first appendix describes the processing of the NIECS data and the creation of an appended NIECS data base. It further describes the creation of marginal electricity and natural gas prices based upon the theory of Chapter Two and describes the use of a network thermal model to provide unit energy consumptions for alternative heating and cooling systems across time.[7]

The second appendix presents the calculation of various conditional moments in the generalized extreme value family. These results extend the analysis given in Dubin and McFadden (1979) for the case of discrete continuous econometric systems where discrete choice is assumed logistic. Finally, this appendix provides the conditional expectations used in selectivity type corrections of dummy endogenous variable systems in which the probability system is nested logistic.[8]

The third appendix considers an efficiency comparison of various two-stage consistent estimation techniques applied to a single equation which is linear in parameters but possibly non-linear in the interaction of a dummy endogenous variable and other exogenous explanatory variables. This class of models covers the demand system estimated in Chapter Four as well as the system of Dubin and McFadden (1979) and Heckman (1979). Asymptotic distributions are derived for each estimator using the methods of Amemiya (1978, 1979).

Footnotes

1. "Annual Report to Congress, Volume Two: Data, "U.S. Department of Energy, Energy Information Administration Report DOE/EIA-0173-(80)/2 (April, 1981), p. 9.

2. Classical studies of aggregate electricity consumption given appliance stocks are Houthakker (1951), Houthakker and Taylor (1970), and Fisher and Kaysen (1962). A number of other studies postulate an adaptive adjustment of consumption to long-run equilibrium, which can be attributed to long-run adjustments in holdings of appliances; see Taylor (1975).

3. The Hartman review describes both single fuel and inter-fuel substitution models. Among the single fuel demand studies based on aggregate data, Hartman includes Acton, Mitchell, and Mowill (1976), Acton, Mitchell, and Mowill (1978), Anderson (1973), Chern and Lin (1976), Hartman and Werth (1979), Mount, Chapman and Tyrell (1973), Wilder and Willenborg (1975), and Wilson (1971).

4. Cross-section studies with this structure are McFadden-Kirschner-Puig (1977), the residential forecasting model of the California Energy Conservation and Development Commission (1979), the micro-simulation model developed by Cambridge Systematics/West for the Electric Power Research Institute described in Cambridge Systematics/West (1979), Goett (1979), and Goett, McFadden, and Earl (1980).

5. Related work in the area of discrete/continuous econometric systems is given in McFadden (1979), Duncan (1980a), Duncan and Leigh (1980), Duncan (1980b), Hay (1979), King (1980), Lee and Trost (1978), McFadden and Winston (1981), and Hausman and Trimble (1981).

6. See Cambridge Systematics/West (1979) for a description of REEPS.

7. See McFadden and Dubin (1982) for details about the thermal model developed to provide capacity and baseline usage of alternative heating and cooling systems in NIECS single family detached dwellings.

8. The nested logit model is described in McFadden (1978, 1979, and 1981).

CHAPTER I

ON THE THEORY AND ESTIMATION OF CONSUMER DURABLE CHOICE AND UTILIZATION[1]

This chapter reviews and extends the economic and econometric models of consumer durable choice, holdings, and utilization. Examples are drawn primarily from the literature on electricity demand and appliance choice but much of the exposition is consistent with a wider realm of household behavior. For instance, the methodology could be used to develop a model of household automobile choice and utilization without substantive modification.

Consumer durable models are usefully classified by their treatment of durable utilization in addition to the frequent distinction between holdings and purchase. Broadly speaking, a purchase model analyzes the decision to acquire a durable stock while a holdings model attempts to explain how the stock evolves during its economic life.

Examples of pure holdings models are Diewert (1974) who uses the classical stock-flow model to analyze the demand for money over time, and Griliches (1960) who uses a stock-adjustment model to estimate the demand for farm equipment. Pure purchase or choice models are considered by Chow (1957) in the context of the demand for automobiles, Cragg and Uhler (1970), Cragg (1971), and Li (1977) for housing choice. Appliance purchase models are considered by McFadden-Kirschner-Puig (1977).

Examples of holdings and utilization models are the classical stock-flow utilization studies of aggregate electricty consumption given applicance stocks by Houthakker (1951), Houthakker and Taylor (1970), and Fisher and Kaysen (1962). Stock-adjustment models with utilization are treated in the work of Balestra and Nerlove (1966) on the demand for natural gas.

Purchase or choice models for durable goods which jointly consider utilization are very recent. Dubin and McFadden (1979), Hartman (1979), and Hausman (1979) all consider discrete choice models of appliance ownership and corresponding utilization.

In general, any model of consumer durable choice should consider:

1) the distinction between the decision to purchase a stock of durable goods and the decision to hold or replace that stock,
2) the inherent "discreteness" of durable goods, e.g., while additional cooling may be provided by an individual room air-conditioner, available units offer only fixed ranges of capacity,
3) the imperfect or non-existence of rental markets for durable re-sale,
4) the sizable transaction and installation costs often connected with the decision to retrofit or upgrade a durable stock,
5) the intertemporal utility maximization problem that results from the inherently dynamic choice of a durable stock and the utilization of that stock over its lifetime,
6) the characterization of any solution to be conditioned on information available to the consumer at the time the decision is made; the modifications to that solution as new information becomes available, e.g., technological innovation or change in the relative costs of alternative fuels, and
7) the link between a durable good and the technology which it often embodies.

Unfortunately, previous literature has failed to incorporate all of these crucial points in a consistent model of durable choice behavior. For example, the classical holdings model of consumer durables as presented in Diewert (1974) assumes perfect foresight, perfect rental markets, and a flow of services that results from a stock of durable goods which depreciates but may be augmented continuously. This capital-theoretic framework fails to integrate the purchase decision with the decision to utilize or change the durable stock. The initial choice of durable stock with given features is crucially important, however, since the realization of levels or rates of change of key economic variables which differ from the consumer's ex-ante predicted values may make the ex-ante optimal durable choice ex-post nondesirable. Faced with low resale values of his durable stock, non-accessibility to markets for re-sale, or high transaction costs involved with the decision to retrofit, the consumer would not be expected to change his durable stock often and perhaps only when very large changes in utility had occurred. Furthermore, prices of durable goods reflect their capitalized rents and hence tend to have values which become significant fractions of consumers' budgets. The resolution of financing large initial set-up costs may directly affect durable choice when some consumers' access to capital markets is limited. This may indirectly affect the choice of other economic goods and thus affect consumer welfare.

The importance of initial purchase is derived from the notion that once a durable stock is purchased it will remain intact for many years. The classical model de-emphasizes the purchase decision by allowing "putty-putty" flexibility in durable stocks.

It would be unfair to say that the classical model cannot treat aspects of transaction costs and limited rental markets. Such factors may be incorporated into stock-flow models but invariably surface in their effects on the "user cost of capital." A change in the user cost of capital induces an immediate and continuous response in the desired level of durable stock.

As an alternative to the classical model, consider the general discrete choice model. The discrete-choice model assumes that the purchase, holding, and replacement decisions correspond to differences in utility values crossing threshold levels. The decision to change the level of durable holdings is viewed as a discrete movement from one durable portfolio combination to another. This change is typically costly and occurs infrequently for the usual consumer.

The discrete and classical models of individual choice behavior differ in that the former does not assume that the stock of durable goods can be changed continuously. Thus differences between desired and actual stocks are not instantaneously or adaptively actualized as in the classical model. Finally, depreciation itself is often a stochastic phenomenon which represents durable failure and necessitates a repair or replacement decision on a very discontinuous basis. These distinctions are potentially important since they may imply rather different choice behavior by consumers. A comparison of the predictive abilities of the discrete choice approach with the classical model of durable choice awaits our empirical results.

The bulk of this chapter then is concerned with rigorously developing a theoretical and econometric framework for analyzing durable choice from a discrete choice perspective. We begin the chapter by reviewing several classical models and investigate their extensions. In Section II, we turn to the development of the discrete choice approach by considering two examples.

The first example motivates the characterization of durable selection as the choice of a particular technology for producing household services which yield direct satisfaction to the consumer. This link to household production theory relaxes the assumed proportionality relationship between flows and stocks in the classical model. The second example explores the engineering characterization of durable selection which emphasizes the trade-offs between operating and capital costs. The engineering approach is shown to be the natural dual to a general utility maximization model which incorporates the aspects of discrete choice, household production and the trade-off between operating and capital costs.

In Section IV, we seek conditions on technology and preferences under which household production of durable services follows a two-stage plan. In the first stage, consumers determine optimal production service levels and in the second stage choose input combinations which produce these services at minimum cost. Section V introduces several econometric models of discrete choice and utilization with explicit attention given to the link with the theoretical model and the treatment of stochastic components. A final section provides a summary and conclusions.

II. Classical Models of Consumer Durable Choice

This section reviews the classical stock flow model and the user cost of capital concept. We then modify the stock-flow model to allow a fixed coefficient technology and an element of discreteness in the durable stock.

1. Stock-Flow Model

For simplicity we discuss a two-period consumer choice model with complete markets and perfect information. Assume that in each period, consumers derive utility from consumption of a non-durable good, denoted by q, and from consumption of the flow of services provided by the stock, K, of a durable good. Here we assume that the flow of services is proportional to the stock and denote the intertemporal utility function by $U(q_1, q_2, K_1, K_2)$ where the stock variables replace the flow variables by a change in units. The basic notation to be used in this section is:

q_j = consumption of non-durable good in period j

p_j = spot price of non-durable good in period j

K_j = stock of durable good in period j

S_j = savings in period j

v_j = spot price of durable good in period j

W_j = income in period j

D_j = purchases of durable good in period j

ω = depreciation rate

i = interest rate

In keeping with the spirit of this model, we assume that income is exogenously determined in each period and that spot prices are known with certainty. In this classical framework, the durable good K is defined over a continuous range and is assumed to depreciate continuously at rate ω.

Three equations determine the relationships among the state variables:

(1) $W_1 - p_1q_1 - v_1K_1 = S_1$

(2) $W_2 + S_1(1+i) = p_2q_2 + v_2D_2$

(3) $K_2 = D_2 + (1-\omega)K_1$

Equation (1) states that cash flow in period 1 is income in period 1 less expenditures on durable and non-durable goods in period 1. Equation (2) similarly states that expenditures on durable and non-durable goods in period 2 must equal disposable income defined by income in period 2 and the second period value of the first period cash flow. In (3), the level of durable stock in period 2 is determined by purchases of the durable good in period 2 plus the net (after depreciation) level of stock of durable good from period 1. Note that we set $S_2 = 0$ which is the two period model constraint and have implicitly set $D_1 = K_1$ which implies from (3) that the consumer begins period 1 without any durable stock. This implies a minor asymmetry between periods 1 and 2 which is basic to finite time horizon models.

We combine equations (1) and (2) to obtain:

(4) $W_1 + W_2/(1+i) = p_1q_1 + p_2q_2/(1+i) + v_1K_1 + v_2D_2/(1+i)$

In (4), expenditures are allocated over the two periods so that their present discounted value is equal to wealth, i.e., the present discounted value of income. Combining equation (4) with equation (3) we obtain:

(5) $W_1 + W_2/(1+i) = p_1q_1 + p_2q_2/(1+i) + [v_1 - ((1-\omega)/(1+i))v_2] \cdot K_1$
$+ [v_2/(1+i)] \cdot K_2$

Equation (5) now has the usual form of a budget constraint set for the utility function $U[q_1, q_2, K_1, K_2]$. The "price of K_1, $[v_1 - ((1-\omega)/(1+i))v_2]$, is the "user cost of capital" or "rental equivalent price". Purchasing one unit of durable good has an associated cost of v_1. After one period, $(1-\omega)$ units of the durable stock will remain due to depreciation. The present discounted value of the revenue from reselling the $(1-\omega)$ units of durables at price v_2 is $[(1-\omega)v_2/(1+i)]$. The net price is then clearly the difference.

An essential feature of the stock-flow model of durable holdings is the definition of rental equivalent prices. This is accomplished through rearrangement of the budget constraint set and does not involve the preferences defined by $U[q_1, q_2, K_1, K_2]$. The extension of the definitions of user cost and rental equivalent prices where there are more than two periods is straightforward.

Diewert performs precisely this generalization and estimates rental equivalent prices for durable commodities. He then fits a flexible intertemporal indirect utility function using the defined prices.

Diewert (1974) and others have noted that the concept of user cost may be related to the rate of nominal appreciation or depreciation in capital value of the durable good. Specifically, let $k = (v_2-v_1)/v_1$ so that:

$$(6) \quad [v_1 - ((1-\omega)/(1+i))v_2] = v_1[1 - ((1-\omega)/(1+i))(1+k)].$$

A first-order Taylor approximation implies that the second term can be written as $v_1[i + \omega - k]$. When second period prices are unknown and consumers use estimated values for k, it is possible that the user cost term may be negative. This would, unrealistically, imply optimally unbounded purchase of the durable in the first period.[2] One method of smoothing the connection between the predicted changes in durable stocks implied by changes in the rental equivalent price is to postulate a lag structure in which stocks of durables adjust partially in the direction of the difference between desired and actual holdings. The stock-adjustment variants of the stock-flow model require strong assumptions both in their theory and in their estimation.

The components of user cost v_1, i, ω, and k are in reality specific to a particular consumer and a particular durable type. An important generalization to be considered below is the case of a population of consumers with heterogeneous tastes and with choices defined over a broad range of durable categories.

2. Consumer Choice of Fixed Coefficient Technology with Operating Costs

We now extend the stock-flow model of durable choice to incorporate the effects of operating costs. Here we link the durable choice to the selection of a technology for producing a given end-use service. Consider the classic example of a light bulb which may be regarded as a durable good. That is, it represents the technology for producing so many candle hours of lighting service while requiring the basic fuel input of electricity. In this example it is reasonable to assume that the energy service ratio defining electricity input per unit of service output is constant. This assumption is equivalent to assuming that lighting services are delivered by a fixed-coefficient technology.

To extend the neo-classical durable choice model, define:

x_j = consumption of input commodity

θ = energy service ratio

w_j = spot price of input commodity

Equations (1) and (2) are modified in equations (7) and (8) respectively to include purchase of the input commodity

(7) $W_1 - p_1q_1 - w_1x_1 - v_1K_1 = S_1$

(8) $W_2 + S_1(1+i) = p_2q_2 + w_2x_2 + v_2I_2$

Equation (3) remains unchanged while the technology for constant energy service ratio is:

(9) $x_j = \theta \cdot K_j$ for $j = 1,2$

Although the energy service ratio is assumed constant for the present, it would more generally be related to the rate of depreciation and fuel or durable type, etc. Combining equations (3), (7), (8), and (9) we obtain:

(10) $[W_1 + W_2/(1+i)] = p_1q_1 + p_2q_2/(1+i) + [v_1 - ((1-\omega)/(1+i))v_2 + w_1\theta]K_1 + [(v_2 + w_2\theta)/(1+i)]K_2$

The "price" of K_1, $[v_1 - ((1-\omega)/(1+i))v_2 + w_1\theta]$, consists of the rental equivalent price as defined above plus the term $w_1\theta$ which represents the input price per unit of service.

Provided that production technologies for end-use service exhibit constant returns to scale, it is clear that the user cost concept can be extended to include operating costs. Technologies which do not exhibit constant returns to scale are considered below.

3. Neo-Classical Choice of Discrete Durable Stock

Some attempts have been made to incorporate discreteness in a single-period neo-classical framework.[3] To highlight the salient features of this approach, suppose that consumers either own one unit of durable stock, $K_1 = 1$, or they do not, $K_1 = 0$. Assume that consumers derive utility $U[q_1, K_1]$ from a flow of services assumed proportional to the durable stock and from consumption of a single non-durable good. The one-period budget constraint is:

(11) $W_1 = p_1 q_1 + v_1 K_1$

The durable good is purchased when

(12) $U[(W_1 - v_1)/p_1, 1] > U[W_1/p_1, 0]$

For concreteness, assume $U[q_1, K_1] = (K_1+k_1)^{\alpha} \cdot q_1^{(1-\alpha)}$ with $k_1 > 0$. Then condition (12) implies:

(13) $(1+k_1)^{\alpha} \cdot [(W_1-v_1)/p_1]^{(1-\alpha)} \geq [k_1^{\alpha} \cdot (Y/p_1)^{(1-\alpha)}]$

If we let d_1 be the constant $[(1+k_1)/k_1]^{\alpha(1-\alpha)}$ then condition (13) holds when $W_1 \geq d_1 v_1/(1-d_1)$. The income level $W_1^0 = d_1 v_1/(1-d_1)$ marks a threshold level of expenditure delineating durable and non-durable owners. The generalization of this simple example to a population of consumers with heterogeneous tastes motivates a <u>probabilistic choice system</u>.

To generate a probabilistic choice system we might assume that the behavioral parameter α has a distribution $F_{\alpha}[t]$ in the population. Let $F_{d_1}[t]$ denote the cumulative distribution function for d_1 induced by the distribution of α. Then from (12) we have:

(14) Prob[durable is purchased] = $\text{Prob}[W_1 \geq W_1^0]$

$$= \int_{-\infty}^{W_1/(W_1+v_1)} dF_{d_1}[t]$$

In the next section, we consider the specification of more general probabilistic choice systems for durable-technology choice consistent with the specification of demand for end-use service.

III. Consumer Durable Choice and Appliance Technology

The demand for energy by the household is a derived demand arising through the production of household services. The technology which provides household services is embodied in the household appliance durable. To understand the residential demand for energy we must therefore understand the residential demand for durable equipment.

Assume that a household faces a decision in which a space heating system is being considered. This decision may arise as a result of the installation of a heating system in new construction, as part of a technological upgrading of the existing stock (the "retrofit" decision), or from replacement due to existing system failure. Observational experience suggests that households choose a temperature profile during a 24-hour period which they attempt to attain using their heating system. For some households this may involve setting the thermostat at one temperature during the day and at another level at night. Other households rely on thermostat timers or simply the "feel" of the coldness in the air.

The degree to which a given housing structure loses heat to the colder outside is related directly to the size of the various exposed surfaces and their conductivity to heat flow as well as the absolute temperature differential. Insulation in the walls and ceiling and the presence of storm windows all lower the overall thermal conductivity of the housing shell and hence the requirements on a heating system to maintain a given comfort level. As the temperature differential between inside and outside increases, the capacity of a system for providing delivered BTU's of heat may be reached. Recommended construction practice suggests that a space heating system should provide adequate heating capacity against all but the coldest 1 percent of the heating season.[4] It is thus an engineering decision which determines required capacity.

Given the capacity of the system, households then choose among available technologies and delivery systems. For example, space heating is commonly provided by central forced air, wall units, hot water radiators, etc. Each system is available at a corresponding capital cost. In choosing a given space heating system type, consumers face an economic decision in which they compare the initial dis-utility of purchasing the capital equipment with the future utility of the heating services provided by its operation.

The simultaneous consideration of ex-ante purchase and ex-post utilization apply to a wide variety of appliance durables.[5] Assume that the consumer faces a set B of possible appliance designs. We distinguish between variable parameters, a, and fixed design parameters, K, in the definition of $b = (a,K) \in B$. Examples of characteristics which are fixed in the design and construction of a given appliance and not subject to variation by consumer are capacity, size, voltage, recovery rate, reliability, appearance, durability, and range of operation. Other fixed factors concern the affect of the structure on appliance technology. Examples of structural parameters are the size of the dwelling, the number of rooms, and the thermal integrity of the dwelling.

Variable parameters consist primarily of environmental factors and perhaps the outcome of a random failure of an appliance or a random change in technological performance.

Environmental factors are typically beyond the control of the individual. Structural parameters are variable in the longest run in which major structural changes can be effected. Important exceptions to this include a change in the thermal integrity of the dwelling resulting from installation of insulation or storm windows.

An appliance production plan, $Y = \{Y_t, t = 1, 2, \ldots L\}$ consists of netput vectors $Y_t = (Z_t, -X_t)$ where components of Z_t are positive outputs and components of X_t are positive inputs. The production plan Y is feasible when Y is a member of the restricted technology set V(b) corresponding to design vector $b \in B$. Outputs of a production plan corresponding to a given appliance technology are end-use services which yield direct utility to the individual. Examples of residential services are degree hours of heating or degree hours of cooling, degree hours of maintained water temperature, loads of dishes washed, etc.

Inputs to an appliance technology would include labor, labor and materials for maintenance, and primarily fuel. Fuel input would almost certainly be determined by choice of a fixed design parameter. Joint production is possible and provides a natural framework for the technology of space-conditioning in which one durable good provides both cooling and heating capability.

We assume that individuals maximize an intertemporal utility function $U[Z,Z^o]$ where Z are the outputs of an appliance production plan, and Z^o is a consumption plan in traded commodities Z_t^o, with $Z^o = \{Z_t^o, t = 1, 2, \ldots L\}$. We further assume that individuals contract for inputs on future markets with vector P_X and price vector P_{Z^o} for traded commodities Z^o subject to a budget constraint in wealth W. Suppose further that appliance technology V(a,K) is available to the consumer at cost H[K]. The consumer's problem is then:

$$\max U[Z,Z^o] \text{ subject to:}$$

$$P_X X + P_{Z^o} Z^o \leqq W - H[K] \quad \text{and} \quad Y = (Z,-X) \in V(b) \text{ for } b = (a,K) \in B.$$

We will see that the assumption of a distribution for utility in the population and the finiteness of the set K leads to probability choice systems in which each possible resultant technology has a well-defined selection probability. To illustrate these concepts and elucidate their connection to other work we consider two examples.

Example one considers a choice between two alternative technologies for producing identical final services. Example two considers the choice among a continuum of technologies for producing identical final services, each technology available at a pre-specified price. These examples illustrate that the general ex-ante selection of techology will involve both discrete and continuous choices. Each example also suggests a natural cost minimization dual which takes service levels parametrically.

Example 1

Our first example assumes a one-period world in which consumers have the choice of two technologies for providing an identical end-use service. The isolated choice of a gas or an electric clothes dryer for providing a given service level, e.g., pounds of dry clothes per day, fits into this category.

Suppose that the alternative technologies are given by $Y_1^1 = f_1(x_1;\bar{a})$ and $Y_1^2 = f_2(x_2;\bar{a})$ with respective purchase prices of v_1 and v_2.

Vectors x_1 and x_2 represent inputs to the respective technologies and may be purchased at prices p_1 and p_2. The parameters $\bar{a}$ are assumed fixed in the short run and are independent of technology choice. Conditioning production on the parameters $\bar{a}$ in the function f corresponds to the notion of a restricted technology set used above. Note that the durable appliance technology is available in exactly two varieties in contrast to the classical stock-flow model where capital is assumed to be the input to household production.

We assume that preferences are representable by a single period utility function $U[Y_1,Y_2]$ where Y_1 is the end-use service level provided by either of the alternative technologies and Y_2 is a transferable numeraire or Hicksian commodity.

The consumer's decision problem is to make an <u>ex-ante</u> technology choice recognizing that <u>ex-post</u>, income I will be allocated between expenditures on input commodities and all other goods to achieve maximal utility in goods and services.

The indirect utility corresponding to the choice of the first technology is:

(15) $V[I - v_1, p_1;\bar{a}] = \max U[Y_1^1,Y_2]$ subject to:

$$Y_1^1 = f_1(x_1;\bar{a}) \quad \text{and} \quad p_1x_1 + v_1 + Y_2 \leq I$$

Similarly the indirect utility corresponding to the choice of the second technology is:

(16) $V[I - v_2, p_2;\bar{a}] = \max U[Y_1^2, Y_2]$ subject to

$$Y_1^2 = f_2(x_2;\bar{a}) \quad \text{and} \quad p_2x_2 + v_2 + Y_2 \leq I$$

In principle, indirect utility is conditioned on the utility and production functionals as well as the parameters $\bar{a}$. We have followed the usual convention in suppressing these arguments.

Consumers will choose technology 1 if and only if:

(17) $V[I - v_1, p_1;\bar{a}] \geq V[I - v_2, p_2;\bar{a}]$.

This implies that unconditional indirect utility is given by:

(18) $V^*[I - v_1, I - v_2, p_1, p_2;\bar{a}] = \max (V[I - v_1, p_1;\bar{a}, V[I - v_2, p_2;\bar{a}])$

In this example, ex-ante choice between technologies is discrete. Either technology 1 is purchased or technology 2 is purchased. This choice has an immediate income response through the purchase price v_j. In a multi-period model we will consider the financing aspects of durable purchase.

The budget set in final goods and services corresponding to the first technology is:

(19) $c_1 = \{(Y_1, Y_2)\varepsilon R_+^2 \mid Y_1 = f_1(x_1;\bar{a}) \ ; \ p_1 x_1 + Y_2 + v_1 \leq I \ ; \ x_1 \geqq 0\}$

When the production function $f_1(x_1;\bar{a})$ is invertible, (19) may be written:

(20) $c_1 = \{(Y_1, Y_2)\varepsilon R_+^2 \mid p_1 f^{-1}[Y_1;\bar{a}] + Y_2 \leqq I - v_1\}$

where $f_1^{-1}[Y_1;\bar{a}]$ denotes the assumed non-negative quantity of input x_1 necessary to produce service level Y_1 given the variable parameters $\bar{a}$.

Assume that the technology is smooth so that the marginal rate of substitution and its rate of change can be calculated on the boundary of c_1. From (20):

(21) $dY_2/dY_1 = -p_1/f_1'(x_1; \bar{a}) < 0$ and

(22) $$d/dx_1[dY_2/dY_1] = \frac{f''(x_1;\bar{a}) \ . \ p_1}{[f'(x_1;\bar{a})]^2} < 0$$

where we have assumed for convenience that f is strictly increasing and concave in its first argument and that p_1 is positive. The set c_1 is illustrated in Figure 1.

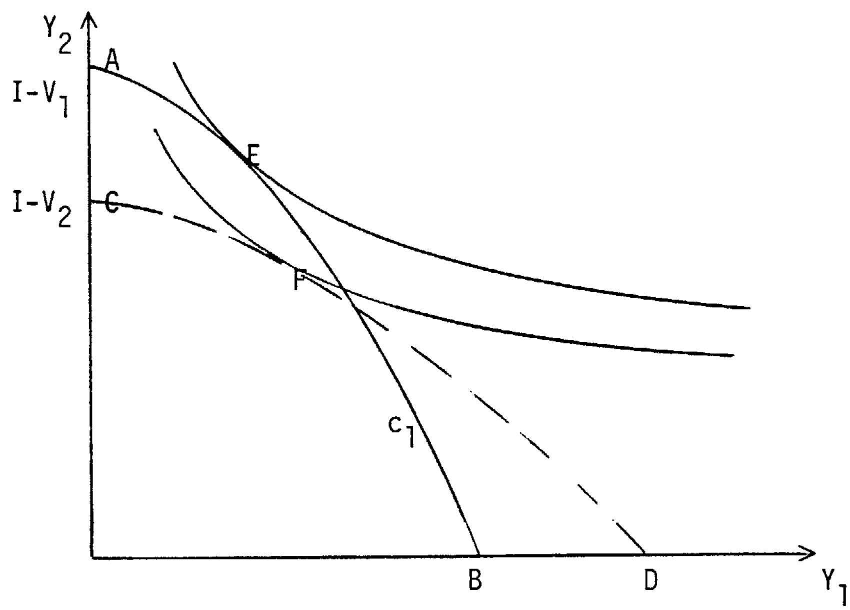


Figure 1

We assume that $f_1(0;\bar{a}) = 0$ so that zero utilization of the input commodity results in point A of the budget set. Strict convexity of the budget set is implied by (22). The budget set corresponding to the second technology is the area beneath the dotted line connecting points C and D. Figure 1 illustrates a situation in which maximal utility in final goods and services is achieved at points E and F corresponding to ex-ante choice of technologies 1 and 2 respectively. In this example, maximal utility would be achieved through choice of technology 1.

The indifference curves for utility at points E and F are drawn to reflect the necessary tangency conditions.

The Lagrangian for (15) (with multipliers λ_1 and λ_2) is:

$$(23)\quad L = U[Y_1^1, Y_2] + \lambda_1[Y_1^1 - f_1(x_1;\bar{a})] + \lambda_2[I - p_1x_1 - v_1 - Y_2]$$

The first-order conditions are:

(24) $L_{x_1} = -\lambda_1 f_1'(x_1;\bar{a}) - \lambda_2 p_1 = 0,$

(25) $L_{Y_2} = U_2[Y_1^1, Y_2] - \lambda_2 = 0,$ and

(26) $L_{Y_1^1} = U_1[Y_1^1, Y_2] + \lambda_1 = 0.$

Combining (24), (25), and (26) we obtain the tangency condition:

$$(27)\quad \frac{-U_1[Y_1^1, Y_2]}{U_2[Y_1^1, Y_2]} = \frac{\lambda_1}{\lambda_2} = \frac{-p_1}{f_1'(x_1;\bar{a})}$$

Equation (27) simply equates the marginal rate of substitution between end-use services, Y_1^1, and all other goods, Y_2, to the marginal cost of producing Y_1^1.

Equation (23) reveals that <u>Roy's identity continues to hold for input or "intermediate" goods</u>. Using the envelope theorem:

(28) $L_I = \lambda_2$ and

(29) $L_{p_1} = -x_1\lambda_2.$ From (28) and (29) we have:

$$(30)\quad \frac{-V_2[I - v_1, p_1]}{V_1[I - v_1, p_1]} = \frac{-L_2[I - v_1, p_1]}{L_1[I - v_1, p_1]} = x_1$$

Dubin and McFadden (1979) have used this result along with simple assumptions about technology to derive a consistent econometric choice and utilization system.

We have, thus far, assumed strict concavity of the production function $Y_1^1 = f(x_1;\bar{a})$ which implies the strict convexity of budget constraint set c_1. When the production function is in fact linear in x_1, the fixed-coefficient technology results. In this case the boundary of c_1

is flat and we may define a service price for end-use consumption which is constant. Furthermore, linearity in the input good x_1 insures that the <u>average efficiency of production</u> defined by the service level achieved per quantity of input utilized is constant.

The appropriate extension of the concept of average efficiency to cases in which production exhibits decreasing returns to scale is the notion of marginal efficiency. We define the <u>marginal efficiency of production</u> resulting from input x as the marginal product of x conditioned on all variable design parameters. This definition implies that the electrical efficiency of providing cooling-degree hours of air-conditioning will depend on climate, usage levels, insulation, capacity of the air-conditioning unit, etc. The quantity $p_1/f_1'(x_1;\bar{a})$ in (27) may be interpreted as the end-use service price for Y_1^1. We see that the end-use service price or marginal cost of Y_1^1 is the price of input commodity x_1 divided by the marginal efficiency of x_1.

This example has considered the choice of alternative technologies with fixed purchase prices for production of an identical end-use service. Our next example considers a similar choice situation but allows service price to vary according to the selection of certain fixed design parameters.

<u>Example 2</u>

Let U[Y] denote the single-period utility derived from consumption of service level Y. Suppose that the technology for Y is given by $Y = f[x;K]$. For simplicity we assume that Y, x, and K are scalars where x represents an input commodity and K represents a fixed design parameter. In the light bulb example, K might be interpreted as a measure of durability, or K might measure an upper limit to cooling capacity or efficiency level for an electric air conditioner. The fixed design component determines the purchase price within the function H[K]. The function H[K] is assumed known in this example but in practice would be estimated from engineering and marketing data.

The consumer's problem is to distribute income, I, optimally between the initial purchase price H[K] and operating cost to achieve maximal utility. This problem can be formulated as:

(31) $\max U[Y]$ subject to $Y = f[x; K]$ and $px + H[K] \leqq I$, which is clearly equivalent to:

(32) $\max U[f[x; K]]$ subject to $px \leqq I - H[K]$

Maximization of (31) conditional on K yields indirect utility

$U[f[I - H(K))/p]; K])$.

Total utility is then $\max_K U(f[(I - H(K))/p ; K])$ which leads to the following first order condition:

$$(33) \quad \frac{f_2[(I - H(K))/p ; K]}{f_1[(I - H(K))/p ; K]} = \frac{H'(K)}{p}$$

From (33) or by inspection one finds that (32) is clearly the dual to the minimization problem:

(34) $\min \; [H(K) + px]$ subject to $Y^o \geqq f(x;K)$ where Y^o represents a pre-chosen service level. The duality between the maximization problem in (32) and the minimization problem in (34) is a consequence of the monotonic transformation of the production function f by the utility function U. The duality exhibited in this example illustrates a deeper issue of separability to be confronted in Section IV.

We consider two specializations of this example which are easily illustrated. Suppose first that $H(K) = rK$ where r is interpreted as the price of attribute K. The maximization problem in (32) is illustrated in Figure 2 where the indifference surface denoted by $\tilde{U}$ is given by:

$$\tilde{U} = \{(x,K) \quad U[f(x;K)] = c\}$$

for some constant level of indirect utility c. The budget set, B, is given by the area below the line $p \cdot x + rK = I$.

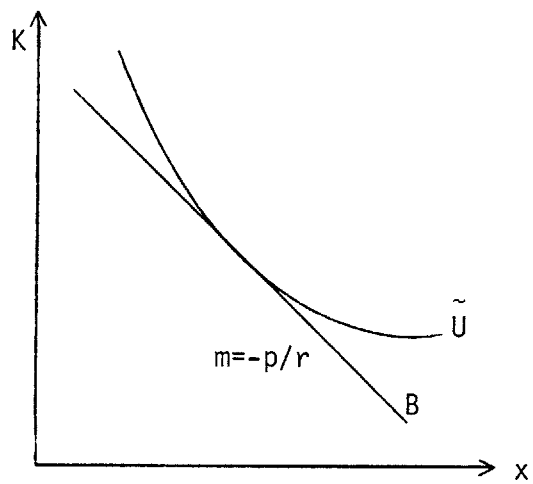


Figure 2

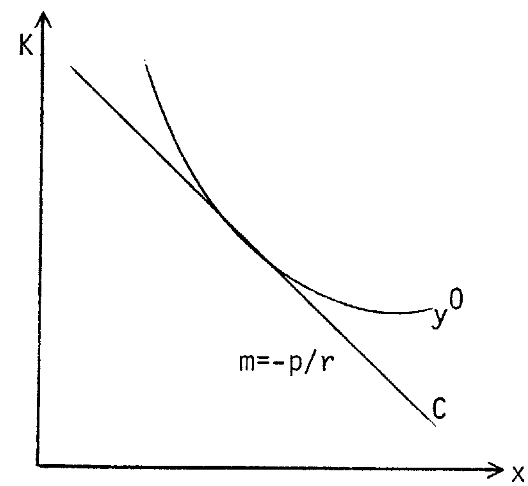


Figure 3

Figure 3 similarly illustrates minimization of isocost, $c = p \cdot x + rK$, subject to the isoquant determined by $y^0 = f(x;K)$. Tangencies in Figures 2 and 3 represent first-order condition (33).

Hartman's (1979) adapatation of Hausman's (1979) theoretical framework considers precisely the minimization problem: $\min\ (p \cdot x + rK)$ subject to $y^0 = f(x;K)$. Hartman specifies the service demand y^0 as a function of exogenous variables and an efficiency adjusted price for fuel input. His methodology, however, begs the separability issues which allow a formal two-stage consistent budgeting decision to be made.

Our second specialization of the maximization problem (31) assumes that the production function $f(x;K)$ has the form $\rho(K)x$. We assume that $\rho(\cdot)$ is positive and strictly increasing in K. Note that f now exhibits a marginal efficiency which is independent of x yet depends explicitly on the fixed design parameter K. Equation (31) is then equivalent to:

(35) $\max U[Y]$ s.t. $(p/\rho(K))Y \leq I - H[K]$

We may write the indirect utility from (35) as $V[I-H(K), p/\rho(K)]$ to underscore a direct trade-off between operating and "capital" costs.

If we let $\tilde{H} = H[K]$ and $\tilde{p} = p/\rho(K)$ then $V^*[\tilde{H},\tilde{P}] = V[I-\tilde{H}, \tilde{p}] = V[I-H(K), p/\rho(K)]$ defines the indirect utility when purchase price is H and service price is p. Figure 4 depicts a level set of the function V*.

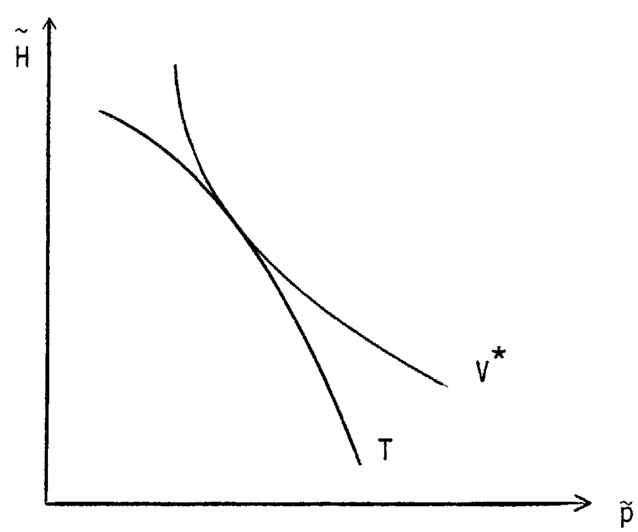


Figure 4

The curvature and slope of the indifference locus in Figure 4 follow by application of Roy's identity and the Slutsky equation. Specifically, the slope of the indifference locus is:

$$(36)\quad \frac{d\tilde{H}}{d\tilde{p}} = \frac{V_2[I-\tilde{H},\ \tilde{p}]}{V_1[I-\tilde{H},\ \tilde{p}]} = -Y[I-\tilde{H},\ \tilde{p}] < 0$$

where the second equality is a consequence of Roy's identity. From equation (36) we have:

$$(37)\quad \frac{d}{d\tilde{p}}\frac{d\tilde{H}}{d\tilde{p}} = Y_1\frac{d\tilde{H}}{d\tilde{p}} - Y_2 = -[Y_1 Y + Y_2] \geq 0$$

where $Y_1 Y + Y_2$ is equivalent to the Hick's compensated price derivative of $Y[I-\tilde{H},\tilde{p}]$ by Slutsky's equation and is therefore nonpositive.

The trade-off between purchase price $\tilde{H} = H[K]$ and $\tilde{p} = p/\rho(K)$ is illustrated by the locus T in Figure 4. The slope of this locus at a point $(\tilde{p},\ \tilde{H})$ is negative if we assume that purchase price is increasing in the attribute K;

$$\frac{d\tilde{H}}{d\tilde{p}} = \frac{d\tilde{H}}{dK} \Big/ \frac{d\tilde{p}}{dK} \quad \text{implies:}$$

$$(38)\quad \frac{d\tilde{H}}{d\tilde{p}} = \frac{-H'(K)\ (\rho(K))^2}{p\ \rho'(K)} < 0 \quad \text{as} \quad H'(K) > 0.$$

The curvature of the locus T will depend on the derivatives of the functions H and ρ and is drawn convex to the origin for illustration only. Note that increasing utility is represented by indifference loci nearer the origin while the feasible price space is determined by the <u>unbounded</u> area above the locus T. It is easy to verify that equating the derivatives (36) and (38) reproduces first-order condition (33) under the maintained assumption $f[x;\ K] = \rho(K)\cdot x$.

Figure 4 suggests a motivation for a dual cost minimization problem which is implicit in the approach of Hirst and Carney (1978):

$$(39)\quad \min_{\tilde{p},\tilde{H}} (\tilde{p}\, y^{0} + \tilde{H}) \quad \text{subject to} \quad (\tilde{p}, \tilde{H}) \in T$$

where y^{0} denotes a predetermined service level.

One may easily verify that (39) produces the first-order condition (33). The minimization problem (39) is illustrated in Figure 5. We have followed the convention of drawing the locus T concave to the origin. A sufficient condition for this curvature is increasing marginal purchase costs as (38) implies:

$$(40)\quad \frac{d}{dK}\left(\frac{d\tilde{H}}{d\tilde{p}}\right) = \frac{-1}{p} \quad \frac{\rho'[H'2\rho\rho' + \rho^{2}H''] - H'\rho^{2}\rho''}{(\rho')^{2}} < 0$$

when $H''(K) > 0$.

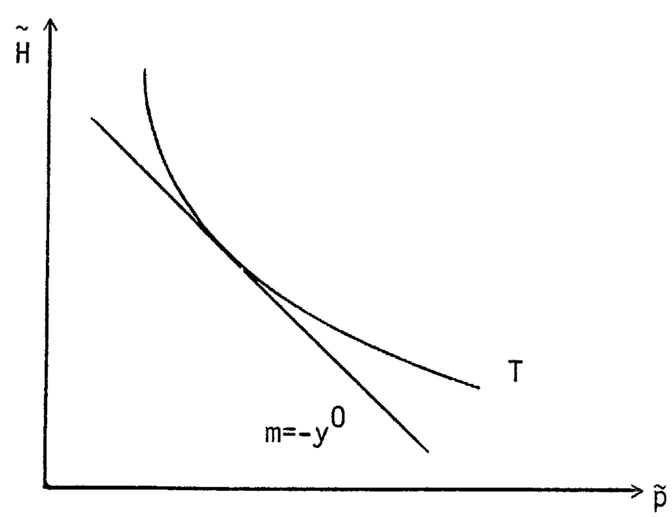


Figure 5

These examples have illustrated how the consumers durable choice problem can be represented in terms of the optimal choice of technology subject to financial and technological constraints. In the next section we derive conditions under which the separability in utility implied by appliance-production technologies permits a consistent two-stage or "tree" budget program. Under the two-stage budgeting procedure, consumers first determine optimal production service levels and then choose input combinations which produce these service levels at minimum cost.

IV. Appliance Technology and Two-Stage Budgeting

The examples pressented in section III make clear the observation that household energy demand is a derived demand for basic fuel inputs to appliance technologies. Conditions under which the optimal allocation of inputs to appliance technologies may be separated by appliance type are now examined The separability condition has very strong implications for the form of the production technology and for the final form of the indirect utility function.

Intertemporal Separability

We consider the intertemporal utility maximization problem allocating intermediate goods to the production of final services, over some fixed horizon L. For convenience, we assume that utility, $U = U[U_1, U_2, \ldots, U_L]$, is weakly intertemporally separable with U_t being the utility of goods and services in period t.

The intertemporal utility maximization problem allocates wealth W among the L periods to:

(41) maximize $U = U[U_1, U_2, \ldots, U_L]$ subject to $\sum_{t=1}^{L} E_t(U_t, P_t) \leq W$ where

$E_t(U_t, p_t)$ = present discounted value of the minimum expenditure necessary to achieve utility level U_t at price P_t.

The demand for goods and services in period t will in general depend on all prices p_1, $p_2, \ldots, p_L$ and wealth W. To achieve demand separability, one must either solve a broad group allocation problem which determines total expenditure in each period or else assume that budget constraints between expenditure groups are set exogeneously. When the intertemporal allocation problem can be solved using appropriate temporal price indices a perfect aggregation solution is said to exist.

Gorman (1959) determined the necessary and sufficient conditions for perfect aggregation such that the consumer need not know the actual prices of the individual goods in order to carry out his preliminary allocation, as long as he knows the values of the price indices and his own income. The existence of unconditional group price indices requires that the utility function be homothetically separable or strongly separable in Gorman polar form.

An implication of the Gorman proposition noted by Blackorby, Lady, Nissen, and Russel (1970) is that when the utility function is homothetically separable perfect aggregation implies a consistent two-stage budgeting procedure.

In the first stage, consumers solve:

(42) $\max U[U_1, U_2, \ldots, U_L]$ subject to

$$\sum_{t=1}^{L} P_t(p_t) \cdot U_t \leqq W$$

Note that (42) has the usual form of utility maximization subject to a budget constraint with U_t interpreted as quantities purchased at prices $P_t(p_t)$. The second stage uses the quantities U_t implied by (42) to define broad temporal expenditures $I_t = U_t \cdot P_t(p_t)$. Second stage commodity demand satisfies:

(43) $\max U_t(x_t)$ subject to $p_t \cdot x_t \leqq I_t$

Gorman posed the perfect aggregation problem for allocation of expenditure among broad groups of commodities within a single period. We have applied his result in the context of allocaton of inter-temporal expenditure.

For the present, we follow Hausman (1979) and assume that expenditure levels are pre-determined. Demand for goods and services within a period thus become a function only of prevailing prices and expenditure.

Household Production and Separability[6]

We write the utility function in (41) as:

$$(44) \quad U(x) = U[f_1^{i_1}(x_1; \bar{a}),\ f_2^{i_2}(x_2; \bar{a}), \ldots,\ f_n^{i_n}(x_n; \bar{a}),\ x_{n+1}]$$

where:

$f_j^{i_j}(x_j; \bar{a})$ = production of end-use service Y_j by technology type i_j, $j = 1,2,\ldots n$;

x_j = vector of input commodities for production of end-use service j, $j = 1,2,\ldots n$;

$\bar{a}$ = vector of variable parameters;

x_{n+1} = vector of non-produced commodities.[7]

Equation (44) assumes that utility is weakly separable between the end-use service commodities Y_j, $j = 1, 2,\ldots, n$ and all other goods x_{n+1}. The index i_j represents a particular technology type for the production of end-use service Y_j. Note that the production functions $f_j^{i_j}(x_j; \bar{a})$ generically separate the commodities x_j for $j = 1, 2,\ldots,n$. The partition is termed generic because the same physical commodity is often an input for several distinct technologies. This interpretation regards electricity used as an input to clothes drying as distinct from electricity used as an input for space heating yet both inputs are priced identically. Total electricity demanded is the sum of electricity demanded in each end-use. We suppose that the input commodities x_j are available at prices p_j and that p_{n+1} is the price vector for all other goods x_{n+1}. The budget constraint for traded commodities is:

$$(45) \quad \sum_{j=1}^{n} p_j x_j + p_{n+1} x_{n+1} \leqq I$$

where I denotes pre-determined total expenditure for the given period.

Conditional on the choice of technologies (e.g., $i_j = i_j^0$, $j = 1, 2, \ldots n$) consumers must allocate resources to maximize (44) subject to (45). Let $c_j^{i_j}(Y_j, p_j; \bar{a})$ be the cost function dual to the production function $f_j^{i_j}(x_j; \bar{a})$. We can recast the optimization problem using the cost functions as:

$$(46) \quad \max u[Y_1, Y_2, \ldots, Y_n, x_{n+1}]$$

$$\text{subject to} \quad \sum_{j=1}^{n} c_j^{i_j}(Y_j, p_j; \bar{a}) + p_{n+1} x_{n+1} \leqq I$$

By direct analogy to Gorman's proposition, we see that necessary and sufficient conditions for a consistent two-stage budgeting solution to (46) in which consumers first determine optimal service levels and then choose input combinations which produce these service levels at minimum cost require that production be homothetic. A stronger condition, employed by Muellbauer (1974) and Pollak and Wachter (1975), assumes that the production technologies exhibit constant returns to scale. For the purposes of this discussion we adopt this assumption but note that the essential features of the argument are unchanged provided a new utility indicator is defined which is consistent with renormalized production functions.[8]

Under constant returns to scale in production the cost functions have the simple form $c_j^{i_j}(Y_j, p_j; \bar{a}) = \theta_j^{i_j}[p_j; \bar{a}] \cdot Y_j$, where the unit cost functions $\theta^{i_j}[\cdot; a]$ are perforce linearly homogeneous.

The optimization problem in (46) becomes:

(47) $\max U[Y_1, Y_2, \ldots, Y_n, x_{n+1}]$

$$\text{subject to } \sum_{j=1}^{n} \theta_j^{i_j}[p_j; \bar{a}] \cdot Y_j + p_{n+1} \cdot x_{n+1} \leqq I$$

from which indirect utility is:

(48) $V[\theta_1^{i_1}(p_1; \bar{a}), \theta_2^{i_2}(p_2; \bar{a}), \ldots, \theta_n^{i_n}(p_n; \bar{a}), p_{n+1}, I]$

where V is dual to U in (44).

We see from (48) that indirect utility satisfies a price partition which corresponds to the commodity partition assumed in (44). The crucial element of the derivation is that the utility function U in (44) is homothetically separable in appliance technologies.

The functions $\theta_j^{i_j}(p_j; \bar{a})$ have a straightforward interpretation as the unit costs of producing end-use service j. As the notation reflects, the unit costs or service prices will depend on choice of technology type (i_j) and all variable factors $\bar{a}$. By Shephard's Lemma we can determine optimal input factors from the gradient of the cost function:

(49) $x_j = [\partial\theta_j^{i_j}(p_j; \bar{a})/\partial p_j] \cdot Y_j$

Equation (49) demonstrates that the input to service ratios x_{jk}/Y_j

for input k, are independent of service level. Let V_j and V_I denote the derivatives of (48) by the j-th service price and by income respectively. Roy's identity applied to (48) determines optimal service levels in the first stage of the two-stage budget procedure:

$$(50)\quad Y_j = \frac{-V_j[\theta^{i_1}(p_1;\bar{a}),\ldots,\theta^{i_n}(p_n;\bar{a}),p_{n+1},I]}{V_I[\theta^{i_1}(p_1;\bar{a}),\ldots,\theta^{i_n}(p_n;\bar{a}),p_{n+1},I]}$$

To derive the total demand for a given input, we use (49) and (50) to determine input utilization by end-use and then sum across end-uses. Suppose, by way of an example, that each technology uses electricity and that the price of electricity appears as an argument in the functions $\theta^{i_j}(p_j;\bar{a})$. Total demand for electricity, x_e, satisfies:

$$(51)\quad x_e = -\sum_{j=1}^{n} \frac{\partial\theta_j^{i_j}(p_j;\bar{a})}{\partial p_e}\cdot\left(\frac{V_j}{V_I}\right)$$

Equation (51) exemplifies the conditional structure of energy demand. Econometric estimation of (51) must recognize the endogeniety of appliance technology selection. Consider a special case of the generalized Gorman polar form for indirect utility:

$$(52)\quad V[\tilde{p},I] = \frac{I-a(\tilde{p})}{b(\tilde{p})}$$

Application of Roy's identity to (52) yields:

$$(53)\quad Y_i = a_i(\tilde{p}) + \frac{b_i(\tilde{p})}{b(\tilde{p})}\cdot(I-a(\tilde{p}))\quad \text{where}$$

(54) $a_i(\tilde{p}) = \frac{\partial a(\tilde{p})}{\partial \tilde{p}_i}$ and $b_i(\tilde{p}) = \frac{\partial b(\tilde{p})}{\partial \tilde{p}_i}$

Note that (53) implies linear Engel curves which do not pass through the origin. From equations (51) and (53), electricity demand satisfies:

$$(55) \quad x_e = \sum_{j=1}^{n} \frac{\partial \theta_j^{i_j}(p_j;\bar{a})}{\partial p_e} \, a_j(\tilde{p}^{i*}) + \frac{b_j(\tilde{p}^{i*})}{b(\tilde{p}^{i*})} \cdot (I-a(\tilde{p}^{i*}))$$

where $\tilde{p}^{i*} = [\theta_1^{i_1}(p_1;\ \bar{a}),\ \theta_2^{i_2}(p_2;\ \bar{a}), \ldots,\ \theta_n^{i_n}(p_n;\ \bar{a}),\ p_{n+1}]$

and where $i* = (i_1, i_2, \ldots, i_n)$ indexes a given portfolio of technologies. A Gorman form for indirect utility in each period and strong intertemporal separability imply that the two-level budgeting procedure can be executed over the L-period time horizon using intertemporal price indices. An example of an indirect utility function exhibiting strong intertemporal separability is:

$$(56) \quad V* = \sum_{t=1}^{L} \delta_t V[p_t^{i*},\ I_t] = \sum_{t=1}^{L} \delta_t G_t[<I_t/p_{jt}^{i*}>]$$

where the parameter δ_t measures the individual's discount rate. Roy's identity applied to (56) demonstrates that service demand in period t is solely dependent on prices and income in period t and independent of the parameter δ_t.

We now consider the financing of durable purchases. Assume that appliance portfolio, i* is purchased as price H^{i*}. Let W_t denote income in period t. Expenditure I_t in (56) must satisfy the inequality:

$$(57) \quad \sum_{t=1}^{L} \frac{W_t}{(1+R_t)} - H^{i*} \geq \sum_{t=1}^{L} \frac{I_t}{(1+R_t)}$$

where R_t is the t-period discount rate.

Suppose that purchase price, H^{i*}, is allocated to each of the L periods in equal amounts, X, and that the one-period discount rate is identical across periods with $(1+R_t) = (1+R)^t$. Then:

(58) $$\sum_{t=1}^{L} X/(1+R)^t = H^{i*}$$ implies:

(59) $$X = \left(\frac{1-(1+R)^{-L}}{1-(1+R)^{-1}}\right)^{-1} \cdot (1+R) \cdot H^{i*}$$

The economic theory of durable choice does not imply a specific payment plan for amortizing purchase price. This suggests the use of a flexible functional form in discount factors, socio-economic variables, initial purchase price etc., to predict per period payments. Specific payment schemes such as (59) are then testable through appropriate parameter restrictions. In the next section, we investigate econometric specifications for models of durable choice and utilization.

V. Econometric Specification for Models of Durable Utilization

We presented in Section IV a two-level utilization procedure in which service levels, Y_j, are determined by equation (50) and optimal input combinations required to produce Y_j are detemined by (49). Econometric specification for this system requires explicit functional forms for indirect utility, V, and for service levels Y_j. As Roy's identity connects V with Y_j through (50), it is often possible to specify a parametric form for demand and then solve a partial differential equation to find a compatible indirect utility function. This methodology has been successfully applied by Hausman (1979, 1981), Burtless and Hausman (1978), and Dubin and McFadden (1979) for individual demand equations. We now consider the recovery of an indirect utility function from a system of demand equations as required by (50). We follow Dubin and McFadden (1979) and assume that demand is linear in real income I and additive with a function of real prices:

(60) $Y_j = \beta_j I + m_j(p_1, p_2, \ldots p_n) \quad j = 1, 2, \ldots, n$

By Roy's identity we may write the first equation in this system as:

(61) $-\partial V/\partial p_1/\partial V/\partial I = \beta_1 I + m_1(p_1, p_2, \ldots, p_n)$

We apply the implict function theorem and write (61) in differential form as:

(62) $-[\beta_1 I + m_1(p_1, \ldots, p_n)]dp_1 + dI = 0$

Application of the integrating factor $\mu(p_1, p_2, \ldots, p_n, I) = e^{-\beta_1 p_1} \cdot g(p_2, \ldots p_n)$ transforms (62) into an exact differential equation with solution:

(63) $V(p_1, p_2, \ldots, p_n, I) = e^{-\beta_1 p_1} \cdot g(p_2, \ldots, p_n)[I + M(p_1, \ldots, p_n)]$

where:

(64) $M(p_1, p_2,\ldots,p_n) = \int_{p_1} e^{\beta_1(p_1-t)}m_1(t,p_2,\ldots,p_n)dt$

Note that (64) satisfies:

(65) $\partial M/\partial p_1 - \beta_1 M = -m_1$

Roy's identity applied to (71) for the second commodity implies:

(66) $Y_2 = -\partial V/\partial p_2/\partial V/\partial I$

$$= \frac{-e^{-\beta_1 p_1} g(p_2,\ldots,p_n)M_{p_2} - e^{-\beta_1 p_1}[I+M]g_{p_2}}{e^{-\beta_1 p_1} g(p_2,\ldots,p_n)}$$

$$= -M_{p_2} - [I+M]g_{p_2}/g \quad \text{where} \quad M_{p_2} = \partial M/\partial p_2$$

Comparing (66) with (60) we must have $-g_{p_2}/g = \beta_2$ and $-M_{p_2} + \beta_2 M = m_2(p_1,\ldots,p_n)$. Proceeding similarly for commodities $j = 3,\ldots,n$ we find:

(67) $V(p_1, p_2,\ldots,p_n,I) = (e^{-\Sigma\beta_j p_j})[I + M(p_1,p_2,\ldots,p_n)]$

where the function M satisfies the restrictions:

(68) $\beta_j M - M_{p_j} = m_j$ for $j = 1, 2,\ldots,n$.

The restrictions in (68) imply a relationship among the m_j which must be satisfied if (67) is consistent with (60). These restrictions are identical to symmetry of the Slutsky substitution matrix as we now demonstrate. Consider (68) for $j = 1,2$:

(69) $\beta_1 M - M_{p_1} = m_1 \rightarrow e^{-\beta_1 p_1}(\beta_1 M - M_{p_1}) = e^{-\beta_1 p_1} \cdot m_1$ and

(70) $\beta_2 M - M_{p_2} = m_2 \rightarrow e^{-\beta_2 p_2}(\beta_2 M - M_{p_2}) = e^{-\beta_2 p_2} \cdot m_2$

From (69) and (70) we have:

(71) $\partial/\partial p_1[e^{-\beta_1 p_1} \cdot M] = -e^{-\beta_1 p_1} \cdot m_1$ and

(72) $\partial/\partial p_2[e^{-\beta_2 p_2} \cdot M] = -e^{-\beta_2 p_2} \cdot m_2$ from which follow

(73) $\partial/\partial p_1[e^{-\beta_1 p_1 - \beta_2 p_2} \cdot M] = -e^{-\beta_1 p_1 - \beta_2 p_2} \cdot m_1$ and

(74) $\partial/\partial p_2[e^{-\beta_1 p_1 - \beta_2 p_2} \cdot M] = -e^{-\beta_1 p_1 - \beta_2 p_2} \cdot m_2$

Equating the mixed partials of (73) and (74) we have:

(75) $\partial/\partial p_2 [e^{-\beta_1 p_1 - \beta_2 p_2} \cdot m_1] = \partial/\partial p_1[e^{-\beta_1 p_1 - \beta_2 - p_2} \cdot m_2]$ or

(76) $\partial m_1/\partial p_2 - \beta_2 m_1 = \partial m_2/\partial p_1 - \beta_1 m_2$

By Slutsky symmetry we have:

(77) $\partial Y_1/\partial p_2 + \partial Y_1/\partial I \cdot Y_2 = \partial Y_2/\partial p_1 + \partial Y_2/\partial I \cdot Y_1$ which implies

(78) $\partial m_1/\partial p_2 + \beta_1 Y_2 = \partial m_2/\partial p_1 + \beta_2 Y_1$ or

(79) $\partial m_1/\partial p_2 + \beta_1[m_2 + \beta_2 I] = \partial m_2/\partial p_1 + \beta_2 [m_1 + \beta_1 I]$ so that

(80) $\partial m_1/\partial p_2 + \beta_1 m_2 = \partial m_2/\partial p_1 + \beta_2 m_1$

Comparing (76) to (80) we find that conditions (68) are equivalent to symmetry of the substitution matrix. Additional integrability restrictions (homogeneity, summability, non-negativity, and negative quasi semi-definiteness) are imposed on M by the requirement that $V(p_1, p_2, \ldots, p_n, I)$ be an indirect utility function.

Equation (67) is a member of the generalized Gorman polar family as can be seen from (52). In this case the demand for electricity in (55) has the form:

$$(81) \quad x = \sum_{j=1}^{n} \psi_e^{i_j}[\beta_j I + m_j(\tilde{p}^{i*})] \quad \text{where}$$

$$\psi_e^{i_j} = \partial\theta_j^{i_j}(p_j; \bar{a})/\partial p_e.$$

Recall that $\psi_e^{i_j}$ are the derivatives of the unit costs of producing end-use service j with respect to the price of electricity conditioned on discrete choice of durable i_j. The $\psi_e^{i_j}$ may be linear-in-parameter expressions in weather and appliance characteristics as well as the relevant set of input prices. An alternative form for the service equation is:

$$(82) \quad Y_j = a_j I/p_j^{i*} \quad \text{which implies the form:}$$

$$(83) \quad x_e = \sum_{j=1}^{n} \psi_e^{i_j} \cdot (a_j/p_j^{i*})I \quad \text{for electricity demand.}$$ This form is restrictive in the modeling of service demand. A less restrictive system is generated under the assumption that V is given by the linearly homogeneous translog form so that:

$$(84) \quad Y_j = (I/p_j^{i*}) \quad a_j + \sum_{k=1}^{K} b_{kj} \log(p_k^{i*}/I) \quad \text{so that electricity}$$

demand becomes:

$$(85) \quad x_e = \sum_{j=1}^{n} \psi_e^{i_j}(I/p_j^{i*}) \; a_j + \sum_{k=1}^{K} b_{kj}\log(p_k^{i*}/I)$$

Whichever specification is chosen, attention must be given to the placement of random components. One approach assumes that the demand equations at the various levels represent the behavior of the average individual. Deviations from the average may be represented by assuming a distribution for the behavioral parameters; estimation should enforce this assumption throughout the equation system. A simpler technique assumes that all random deviations from average behavior are captured in an additive stochastic disturbance. Finding indirect utility functions which are compatible with partial demand systems with additive disturbances is not always feasible. Dubin and McFadden (1979) have had success with the Gorman polar form to which we now return.

Suppose we modify the Gorman form (52) as

(86) $$V[p, I] = \frac{(I - a(p) + \xi_1/\theta)}{(b(p) + \xi_2)} + \xi_3$$

where ξ_1, ξ_2, and ξ_3 are random components. Roy's identity implies:

(87) $$Y_j = a_j(p) + \frac{(I - a(p) + \xi_1/\theta)b_j(p)}{(b(p) + \xi_2)}$$

If $b_j(p) = 0$ and $a(p) = a_o(p) + \sum_{j=1}^{n} p_j n_j$ then

(88) $$Y_j = a_{oj}(p) + n_j.$$

Equation (88) exhibits the additive disturbance structure when n_j is interpreted as a random component but is limited in its applicability due to the absence of income effects. If $b_j(p) \neq 0$ then (87) will be rather inconvenient for linear estimation techniques unless $\xi_2 \equiv 0$ and $b_j(p)/b(p) = \beta_j$, β_j constant. Under these assumptions, (87) implies:

(89) $$Y_j = a_{oj}(p) + \beta_j I - \beta_j a(p) + \beta_j \xi_1/\theta + n_j$$

From equation (55) electricity demand satisfies:

$$(90)\quad x_e = \sum_{j=1}^{n} \psi_e^{ij}[a_{Oj}(p) - \beta_j a(p) + \beta_j I + \beta_j \xi_1/\theta + \eta_j]$$

$$= \sum_{j=1}^{n} \psi_e^{ij}[a_j(p) - \beta_j a(p) + \beta_j I] + \sum_{j=1}^{n} \psi_e^{ij}\beta_j \xi_1/\theta + \sum_{j=1}^{n}\psi_e^{ij}\eta_j \quad \text{so that:}$$

$$(91)\quad x_e = \sum_{j=1}^{n} \psi_e^{ij}[m_j(p) + \beta_j I] + \xi_1^* \quad \text{where} \quad \xi_1^* = \xi_1 + \sum_{j=1}^{n}\psi_e^{ij}\eta_j \quad \text{and}$$

$$\theta = \sum_{j=1}^{n}\psi_e^{ij}\beta_j$$

We now consider the joint estimation of durable choice and utilization. However, we relax the assumption that the additive error component ξ_1 in (91) appears consistently in intertemporal utility and suppose instead that random variations in intertemporal utility, V*, are summarized through an additive disturbance ε^{i*} whose distribution depends on the chosen portfolio i*.

Suppose that intertemporal utility is given by V* in (56) and V in (52). Then:

$$(92)\quad V_{i*}^* = \sum_{t=1}^{L} \delta_t \, e^{-\Sigma\beta_j p_{jt}^{i*}} \; (I_t - a(p_t^{i*})) + \varepsilon^{i*}$$

The probability that portfolio i* is chosen satisfies:

$$(93)\quad P_{i*} = \text{Prob}[V^*_{i*} \geq V^*_{j*},\ j* \neq i*]$$

$$= \text{Prob}[W^{i*} + \varepsilon^{i*} \geq W^{j*} + \varepsilon^{j*},\ j \neq i*]$$

$$= \text{Prob}[\varepsilon^{j*} - \varepsilon^{i*} \leq W^{j*} - W^{i*},\ j \neq i*] \qquad \text{where}$$

$$(94)\quad W^{i*} = \sum_{t=1}^{L} \delta_t [e^{-\Sigma\, \beta_j p^{i*}_{jt}}]\,(I_t - a(p^{i*}_t))$$

Finally, demand for electricity from (91) satisfies:

$$(95)\quad x_{et} = \sum_{j=1}^{n} \psi^{i_j}_{et}\,[m_j(p^{i*}_t) + \beta_j I_t] + \xi^*_{1t}$$

for t = 1, 2, ..., L.

Estimation of the system (93) combined with (95) should account for the endogeneity of variables indicating portfolio choice i*. For a detailed review and comparison of available estimation techniques the reader may consult Appendix II, Appendix III, and Dubin and McFadden (1979).

VI. Summary and Conclusions

This chapter has developed a theoretical and econometric framework for analyzing durable choice and utilization. After identifying the essential feature of any model in durable choice behavior, we formulated an ex-ante ex-post utility maximization model which incorporates the aspects of discrete choice, household production, and the trade-off between operating and capital costs. We then illustrated how the theoretical model could be translated into an estimable econometric system. Empirical implementation of the model will, among other things, permit calculation of the time path of energy conservation resulting from alternative economic policies such as mandatory building standards, appliance efficiency standards, or energy price regulation. Slight modification of the model will enable one to rigorously analyze particular issues that relate to the choice and utilization of other durables such as automobiles.

Footnotes

1. The author gratefully acknowledges the very useful comments of his colleagues, Tom Cowing, Peter Navarro, Rhonda Williams, Nigel Wilson, and Cliff Winston.

2. Note that capital market imperfections limit the availability of financing for new purchases due to equity requirements and the dependence of i on the level of borrowing.

3. See McFadden (1974).

4. See McFadden and Dubin (1981) for a detailed account of the construction of a thermal load model for single-family residences.

5. The ex-ante ex-post decision framework is considered in the context of optimal plant design by Fuss and McFadden (1978).

6. See Becker (1965) and Muth (1966) for alternative characterization of the household as a production unit.

7. We drop the subscript t to avoid excessive notation.

8. A production function $f(x)$ is homothetic when $f(x) = g(h(x))$ g monotonic and h linearly homogeneous. If the utility function is given by $U[f(x)]$ then $\tilde{U}(Z) = U[g(Z)]$ is consistent with the linearly homogeneous function $Z = h(x)$.

Chapter II

RATE STRUCTURE AND PRICE SPECIFICATION IN THE DEMAND FOR ELECTRICITY

Recent studies in the demand for electricity have raised again the question of price specification. The early work of Houthakker (1951a) discussed demand subject to a quantity dependent rate structure as compared to the classical situation of parametrically given prices. Taylor (1975), in his survey of the electricity demand literature, reviews the rate structure problem and indicates a simple procedure which converts the complex optimization problem of the consumer to the standard case of a linear budget constraint set in marginal prices. Modifications to the Taylor (1975) procedure were noted by Berndt (1978) and Nordin (1976).

A behavioral question is whether consumers can detect prevailing marginal rates in the presence of automatic appliances and billing cycle variations. An alternative hypothesis suggests that consumers respond to a summarizing statistic for the quantity dependent rate structure such as average price.

This chapter reviews the theory of price specification and considers the comparative static analysis of demand subject to a declining block rate schedule. We further investigate the statistical endogeneity of prices whose construction requires utilization of the observed consumption level, and determine price specification within a sample of 744 households surveyed in 1975 by the Washington Center for Metropolitan Studies.[1] We finally consider the construction of marginal prices using the WCMS data and monthly billing data from the recent National Interim Energy Consumption Survey (NIECS) of 1978.

II. Specification of Price: Theory

1. Quantity Dependent Rate Structures

We begin by reviewing the case of a declining block rate structure and derive a simple relation among quantity, average price, marginal price, and the rate structure premium. Let B be the total expenditure on electricity and Q the amount of electricity consumed. A typical rate structure has the form:

$$B = C \qquad \text{for } 0 \leq Q \leq X_1$$

$$B = C + \pi_1(Q-X_1) \qquad \text{for } X_1 < Q \leq X_2$$

$$B = C + \sum_{j=1}^{r-1} (X_{j+1} - X_j)\pi_j + \pi_r(Q - X_r) \qquad \text{for } X_r < Q \leq X_{r+1},\ 1 < r \leq n$$

where X_i denote the lower block boundaries and where we have set $X_{n+1} = +\infty$. The constant C is the connect charge and π_j is the price of electricity in block j. Suppose measured consumption, Q*, lies in the rth block so that $X_r < Q^* \leq X_{r+1}$ and total expenditure, B*, is

$$C + \sum_{j=1}^{r-1} (X_{j+1} - X_j)\pi_j + \pi_r(Q^* - X_r).$$ We then define the measured

average price as B*/Q*, the measured marginal price as π_r, and the rate structure premium (RSP) as the difference between total expenditure and the cost of purchasing the quantity Q* at the marginal rate π_r. $RSP = B^* - \pi_r Q^*$. Dividing by quantity we obtain the simple relation average price = marginal price + RSP/Q*. Taylor (1975) shows that the rate structure premium is an adjustment to income such that consumers choose quantity level q* at price π_r and income level Y - RSP.

A declining block rate schedule implies an expenditure function or outlay schedule which increases in linear segments, the slope of each succeeding segment being smaller than the one preceding it. More generally let B(Q) be any quantity dependent expenditure funtion. The marginal price at quantity Q is B'(Q) so that the corresponding rate structure premium adjustment is B(Q) - B'(Q)Q. If V(P,Y) is the indirect utility at prices p and income level Y then the consumer's optimal choice of quantity subject to the expenditure function B(Q) solves the problem:

$$\max_{Q} V[B'(Q), Y - [B(Q) - B'(Q)Q]].$$

The first-order condition implies that optimal Q is given as the solution to Roy's identity:

$$Q = -\frac{V_p[B'(Q), Y - (B(Q) - B'(Q)Q)]}{V_Y[B'(Q), Y - (B(Q) - B'(Q)Q)]}$$

2. Comparative Static Analysis of Demand Subject to a Declining Block Rate Structure

We now consider the comparative static analysis of demand subject to a declining block rate structure. Let U[q, Z] denote the utility derived from the consumption of electricity q and a Hicksian or numeraire commodity Z. We assume a two-tier tariff for electricity with the price of electricity π given by

$$(1) \qquad \pi = \begin{cases} \pi_1 & \text{for } 0 \leq q \leq X \\ \pi_2 & \text{for } X < q \qquad \text{with } \pi_1 > \pi_2. \end{cases}$$

Normalizing the price of the numeraire commodity to equal one the budget constraint satisfies:

(2) $\pi_1 q + Z \leq y$ for $q \leq X$

$\pi_1 X + (q - X)\pi_2 + Z \leq y$ for $X < q$

where y denotes income.

We illustrate the declining tariff in Diagram 1 and the corresponding budget set in Diagram 2.

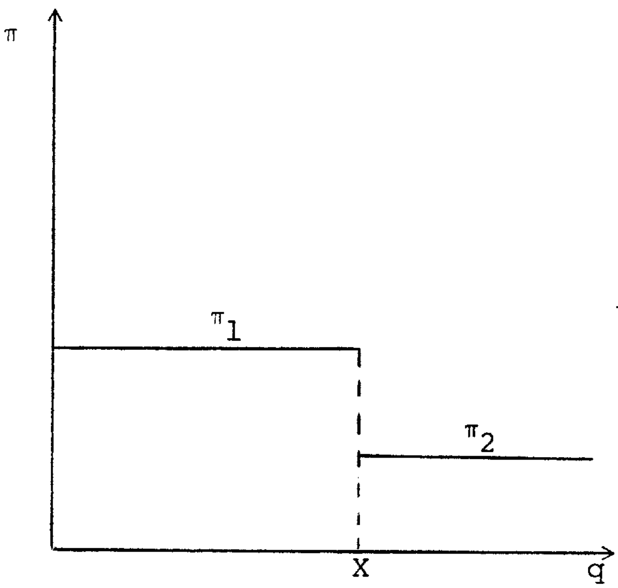


Diagram 1

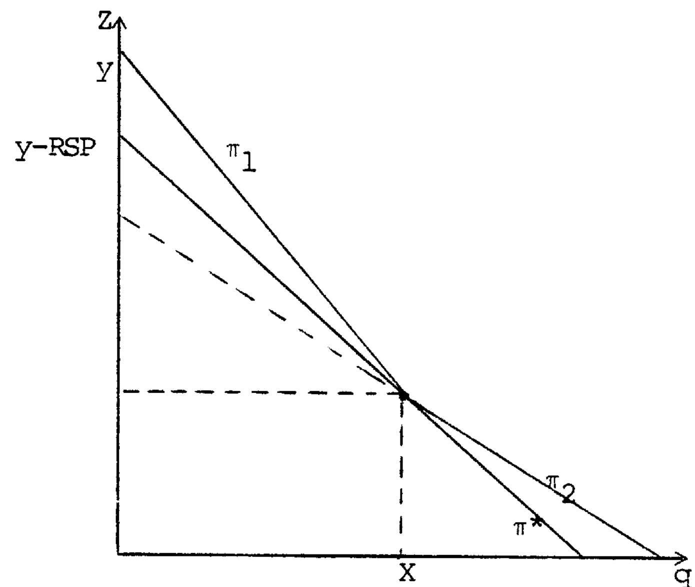


Diagram 2

Denote by $D[\pi, y; \beta]$ the Marshallian or uncompensated demand for electricity where β is a vector of behavioral parameters and let π^* denote the price at which demand equals the lower block boundary, i.e., $D[\pi^*, y; \beta] = X$. Let q_1 denote demand along the segment with slope π_1 and let q_2 denote demand along the segment with slope π_2.

Demand along the first budget segment satisfies

(4) $q_1 = D[\pi_1, y; \beta]$ for $(\pi_1, \pi_2, y) \varepsilon S_1$

while demand in the second segment satisfies

(5) $q_2 = D[\pi_2, y - (\pi_1 - \pi_2)X; \beta]$ for $(\pi_1, \pi_2, y) \varepsilon S_2$

The term $(\pi_1 - \pi_2)X$ is the rate structure premium adjustment for demand in the marginal or tail-end block. We now derive certain results concerning local price response.

Lemma 1

Suppose the uncompensated demand for electricity is decreasing in price and increasing in income then:

1a) $\partial q_1/\partial \pi_1 < 0$ for $(\pi_1, \pi_2, y) \in S_1$

1b) $\partial q_2/\partial \pi_1 < 0$ for $(\pi_1, \pi_2, y) \in S_2$

1c) $\partial q_2/\partial \pi_2 < 0$ for $q_2 \geq X$ and $(\pi_1, \pi_2, y) \in S_2$

Proof Lemma 1

1a) By assumption demand is downward sloping.

1b) $\partial q_2/\partial \pi_1 = (D_Y)(-X) < 0$ since we have assumed that electricity is a normal good.

1c) $\partial q_2/\partial \pi_2 = D_\pi + D_Y X \leq D_\pi + D_Y q_2$ since $X \leq q_2$. Finally $D_\pi + D_Y q_2 < 0$ since $D_\pi + D_Y q_2$ equals the partial derivative with respect to price of the Hicksian or compensated demand function (by Slutsky's relation) and is thus negative.

Remarks: For $\pi_1 \geq \pi^*$, $q_1 \leq X$ by Lemma 1a. For $\pi_1 < \pi^*$, $q_1 > X$ so that optimal demand falls outside the range in which π_1 is the prevailing price. Furthermore $\partial q_2/\partial \pi_2 < 0$ for $X \leq q_2$ implies that for $\pi_2 < \pi^*$, $q_2 > X$. The pattern of prices in which $\pi_2 < \pi^* \leq \pi_1$ implies that q_1 and q_2 are each feasible.

Let $V(\pi, y)$ be the indirect utility function corresponding to the problem $\max_{q,Z} U[q, Z]$ subject to $\pi q + Z \leq y$. For $\pi_2 < \pi^* \leq \pi_1$, the budget segment with price π_1 is optimal when $V(\pi_1, y) > V(\pi_2, y - (\pi_1 - \pi_2)X)$. It is clear that combinations of π_1 and π_2 exist which satisfy $\pi_2 < \pi^* \leq \pi_1$ and imply equal indirect utility so that demand for electricity is multi-valued. For the set of prices which imply equal indirect utility a trade-off exists where an increase in π_1 may be compensated by a decrease in π_2. We have the following result:

Lemma 2

Let $S = \left\{(\pi_2, \pi_1) \mid V(\pi_1, y) = V(\pi_2, y - (\pi_1 - \pi_2)X)\right\}$ for $\pi_2 < \pi^* \leq \pi_1$. Then $\partial\pi_1/\partial\pi_2 < 0$ for $(\pi_2, \pi_1)\varepsilon S$ and for $V_y(\pi_1, y) < V_y(\pi_2, y - (\pi_1 - \pi_2)X)$.

Proof Lemma 2

For $(\pi_2, \pi_1)\varepsilon S$,

$(\partial\pi_1/\partial\pi_2) \cdot V_{\pi_1} = V_{\pi_2} + V_{y_2}[(-X)(\partial\pi_1/\partial\pi_2) - 1]$. Then

$(\partial\pi_1/\partial\pi_2)(V_{\pi_1} + V_{y_2}X) = (V_{\pi_2} + V_{y_2}X)$ which implies

$$(\partial\pi_1/\partial\pi_2) = (V_{\pi_2} + V_{y_2}X)/(V_{\pi_1} + V_{y_2}X)$$

$$= (X - q_2)/(X - q_1(V_{y_1}/V_{y_2})) < 0$$

for $q_1 < X$ and $q_2 > X$. Q.E.D.

To complete the static analysis we need the following result which indicates the direction of change in indirect utility from changes in price.

Lemma 3

Let $V_1 = V[\pi_1, y]$ and $V_2 = V[\pi_2, y - (\pi_1 - \pi_2)X]$

3a) $\partial V_1/\partial\pi_1 < 0$

3b) $\partial V_2/\partial\pi_1 < 0$

3c) $\partial V_2/\partial\pi_2 < 0$ for $X \leq q_2$.

3d) $\partial(V_2 - V_1)/\partial\pi_2 < 0$ for $\pi_2 < \pi^* \leq \pi_1$ and $V_{y_1} < V_{y_2}$.

Proof Lemma 3

3a) $\partial V_1/\partial\pi_1 = V_\pi(\pi_1, y) < 0$ (monotonicity property of indirect utility function).

3b) $\partial V_2/\partial\pi_1 = V_{y_2}(-X) < 0$

3c) $\partial V_2/\partial\pi_2 = V_\pi + V_{y_2} X \leqq V_\pi + V_{y_2} q_2 < 0$ for $X \leq q_2$.

3d) $\partial(V_2 - V_1)/\partial\pi_1 \quad = -[V_{\pi_1} + V_{y_2} . X]$

$$= - V_{y_2} . [X - q \;\; . (V_{y_1}/V_{y_2})] < 0$$

as $V_{y_1}/V_{y_2} < 1$ and $q_1 < X$. Q.E.D.

We now collect the results in the following theorem.

Theorem 1 (Two-Tier Declining Block Rate Comparative Statics)

Let π^* be defined by $D[\pi^*, y; \beta] = X$. Define the functions $\pi_1^*(\pi_2)$ and $\pi_2^*(\pi_1)$ by

$V(\pi_1^*, y) = V(\pi_2, y - (\pi_1^* - \pi_2)X)$ and

$V(\pi_1, y) = V(\pi_2^*, y - (\pi_1 - \pi_2^*)X)$ respectively.

Then equilibrium occurs in the first segment for:

$$S_1 = \left\{(\pi_2, \pi_1) \mid \pi^* \leqq \pi_1 \text{ and } \pi_2^* (\pi_1) \leqq \pi_2 \leqq \pi_1\right\}$$ and

equilibrium occurs in the second segment for:

$$S_2 = \left\{(\pi_2, \pi_1) \mid 0 \leqq \pi_2 \leqq \pi^* \text{ and } \pi_2 \leqq \pi_1 \leqq \pi_1^* (\pi_2)\right\}.$$

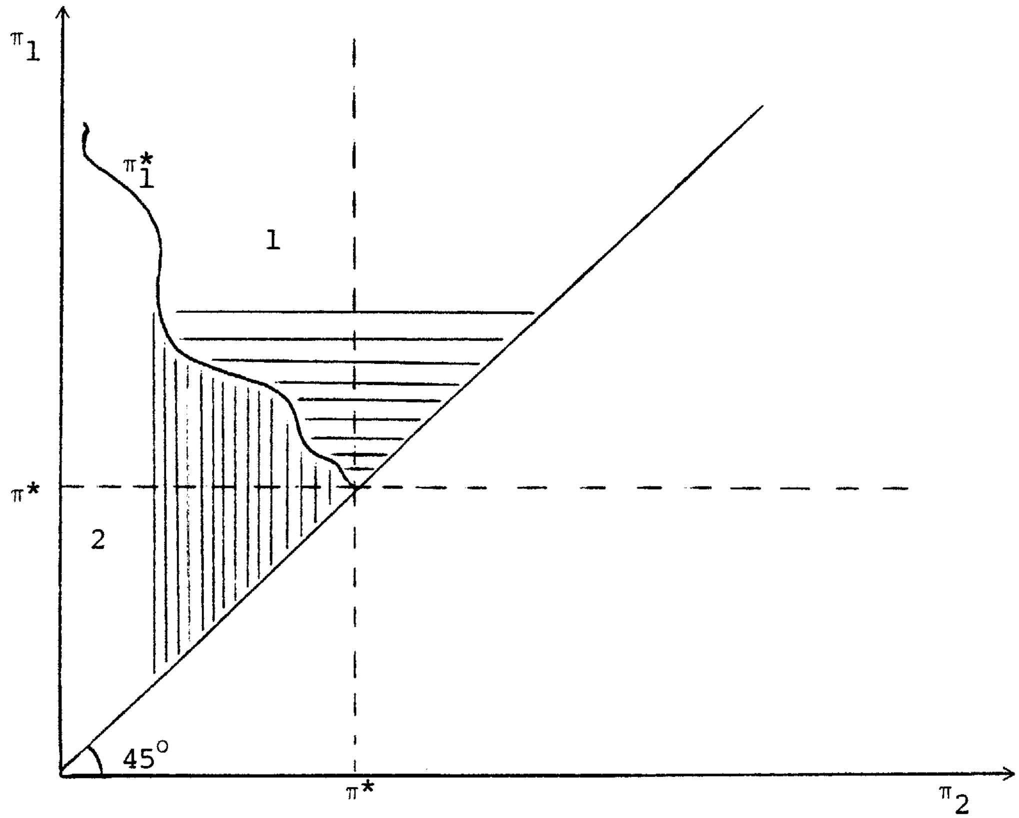


Diagram 3

Proof Theorem 1

The shaded region above the diagonal line in Diagram 3 represents the set of feasible declining block rate structures. The curve with declining slope which intersects the (π^*, π^*) point is the set S of Lemma 2. Suppose we begin at a point on the curve S and increase π_2 while leaving π_1 unchanged. Since we are in a region in which both budget segments are feasible, Lemma 3c implies that the increase in π_2 decreases the utility V_2. As we began at a point of equal utility and

V_2 has decreased while V_1 remains constant it must be the case that budget segment one is preferred to budget segment two as indicated in the Diagram.

Similarly consider a decrease in π_1 leaving π_2 constant. In this case, Lemma 3d applies so that $V_2 - V_1 > 0$ and budget segment two becomes optimal. In the southwest quadrant above the 45° degree line, demand occurs in the second budget segment since optimal demand for prices $\pi_1 < \pi^*$ exceeds the block boundary X. The other quadrants are similarly derived using the results of Lemma 1 and Lemma 3. Q.E.D.

Note that the price pairs below the diagonal imply increasing or non-decreasing block rate schedules which correspond to convex budget sets. The triangular area in the southwest quadrant below the diagonal implies optimal demand in the second budget segment while the area below the diagonal in the northeast quadrant implies demand in the first budget segment. The southeast quadrant which includes the boundary $\pi_1 = \pi^*$ but excludes the boundary $\pi_2 = \pi^*$ implies optimal demand at the block boundary X. We further note that the set S of equal utility points has measure zero in the price space of Diagram 3.

We now use Diagram 3 to answer simple comparative static problems. Suppose for example that we increase the lower block boundary. Diagram 4 illustrates that the partition moves to an intersection with the 45° line at the point $(\pi^{*\prime}, \pi^{*\prime})$ with $\pi^{*\prime} < \pi^*$ since $X' > X$.

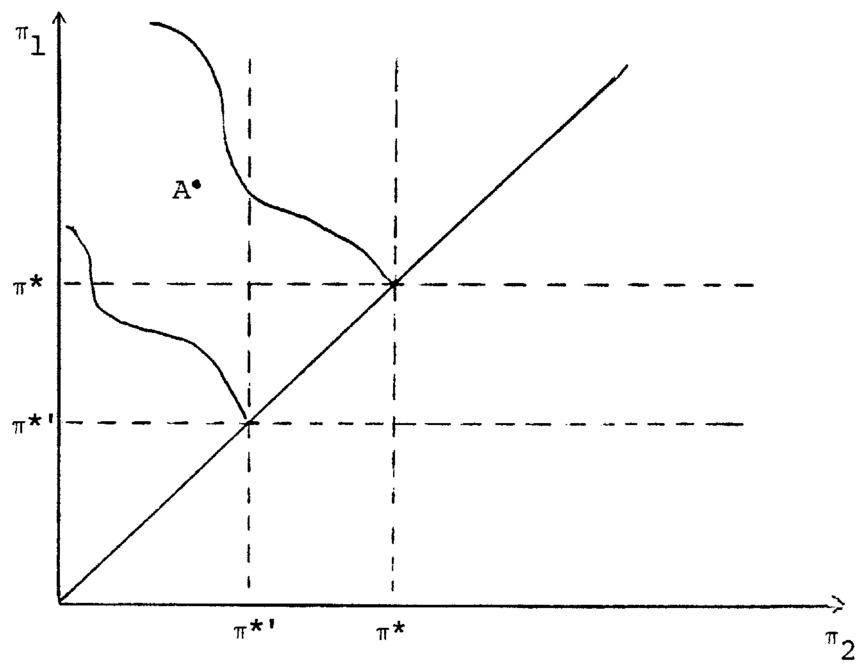


Diagram 4

Suppose equilibrium had occurred initially at the point A. The discontinuous change in lower block boundary from X to X' implies that the price pair at point A now corresponds to optimal demand in budget segment one versus the initial equilibrium in budget segment two.

Finally we note that our comparative static analysis as developed in Theorem 1 applies to the more general case of multiple tier declining block rate schedules where we interpret π_2 as the marginal rate and let π_1 be the intramarginal average price, i.e., the average price up to but not including the marginal block.

III. Specification of Price: Empirical Results

We now address the issue of price specification with an econometric analysis of the 1975 survey of 1502 households carried out by the Washington Center for Metropolitan Studies (WCMS) for the Federal Energy Administration. Individual household locations (identified at the level of primary sampling units) permitted matching of actual rate schedules used in 1975 to each household. The use of disaggregated data is necessary to avoid the confounding effects of misspecification due to aggregation bias or due to approximation of the rate data.

We resolve four empirical issues related to the estimation of the demand for electricity: (1) measured average price and measured marginal price are statistically endogenous so that least squares techniques are not appropriate for the determination of price elasticities, (2) while the rate structure premium adjustment has established theoretical merit its statistical contribution is negligible, (3) consumer behavior in the demand for electricity follows the marginal price specification rather than the average price specification, and (4) estimates of price responsiveness are not statistically different using the tail-end price rather than the true marginal rate.

1. Endogeneity of Measured Prices

The general proposition is that explanatory variables which utilize the observed consumption level introduce correlation between those variables and the error term. To illustrate the direction of least squares estimation bias write the demand for electricity equation as $Q = \beta p + Z\delta + \varepsilon$ where p is the measured marginal price with coefficient

β, Z is a vector of socioeconomic variables with coefficient vector δ and ε is the equation error. For simplicity assume that p is uncorrelated with Z so that $\hat{\beta}_{LS} = \beta + p'\varepsilon/p'p$. An unobserved increase in electricity consumption induces a decrease in price so that we expect an a priori negative correlation between p and ε. The formula for $\hat{\beta}_{LS}$ shows that least squares over estimates in absolute magnitude the price-response coefficient β.[2]

McFadden (1977) and Hausman et al. (1979) have demonstrated that an instrumental variable estimation technique provides consistent estimates of the electricity demand equation where instruments are constructed utilizing predicted rather than actual consumption to determine measured prices. In forming predicted consumption levels all endogenous variables are purged from the set of explanatory variables. One must insure that the instruments so constructed are not exact linear combinations of the exogenous variables included in the demand for electricity equation. This is usually not a problem given the non-linearity of the rate schedule and given the existence of other prices which are exogenous. The tail-end block price, for example, will be used in exactly this role.

To establish empirical verification of the hypothesis of endogeneity of measured price we apply the specification test due to Wu (1973) and recently discussed in Hausman (1978).

The methodology consists of isolating a group of explanatory variables whose endogeneity is under test. Using the result that the least squares estimator has zero asymptotic covariance with its difference from the instrumental variable estimator, we are able to form a simple statistic which is asymptotically chi-squared under the null hypothesis of statistical exogeneity for the test group.

To illustrate the test write the demand for electricity in schematic form as $Q = X\beta + Z\gamma + \varepsilon$ where X is a k-vector of price and income terms under various specifications and Z is a group of assumed exogenous variables. The variables in X will in general be suspect of endogeneity. The test statistic is then:

$$T = (\hat{\beta}_{IV} - \hat{\beta}_{LS})'[V[\hat{\beta}_{IV}] - V[\hat{\beta}_{LS}]]^{-1}(\hat{\beta}_{IV} - \hat{\beta}_{LS}) \overset{A}{\sim} \chi^2(k)$$

where V is the estimated variance covariance matrix and k is the number of coefficients in β.

The dependent variable in each estimated equation is monthly consumption of kilowatt hours of electricity used by the family in 1975. The socioeconomic variables include appliance ownership dummies for the electric dishwasher, electric washing machine, food freezer, electric range, color television, black and white television, electric clothes dryer, and central air conditioner. To capture the effects of climate, the annual number of cooling degree days (the number of days in which the daily average temperature was greater than 65°) and this number multiplied by respectively the central air conditioner dummy and the number of room air conditioners were included as well as scale variables for the number of rooms, the number of persons, and the number of room air conditioners.[3]

Price terms included the average price, measured marginal price and the tail-end block rate. These rates are used below in various combinations and are taken from the rate schedules prevailing in the winter of 1975.

In Table 1 we present the mean values of all variables. To demonstrate the bias induced by least squares under the marginal price

Table 1:

VARIABLE NAME[a]	DESCRIPTION	MEAN
AKWH75	monthly consumption of electricity in 1975	916.5
RATE	measured marginal price in 1975	.02427
AVPRICE	measured average price in 1975	.03128
WMPE75	winter tail-end block price for electricity in 1975	.02138
INCOME	monthly income of household head	1322
RSP	measured rate structure premium	5.151
WHE	electric water heat dummy	0.2728
SHE	electric space heat dummy	0.1411
ROOMS	number of rooms in household	6.078
PERSONS	number of persons in household	3.550
CAC	central air-conditioning dummy	0.2890
CDDCAC	(annual cooling degree days) * (CAC)	463.7
RACNUM	number of room air conditionerss	.4382
CDDRACNUM	(annual cooling degree days) * (RACNUM)	642.3
AUTOWSH	automatic washing machine dummy	0.8898
AUTODSH	automatic dishwasher dummy	0.4921
FOODFRZ	food freezer dummy	0.5323
ELECRNGE	electric range dummy	0.6411
ECLTHDR	electric clothes dryer dummy	0.4990
BWTV	black and white television dummy	0.5806
CLRTV	color television dummy	0.7446

[a]A subsample of the original 1502 observations was selected so that all price and income data were positive and so that complete information was available for each individual.

specification we compare the least squares and instrumental variable estimates of the equation: $Q = \alpha$ (measured marginal price) $+ Z\delta + \varepsilon$. For brevity we report the coefficient estimates on the variables: measured marginal price, income, electric water heat and electric space heat in Table 2. At sample means the price elasticity implied by least squares is -0.266 while the instrumental variable estimates imply a price elasticity of -0.159. The direction of the bias agrees with our a priori expectation that least squares will overestimate in magnitude the price sensitivity coefficient.

Taylor reports both short-run and long-run price and income elasticities. Of nine estimates of residential elasticities two used marginal price. Each of the studies by Houthakker (1951a, 1951b) reports short-run elasticities of approximately -0.90.[4] Both our least squares and instrumental estimates are well below this estimate in magnitude but are entirely consistent with other estimates of electricity demand price elasticity using an average price specification.[5]

The Hausman statistic for the endogeneity test of measured marginal price is computed to be 34.18. This well exceeds the critical value for a Chi-squared test of any size given the single degree of freedom. We note that the respective income elasticities for least squares and instrumental variables are 0.118 and 0.109. Both estimates are consistent with those obtained in previous studies.

If the same test is repeated using measured average price in place of measured marginal price we find price elasticities for least squares and instrumental variables of respectively -0.437 and -0.416. Note that the direction of bias is the same as that obtained with measured marginal price--a general increase in price sensitivity magnitude. Income

Table 2:

VARIABLE[a]	LS ESTIMATES	IV ESTIMATES
Measured Marginal Price	-10050.	-6006.
	(-5.909)[b]	(-3.269)
Income	.08169	.07570
	(3.330)	(3.071)
WHE	405.6	404.5
	(10.22)	(10.15)
SHE	694.8	714.9
	(14.08)	(14.40)
R^2	.7074	.7051
Number of Observations	744	744
Sum of Squared Residuals	.9094E+8	.9166E+8
Standard Error of Regresion	354.2	355.6

[a]In Tables 2-6 coefficient estimates are not reported for the variables: PERSONS, BWTV, ROOMS, RMCLCAC, CDDCAC, CAC, RACNUM, CDDRACNUM, FOODFRZ, ELECTRNGE, CLRTV, ECLTHDR, AUTODSH, AUTOWSH, and the intercept. The dependent variable is AKWH75.

[b]t-statistics presented in parentheses.

elasticities were robustly estimated at 0.120 and 0.104 for the two procedures. The Chi-squared statistic was computed in this case to be 118.2 which well exceeds the critical value of 3.84 for a 5 percent test. Parameter estimates for the average price specification are reported in Table 3.

In summary we remark that previous studies in the demand for electricity have undoubtedly been subject to the bias illustrated above. The bias has been demonstrated to be statistically significant for the two most common specifications of price and is qualitatively impressive on the order of 67 percent.[6]

2. Rate Structure Premium Adjustment

From Table 1 we see that the mean value of rate structure premium is $3.12 compared to the mean value of income of $1321/month. The negligible value of RSP as compared to INCOME implies that the difference (INCOME - RSP) could not be distinguished from general measurement error in the definition of monthly income. In Table 4 we present instrumental variable estimates of the electricity demand equation using the marginal price specification and income adjusted by the rate structure premium.

Comparison of the estimates in Table 4 with estimates given in Table 2 for instrumental variables demonstrates the qualitative similarity. Based on these results we do not advocate the rate structure premium correction to income in the WCMS data for 1978. This confirms the findings of Hausman et al. (1979) for insignificance of the RSP adjustment.

Table 3:

VARIABLE	LS ESTIMATES	IV ESTIMATES
Average Price	-12810.	-4266.
	(-8.731)	(-2.563)
Income	.08304	.07221
	(3.484)	(2.959)
WHE	388.8	398.1
	(10.05)	(10.06)
SHE	669.2	719.6
	(13.90)	(14.56)
R^2	.7225	.7095
Number of Observations	744	744
Sum of Squared Residuals	.8626E+8	.9029E+8
Standard Error of Regresion	344.9	352.9

Table 4:

VARIABLE	IV ESTIMATES
Measured Marginal Price	-6006.
	(-3.269)
NETINC	.7560E-01
	(3.067)
WHE	404.5
	(10.15)
SHE	715.0
	(14.40)
R^2	.7050
Number of Observations	744
Sum of Squared Residuals	.9167E+8
Standard Error of Regresion	355.6

3. Average versus Marginal Price

Estimation in demand for electricity studies has followed the predominant usage of either marginal or average price. A simple observation will allow us to nest both the marginal and average price specification in a more general model. We have demonstrated above that the difference between measured average price and measured marginal price is the rate structure premium divided by measured consumption. Hence an unrestricted specification of marginal and average prices has the form:

$$Q = (\text{measured marginal price})\alpha_0 + (\text{Rate structure premium/Quantity})\alpha_1 + Z\delta + \varepsilon$$

Clearly when α_0 equals α_1 we have the average price specification. When $\alpha_1 = 0$ we have the marginal price specification.

Ordinary least squares and instrumental variable estimates for the unrestricted model are presented in Table 5. For brevity we report only the coefficient estimates of measured marginal price, rate structure premium/quantity, income, WHE, and SHE. The Hausman statistic of 83.8 with the two degrees of freedom confirms the endogeneity of the explanatory variables measured marginal price and rate structure premium/quantity.

Using the instrumental variables estimates in Table 5 we compute a Wald test of the hypothesis that the coefficients of measured marginal rate and rate structure premium/Q are equal. The test statistic which compares the difference in the estimated coefficients has a value of 7.09 and is distributed chi-squared with one degree of freedom (the number of imposed restrictions). We thus reject the average price specification at the 1 percent critical level. Furthermore the individual t-statistics for the coefficients of measured marginal price and RSP/Q confirm the

Table 5:

VARIABLE	LS ESTIMATES	IV ESTIMATES
Measured marginal rate	-10130.	-6430.
	(-6.158)	(-3.352)
Rate Structure Premium/Q	-22410	10040.
	(-7.236)	(1.777)
Income	.07702	.07846
	(3.248)	(3.068)
WHE	374.9	418.4
	(9.717)	(9.961)
SHE	673.6	722.1
	(14.10)	(14.00)
R^2	.7271	.6840
Number of Observations	744	744
Sum of Squared Residuals	.8481E+8	.9823E+8
Standard Error of Regresion	342.3	368.3

marginal price specification as the former coefficient is significant while the latter is insignificant at the 5 percent level.[7] It is interesting to note that inspection of the least squares estimates would lead one to choose the average price specification over the marginal price specification. Given the differential in sum of squared residuals for the measured marginal price and average price specifications (using the consistent estimates in Tables 2 and 3 respectively) it is likely that a non-nested test (see Pesaran (1974) for example) would also discriminate between the two models. We are thus led to conclude that consumer behavior in the demand for electricity follows the marginal price specification rather than the average price specification.

4. Measurement Error in Marginal Price

We now consider the impact of the measurement-error misspecification which results from the use of the tail-end rate in place of the measured marginal rate. In Table 6 we reproduce the least squares regression results for this specification. Note that least squares estimation provides consistent parameter estimates since the tail-end price is by definition exogenous. The use of the tail-end rate in place of the measured marginal rate introduces measurement error in the price variable. However it is not appropriate to apply the usual measurement error bias formulae since price is expected to reveal significant correlation with the other explanatory variables and since the difference between the two measures of price is not a mean zero random disturbance.

Comparing the estimate of the tail-end price coefficient in Table 6 with the consistent estimate of the measured marginal price coefficient in Table 2, we see that relative to the standard error the difference is

Table 6:

VARIABLE	LS ESTIMATES
WMPE75	-6828.
	(-3.644)
Income	.08299
	(3.277)
WHE	414.1
	(10.26)
SHE	721.7
	(14.51)
R^2	.6988
Number of Observations	744
Sum of Squared Residuals	.9361E+8
Standard Error of Regresion	359.3

not significant. (t = (-6006.) -(-6828.)/1837. = 0.45). This result is confirmed through the inspection of the variables WMPE75 and RATE; the correlation coefficient between the two variables is 0.87 and the mean difference is approximately one-third of a standard deviation. While there is no specific suggestion that the rate schedules in the WCMS data are flat, these estimates suggest that many individuals are close to the tail-end of the rate schedule so that measured marginal rates are well approximated by the tail-end price.[8]

IV. Measurement of Price: Theory and Estimation

This section investigates the construction of marginal price when basic observations are limited to total quantity consumed and total expenditure. We begin with an analysis of eight locations from the WCMS (1975) data set for which precise matching of rate schedules to households was possible. We compare the two-part tariff approximation to the actual rate schedule and attempt to illustrate the qualitative and quantitative bias in each physical location. We then examine seven locations from the National Interim Energy Consumption Survey (1978) NIECS for which only total expenditure and quantities are observed by billing periods. Under the assumption that households within a primary sampling unit are served by a common utility we attempt to distinguish between all electric and seasonal rates.

1. Washington Center for Metropolitan Studies (1975)

In Figures 1-8 we plot expenditure versus quantities for eight WCMS households. The figures are organized in pairs: (1) the plot of expenditures versus quantities and (2) the plot of the prevailing rate schedule. The symbol R denotes points chosen from the rate schedule while symbols A and B denote one and two observations respectively. For each location we give the estimates of the two-part tariff approximations, the actual tail-end price and the appropriate connect charge.

Figure 1a illustrates that 9 of the 10 observations from Boston, MA. correspond to the tail-end price. The estimate of marginal price from the two-part tariff approximation is 0.0373 while the actual tail-end

Boston, Massachusetts 1975

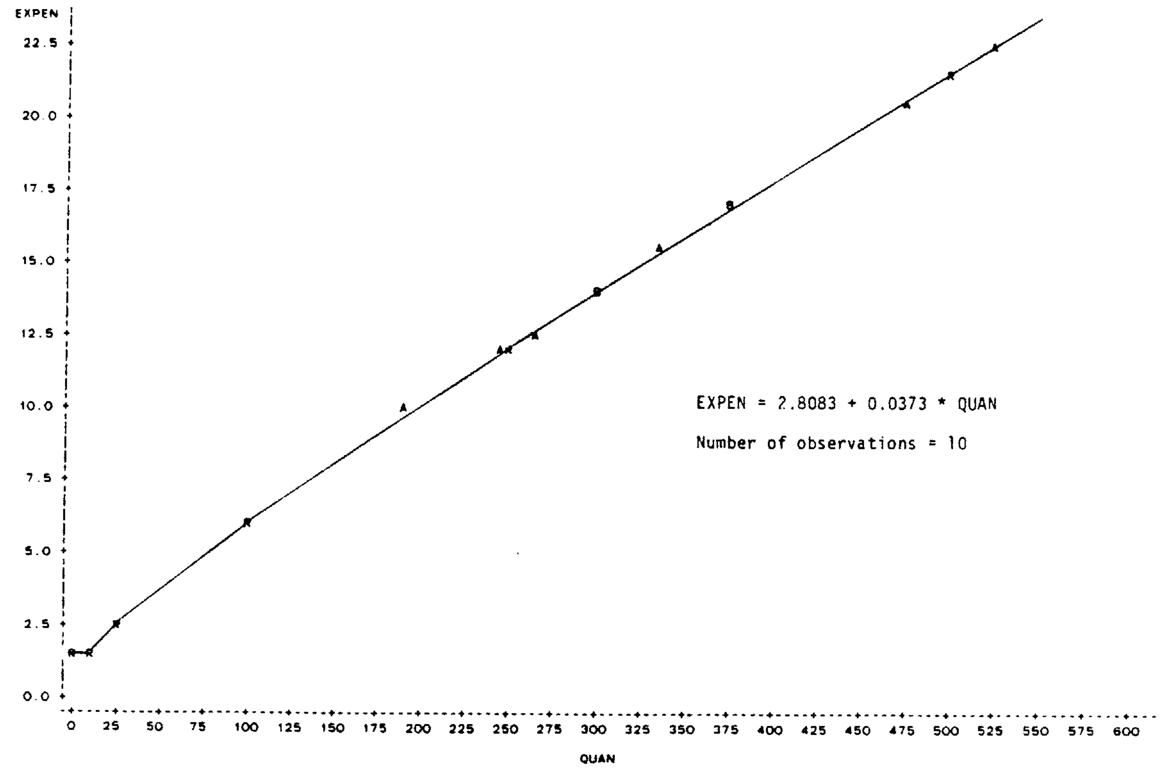


Figure 1a

Boston, Massachusetts 1975

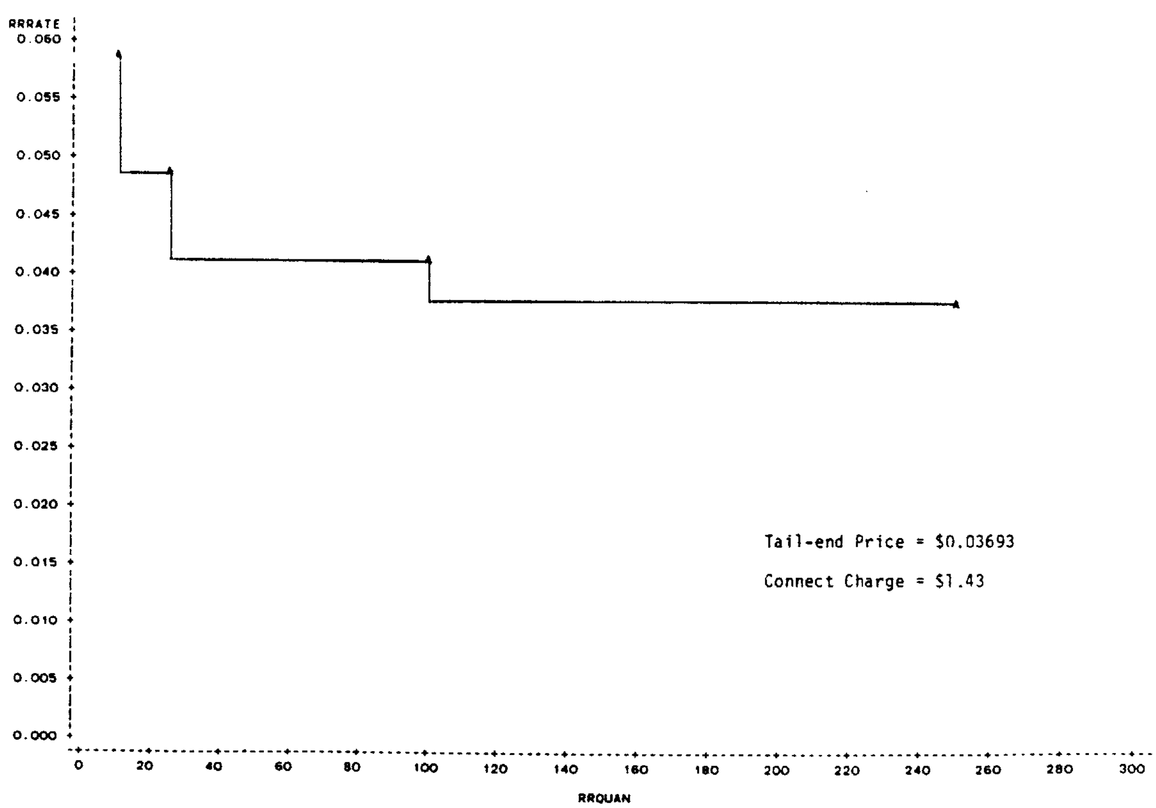


Figure 1b

Chicago, Illinois 1975

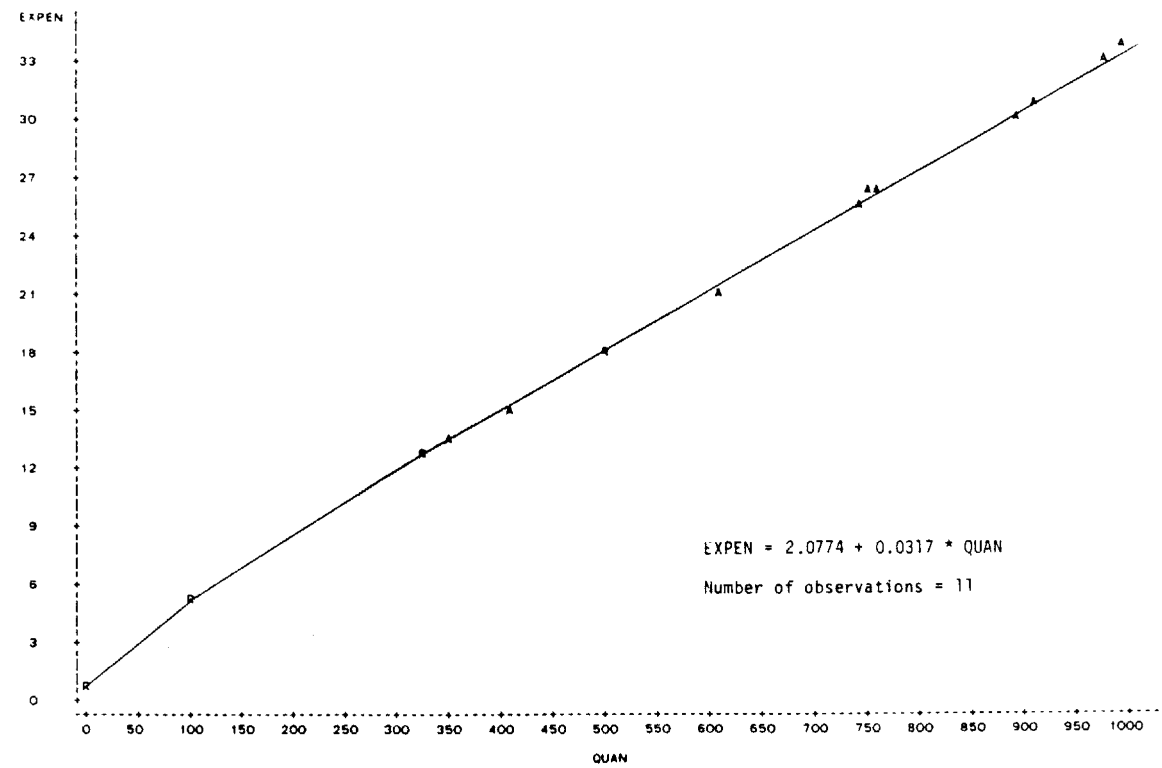


Figure 2a

Chicago, Illinois 1975

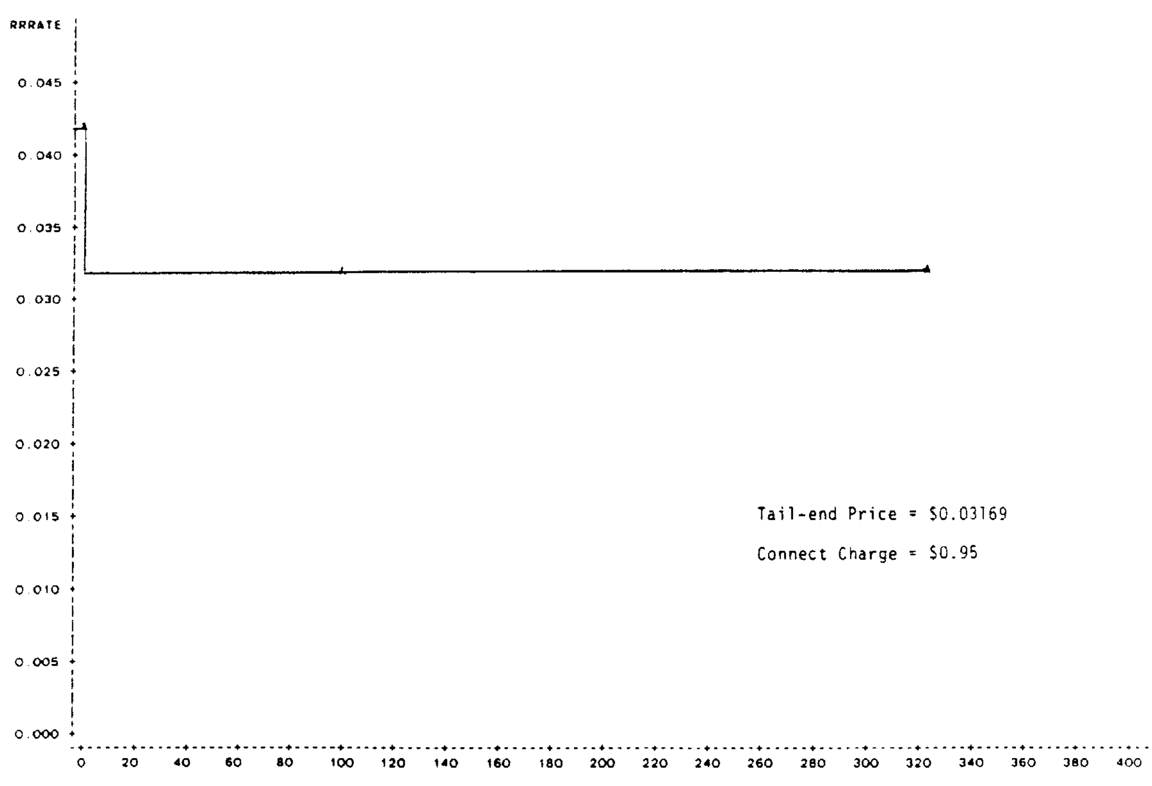


Figure 2b

Springfield, Ohio 1975

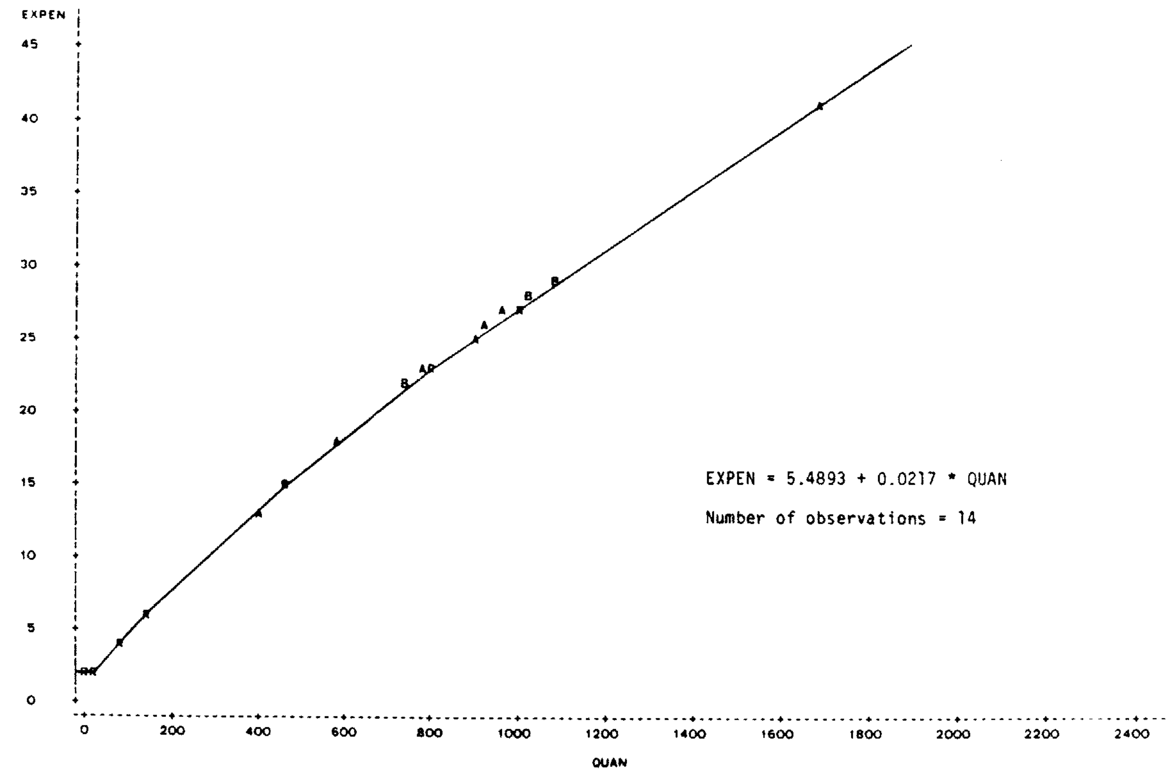


Figure 3a

Springfield, Ohio 1975

Figure 3b

Detroit, Michigan 1975

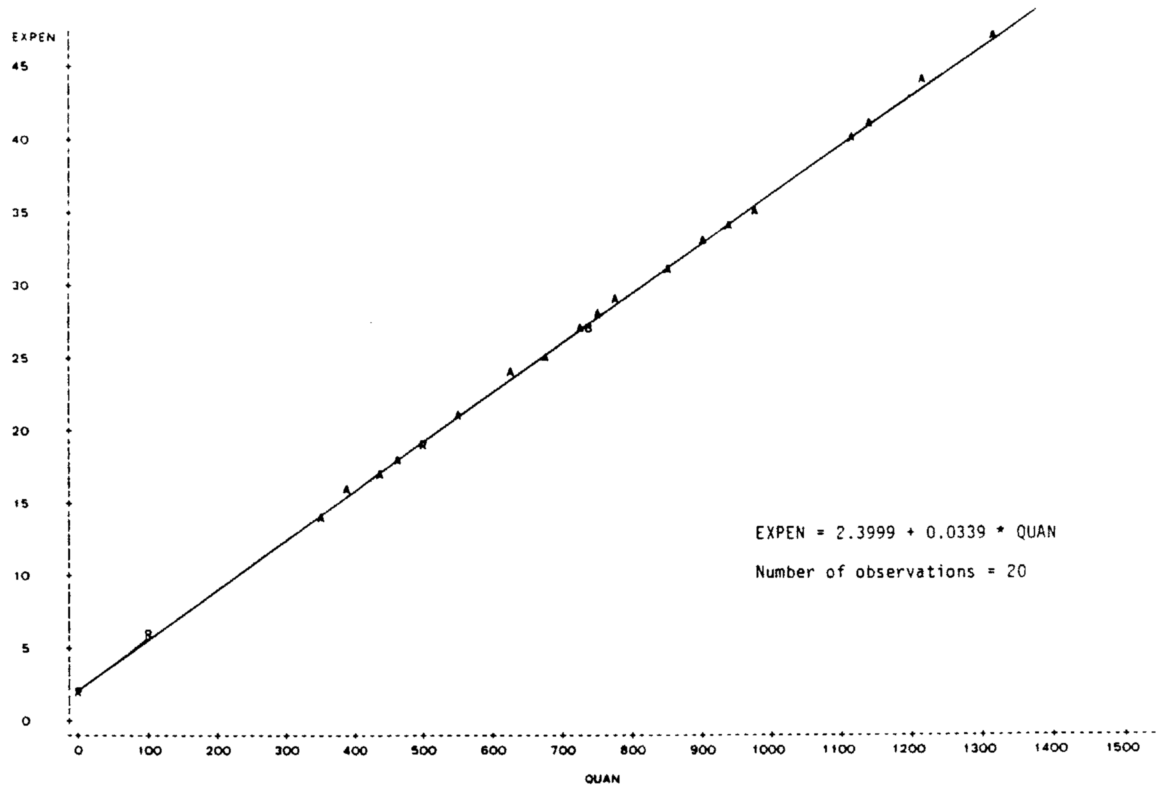


Figure 4a

Detroit, Michigan 1975

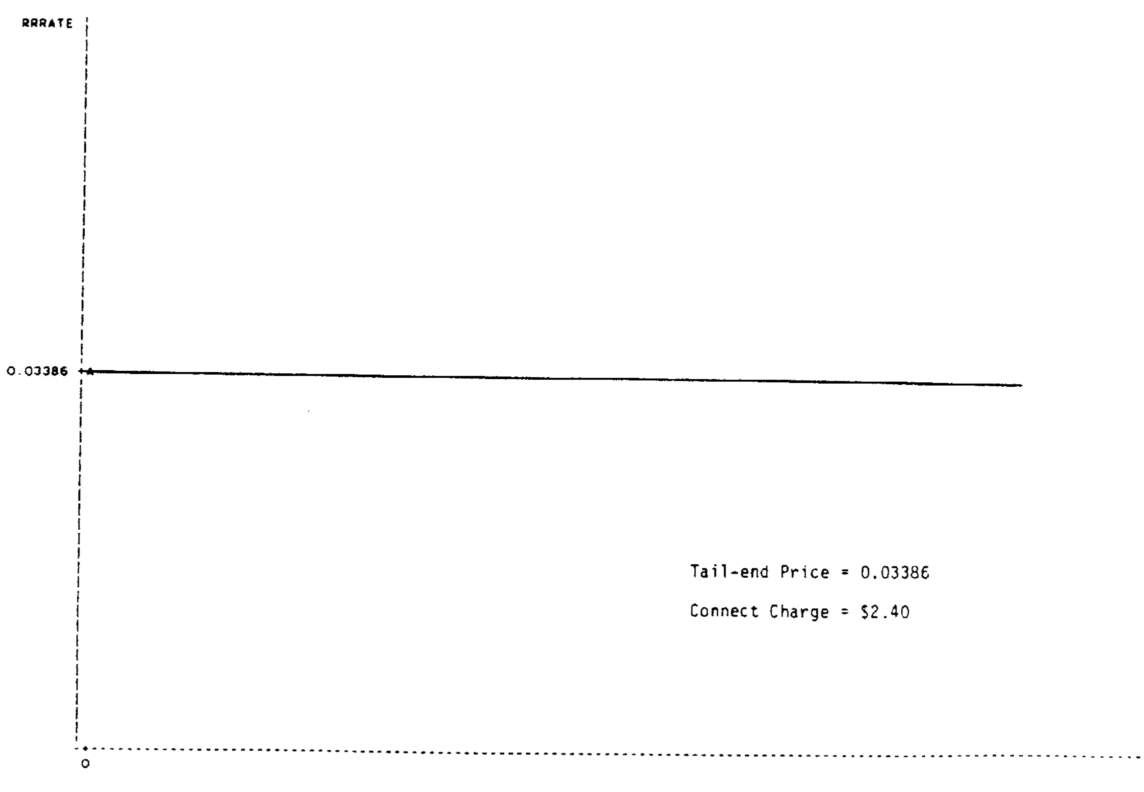


Figure 4b

Pittsfield, Massachusetts 1975

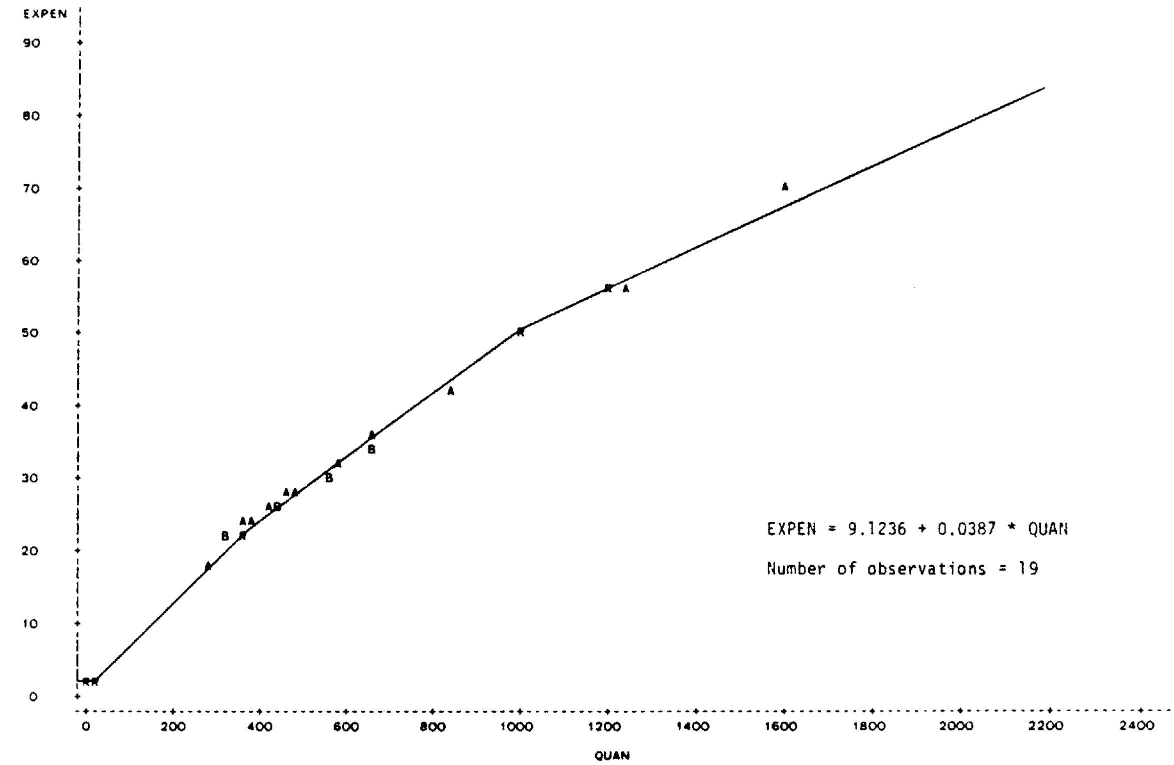


Figure 5a

Pittsfield, Massachusetts 1975

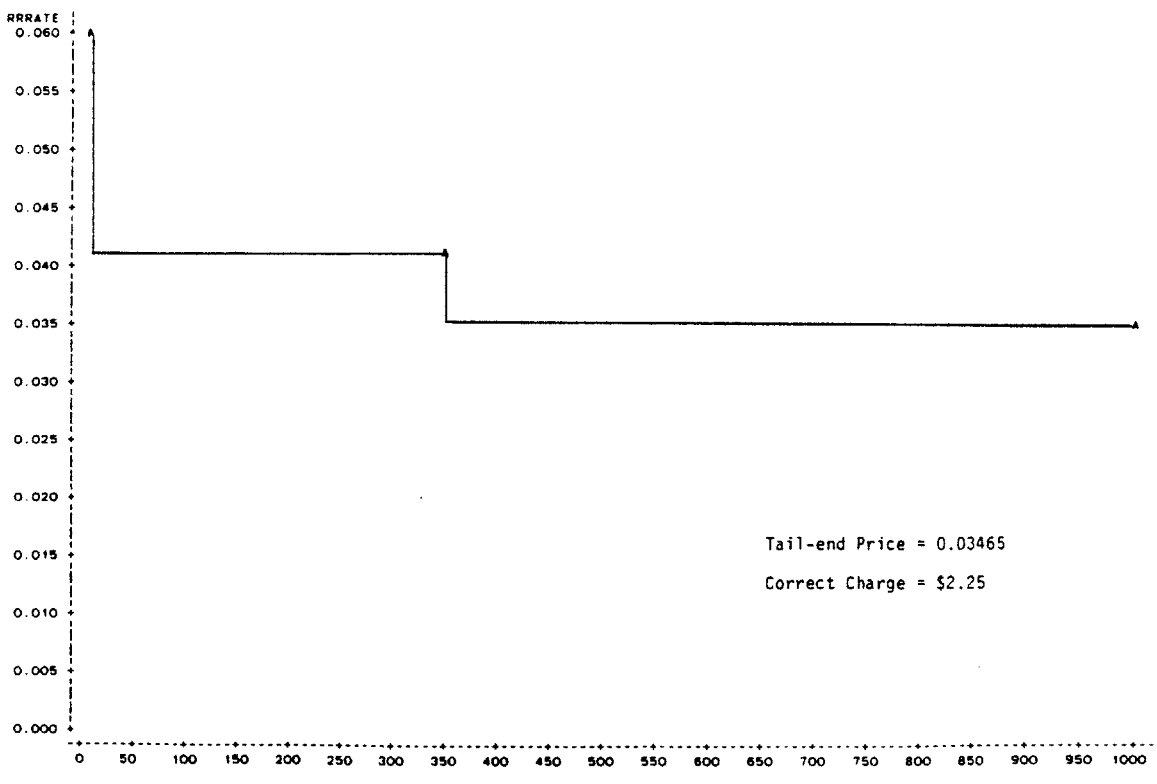


Figure 5b

Dayton, Ohio 1975

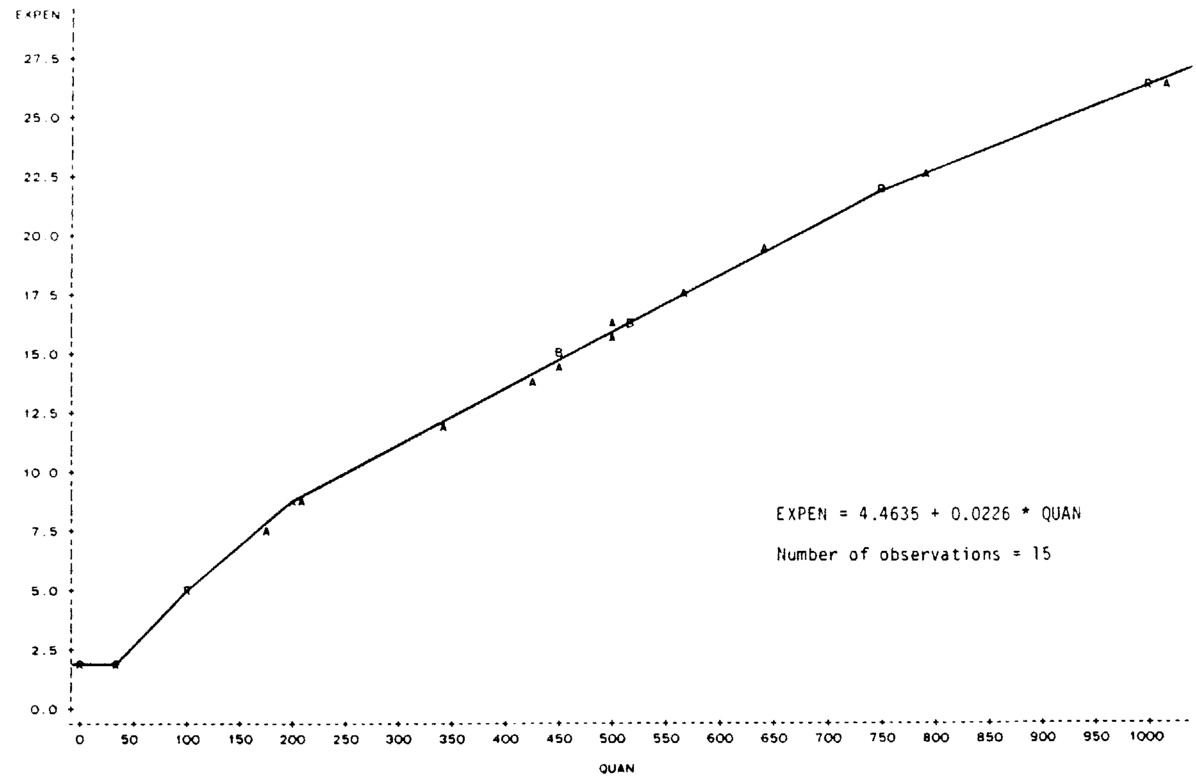


Figure 6a

Dayton, Ohio 1975

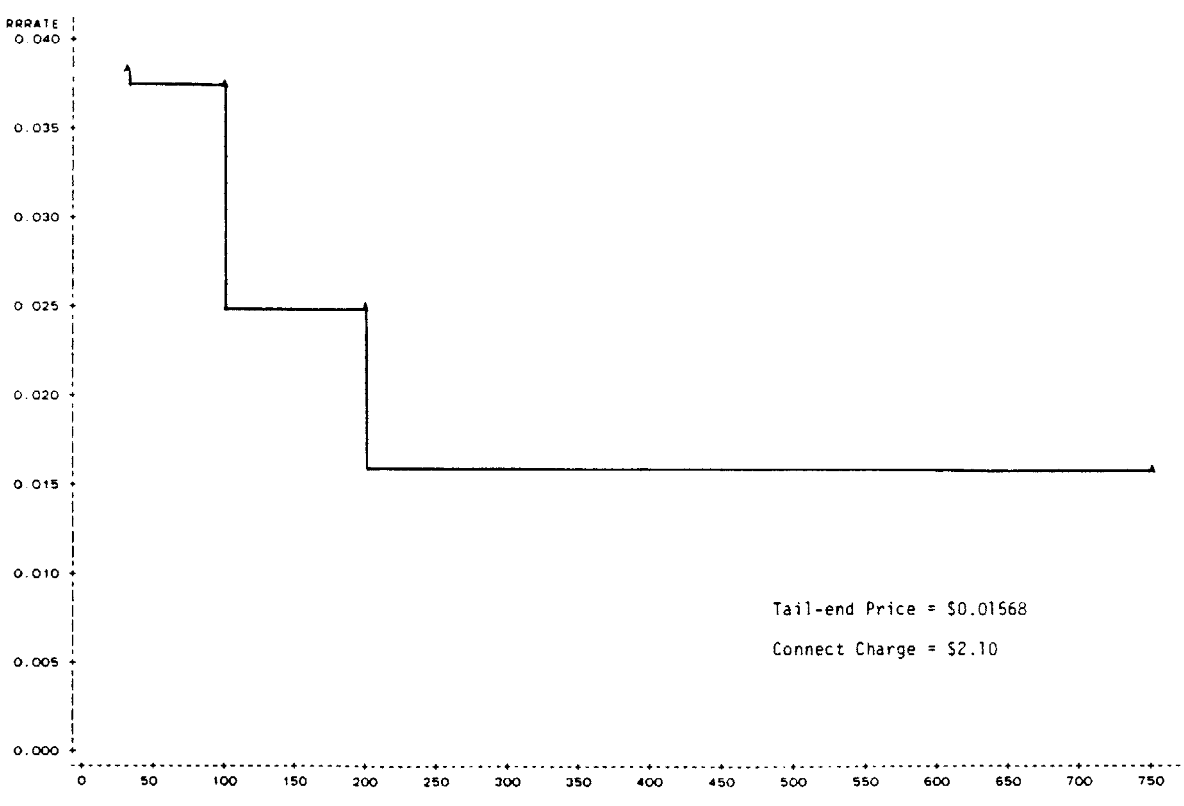


Figure 6b

Buffalo, New York 1975

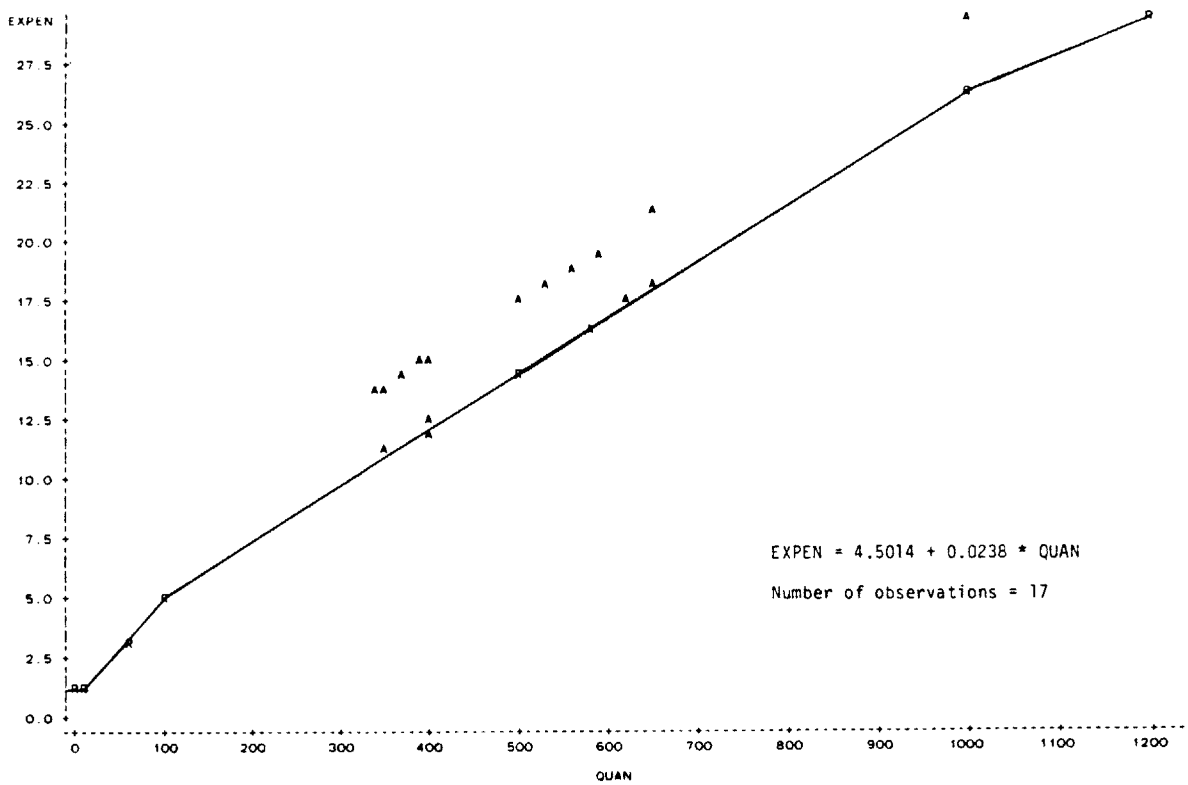


Figure 7a

Buffalo, New York 1975

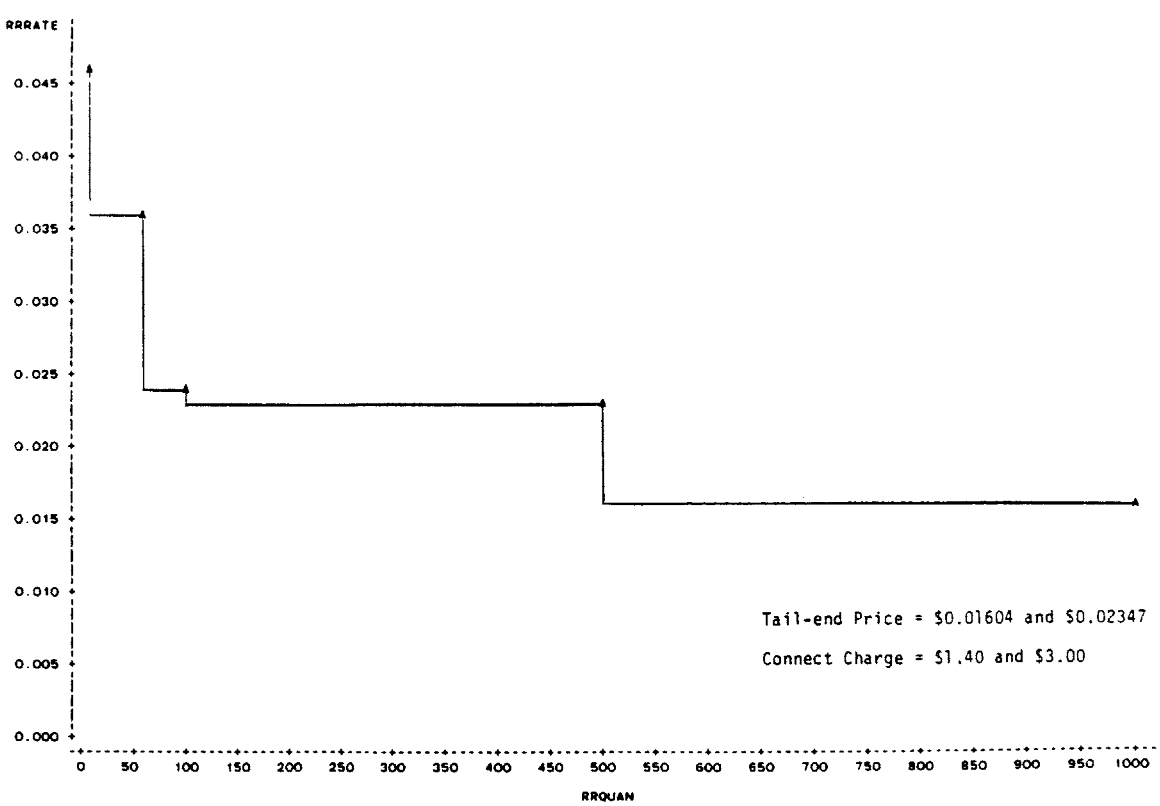


Figure 7b

Cortland, New York 1975

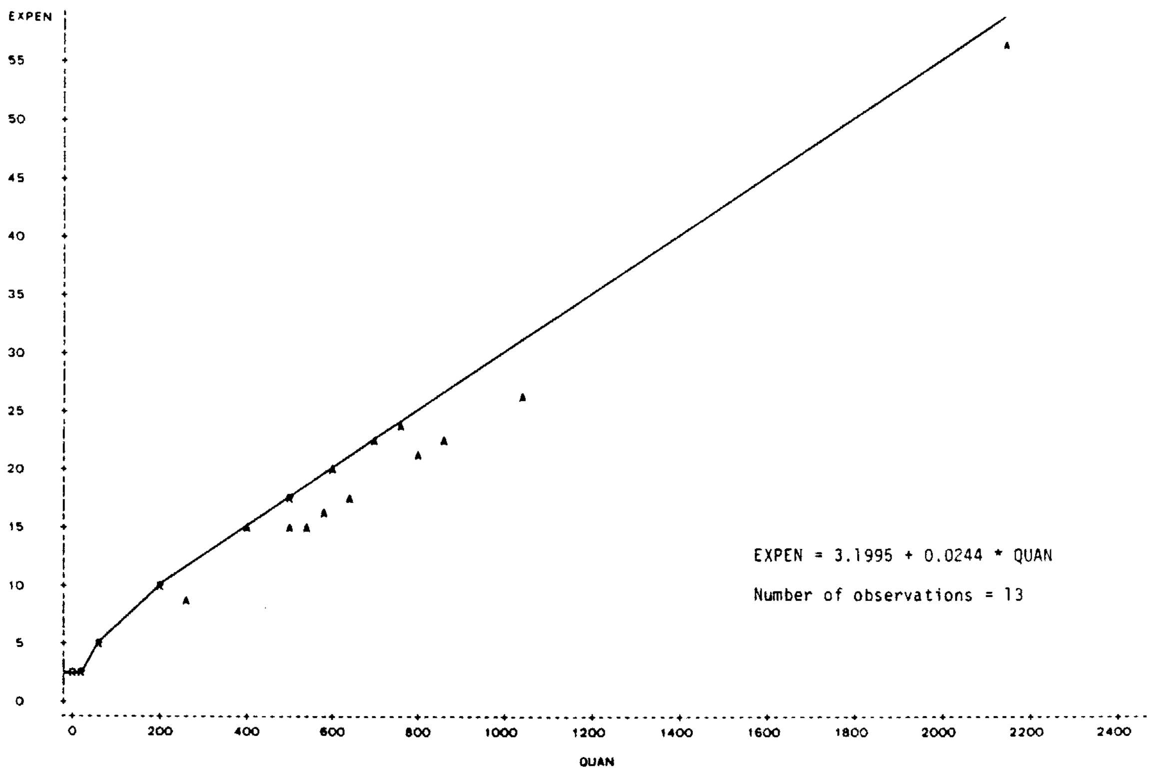


Figure 8a

Cortland, New York 1975

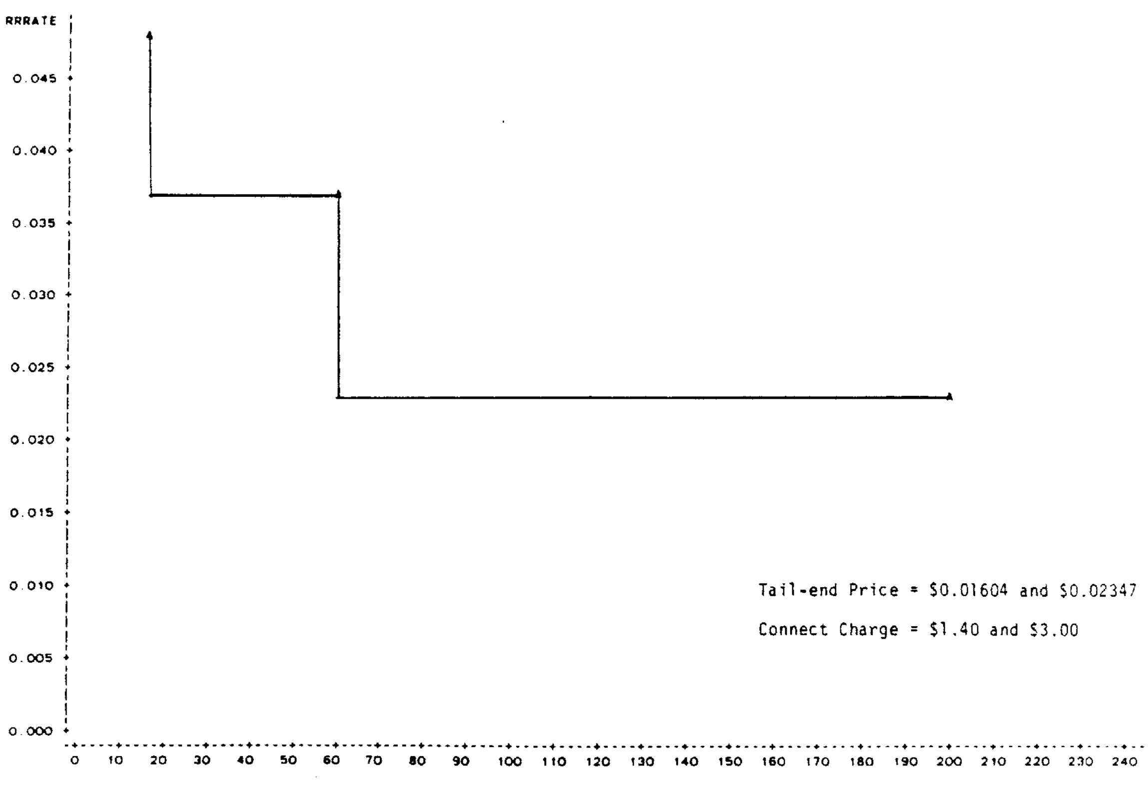


Figure 8b

rate is 0.03693. The standard error of the slope estimate is 0.000215 so that a t-test for significance of the difference is rejected at the 5 percent level. (One-sided test, degrees of freedom = 8, size = 5 percent, t = 1.172.) The situation in Figures 2a and 2b is qualitatively similar. In this case 9 of the 11 observations lie in the tail-end of the rate structure. The estimate of the slope is 0.0317 which is again not statistically different from the tail-end price 0.03169 (one-sided test, degrees of freedom = 9, size = 5 percent).

In Figure 3 fewer observations are in the tail. The estimate of the slope coefficient is 0.0217 while the tail-end price is 0.02043. The t-statistic for the difference is 3.86 which is significant for a one-sided 5 percent test given the 12 degrees of freedom. Figure 4 illustrates a near-perfect fit as the underlying rate schedule is flat. By contrast the distribution of points in Figure 5a suggests that the two-part tariff should not approximate the declining block rate schedule very accurately. In this case the estimated slope coefficient is 0.0387 while the true tail-end rate is 0.03465. With 17 degrees of freedom we reject the hypothesis that the estimate of the tail-end rate and the actual rate are equal (t = 9.01, size = 5 percent, degrees of freedom 17). Figure 6 is qualitatively similar to Figure 5 (t = 11.58 with 13 degrees of freedom).

Figures 7 and 8 illustrate a quite different phenomenon. Clearly two separate rate schedules were operative for Buffalo, New York and Cortland, New York in 1975. Their respective rates are given in Figures 7b and 8b respectively. Inpsection reveals that the two-part tariff approximation to multiple rate schedules is not likely to provide an adequate estimate of any individual marginal rate (t = 3.52 and t = 0.17

for Figure 7, and t = 8.64 and t = 0.92 for Figure 8).

In summary we have seen that the two-part tariff approximation to the declining rate schedule works quite well when many observations lie in the tail-end block. However when more than the one rate schedule prevails within a given primary sampling unit it is possible to estimate incorrectly the tail price. As the eight WCMS locations are not necessarily representative of the complete sample it is not possible to make a statement about general misspecification from only their analysis. The following calculation attempts to bound the estimation error inherent in the use of a two-part tariff approximation for the WCMS data. Essentially misspecification arises because the rate structure premium varies with quantity. If the rate structure premium were constant then the rate structure would be exactly in two-part tariff form. We thus apply a simple misspecification argument to estimate the bias.

Recall that by definition: Expenditure = Rate Structure Premium + Marginal Price * Quantity. For household i we write:

$EXPEN_i = RSP_i + \beta Q_i + \varepsilon_i$ where:

$EXPEN_i$ = expenditure by household i,

RSP_i = rate structure premium for household i,

β = marginal rate

Q_i = quantity consumed by household i,

ε_i = disturbance term.

Rewrite the true model as:

$EXPEN_i = \alpha + \beta Q_i + v_i$ where $v_i = \varepsilon_i + RSP_i - \alpha$

Least-square estimation implies:

$$(\hat{\beta} - \beta) = \sum_i (q_i - \bar{q})(v_i - \bar{v}) \Big/ \sum_i (q_i - \bar{q})^2$$

Since $v_i - \bar{v} = (\varepsilon_i - \bar{\varepsilon}) + (RSP_i - \overline{RSP})$ we have:

$$(\hat{\beta} - \beta) = \sum (q_i - \bar{q})(\varepsilon_i - \bar{\varepsilon}) \Big/ \sum (q_i - \bar{q})^2$$

$$+ \sum (q_i - \bar{q})(RSP_i - \overline{RSP}) \Big/ \sum (q_i - \bar{q})^2$$

so that PLIM $(\hat{\beta} - \beta) = (\frac{\sigma_{RSP}}{\sigma_Q})$ Correl (q, RSP). In the WCMS data the correlation of rate structure premium and quantity is 0.4659 while the standard deviation of rate structure premium and quantity are 2.906 and 646.3 respectively. Hence the two-part tariff approximation bias underpredicts the true marginal rate by 0.002095. Using these estimates the two-part approximation would imply marginal price of +0.02348 relative to the mean value tail-end price of 0.02138. This difference is about 25 percent of one standard deviation in the tail-end price. In conclusion it appears that the two-part tariff approximation adequately represents the declining block rate schedule in the determination of the tail-end block rate for the WCMS data of 1975.

2. National Interim Energy Consumption Survey (1978)

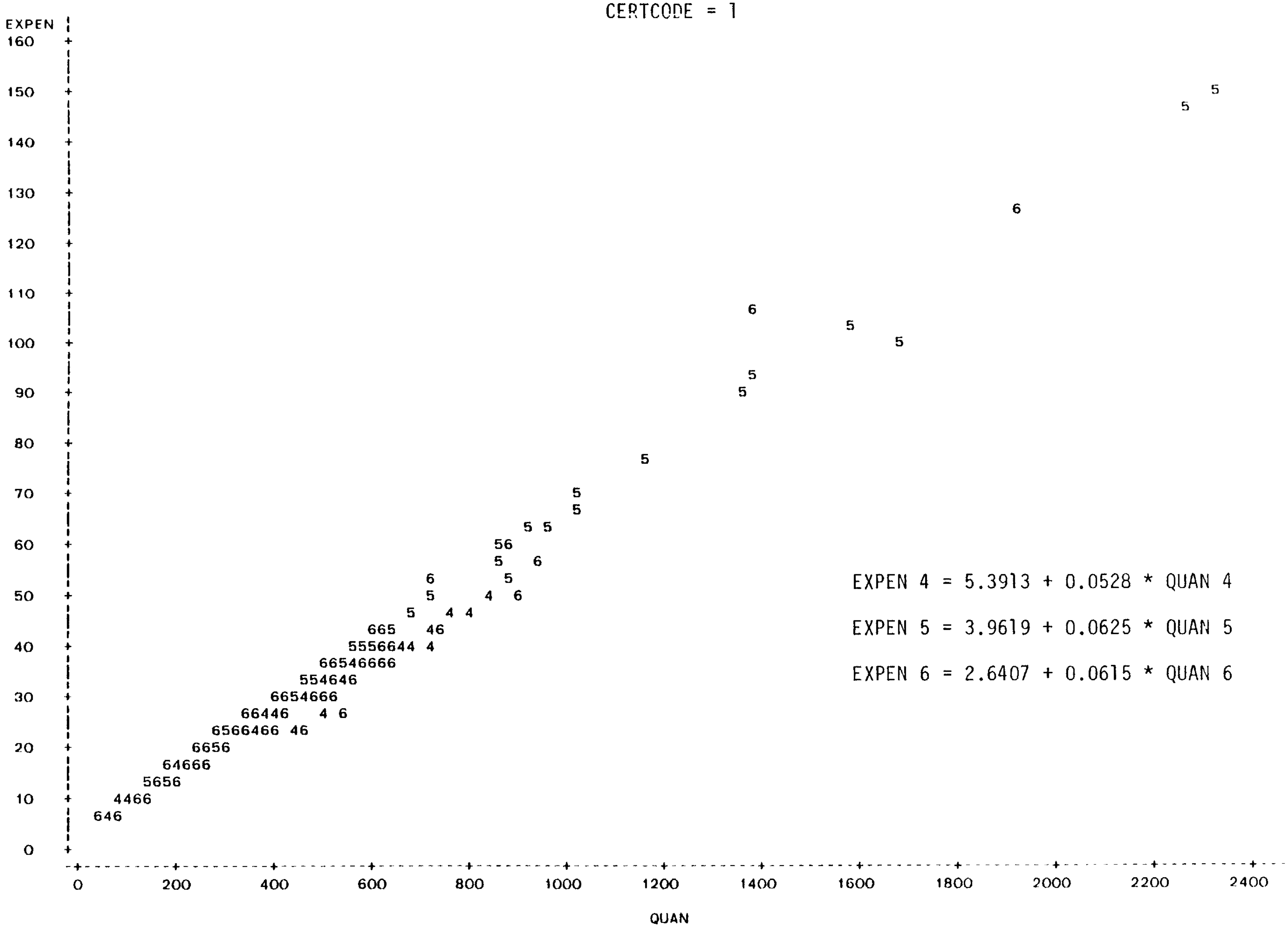
In Figures 9-15 we plot expenditure versus quantity for selected NIECS locations. We have allowed for the following possible rate schedules:

1 - all electric home in the winter

2 - all electric home in the summer

Newark, New Jersey 1978

CERTCODE = 1



EXPEN 4 = 5.3913 + 0.0528 * QUAN 4

EXPEN 5 = 3.9619 + 0.0625 * QUAN 5

EXPEN 6 = 2.6407 + 0.0615 * QUAN 6

Figure 9

Buffalo, New York 1978

CERTCODE = 1

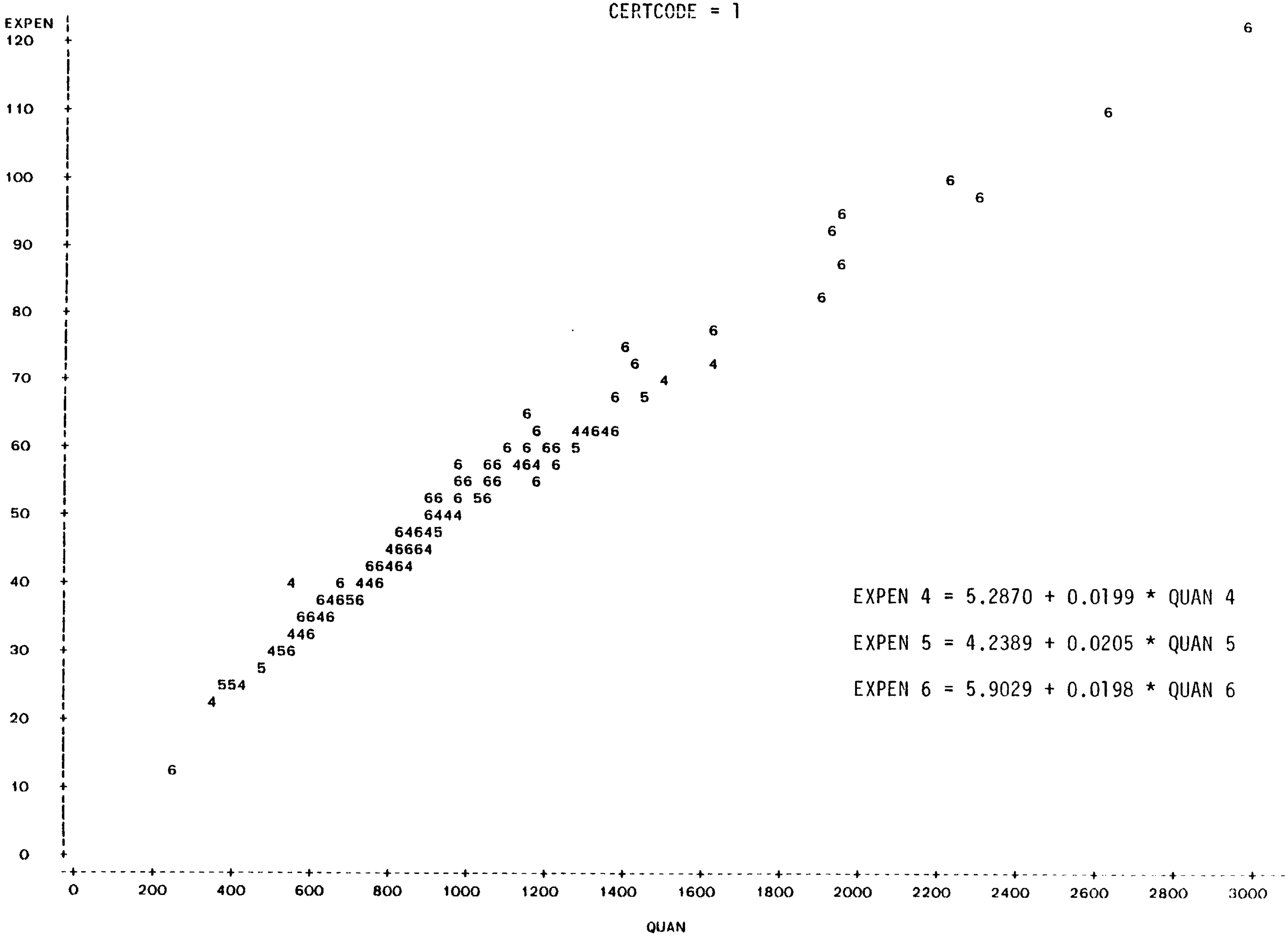


Figure 10

Springfield, Massachusetts 1978

CERTCODE = 1

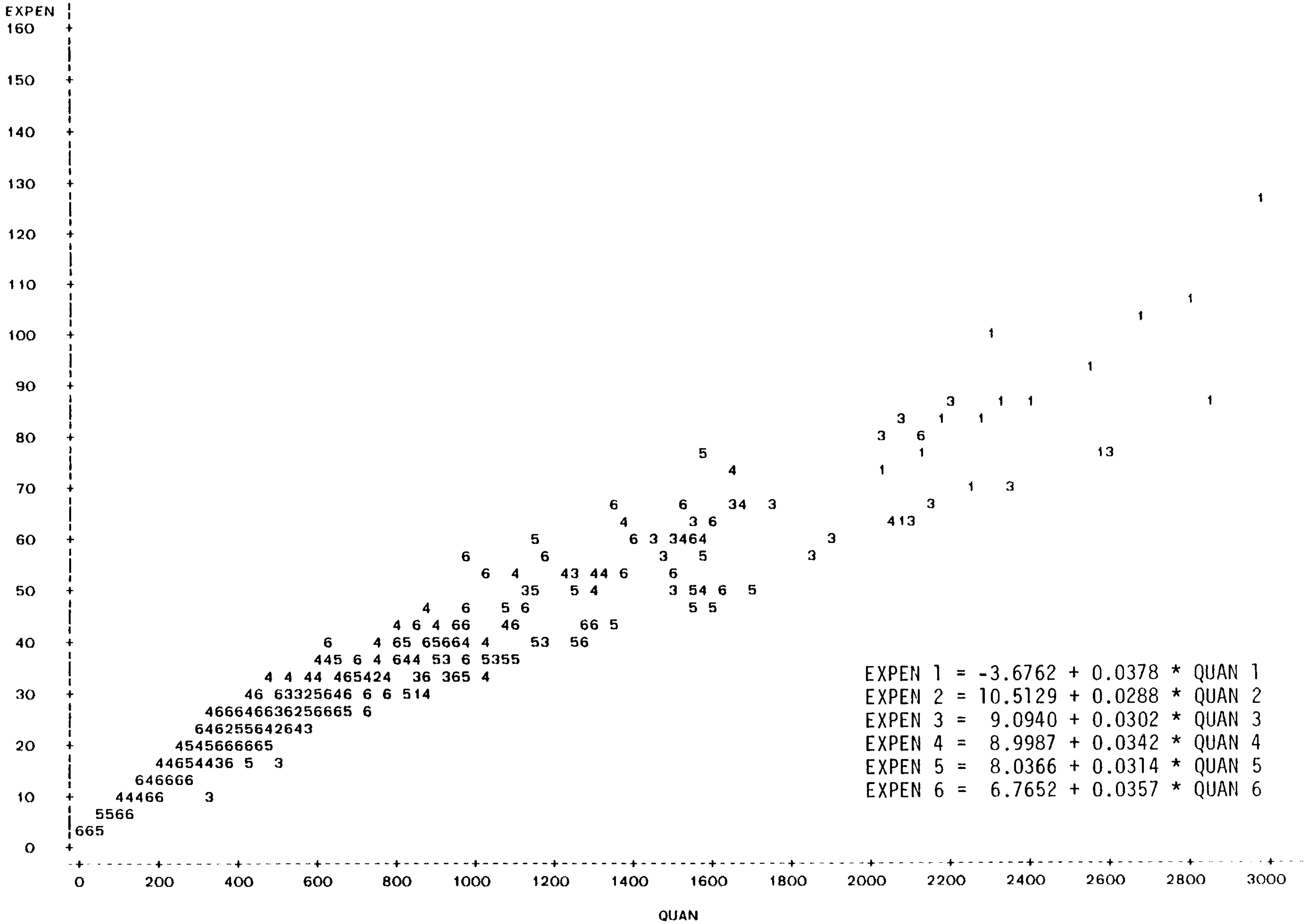


Figure 11

Christian, Illinois 1978

CERTCODE = 3

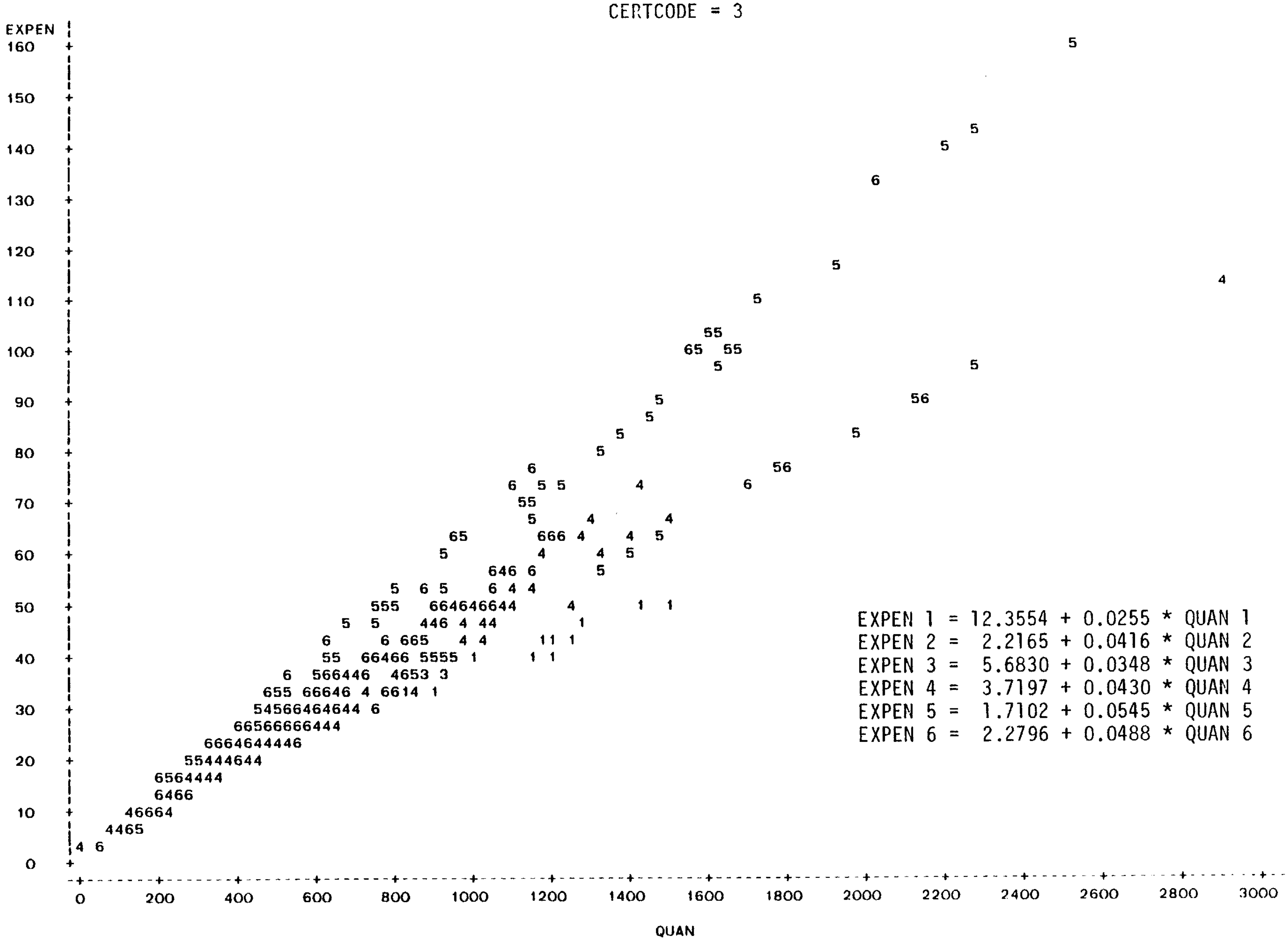


Figure 12

St. Peter-Kasota, Minnesota 1978

CERTCODE = 1

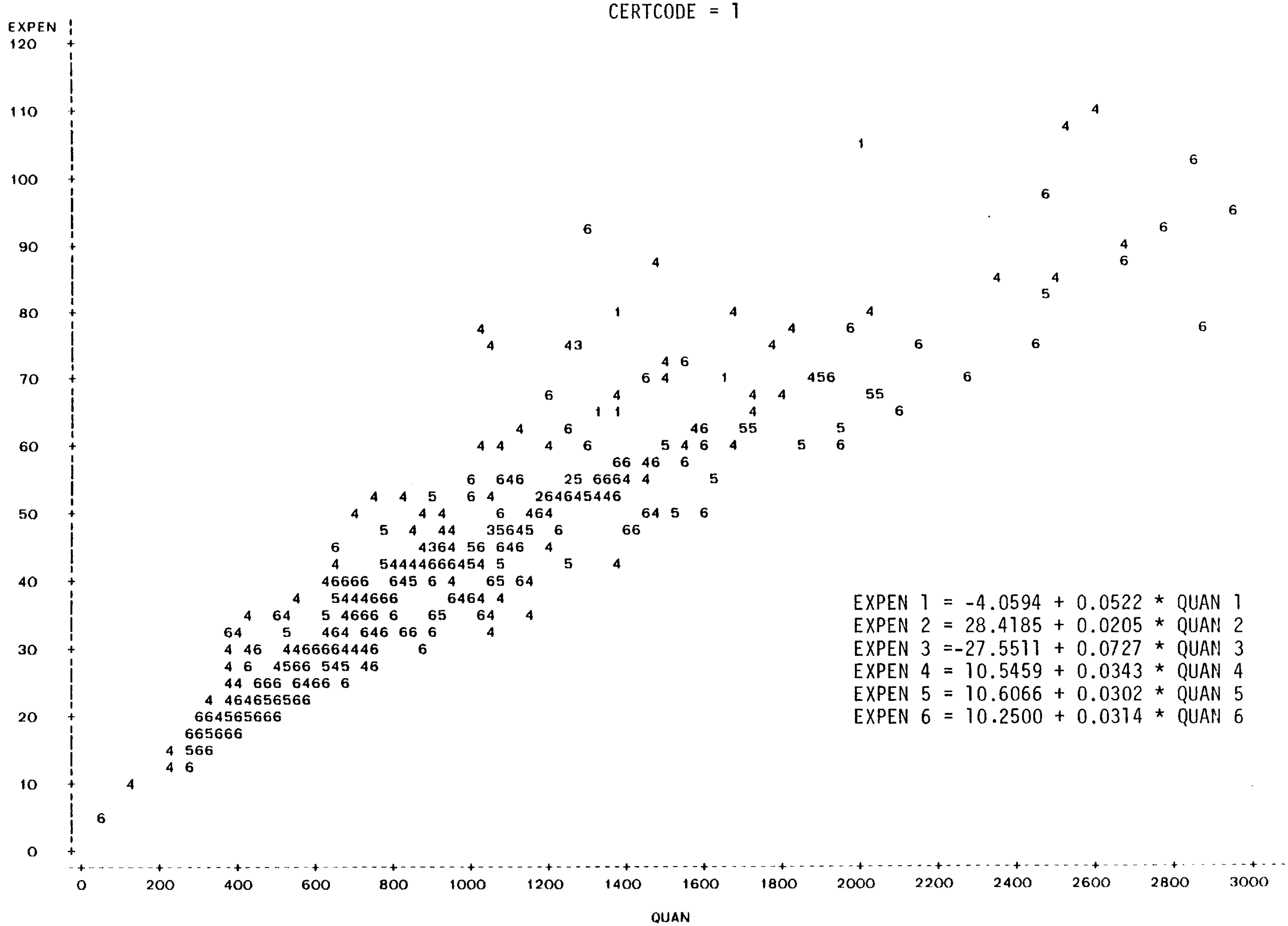


Figure 13

Flat River, Missouri 1978

CERTCODE = 1

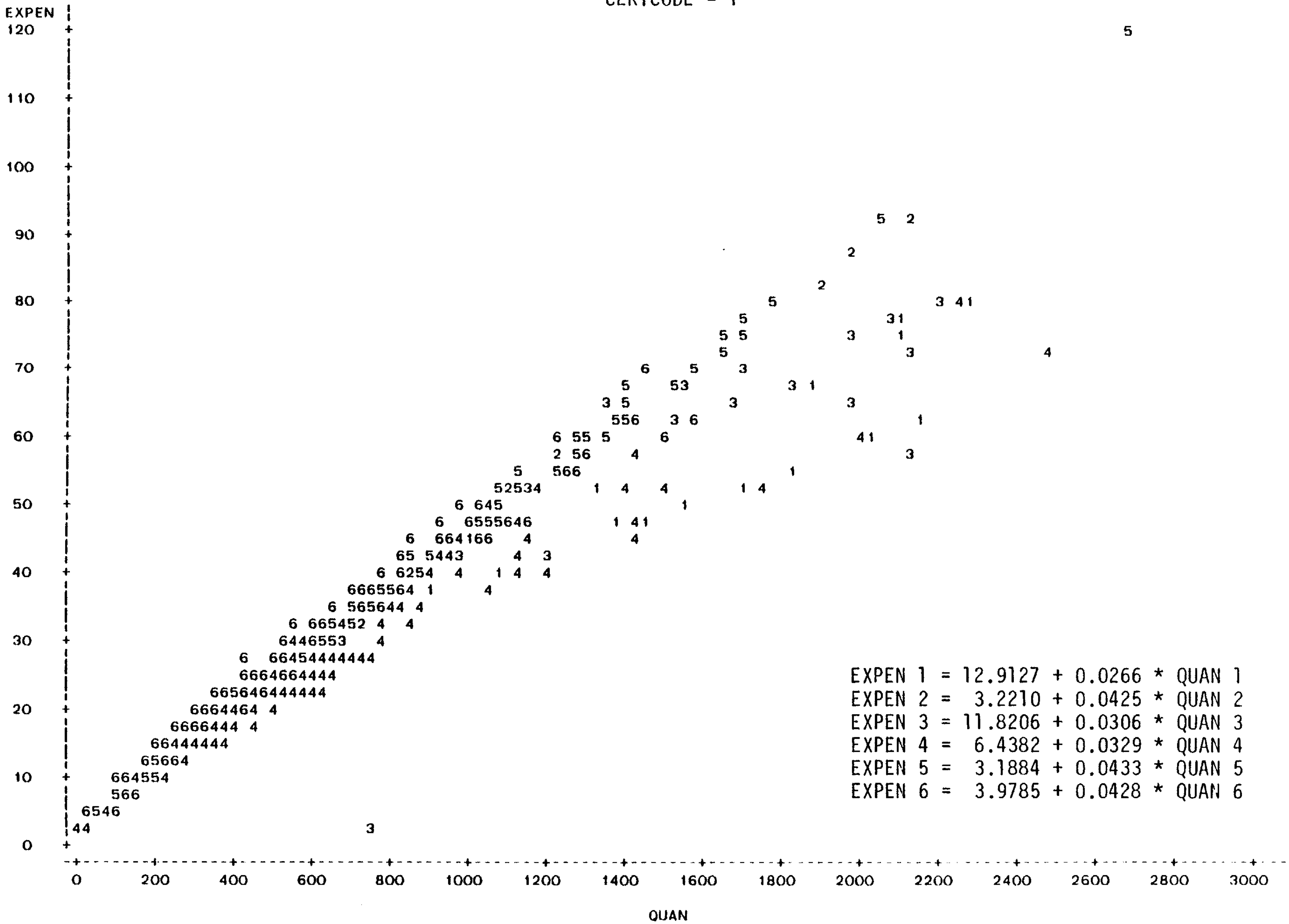


Figure 14

South Bend, Indiana 1978

CERTCODE = 1

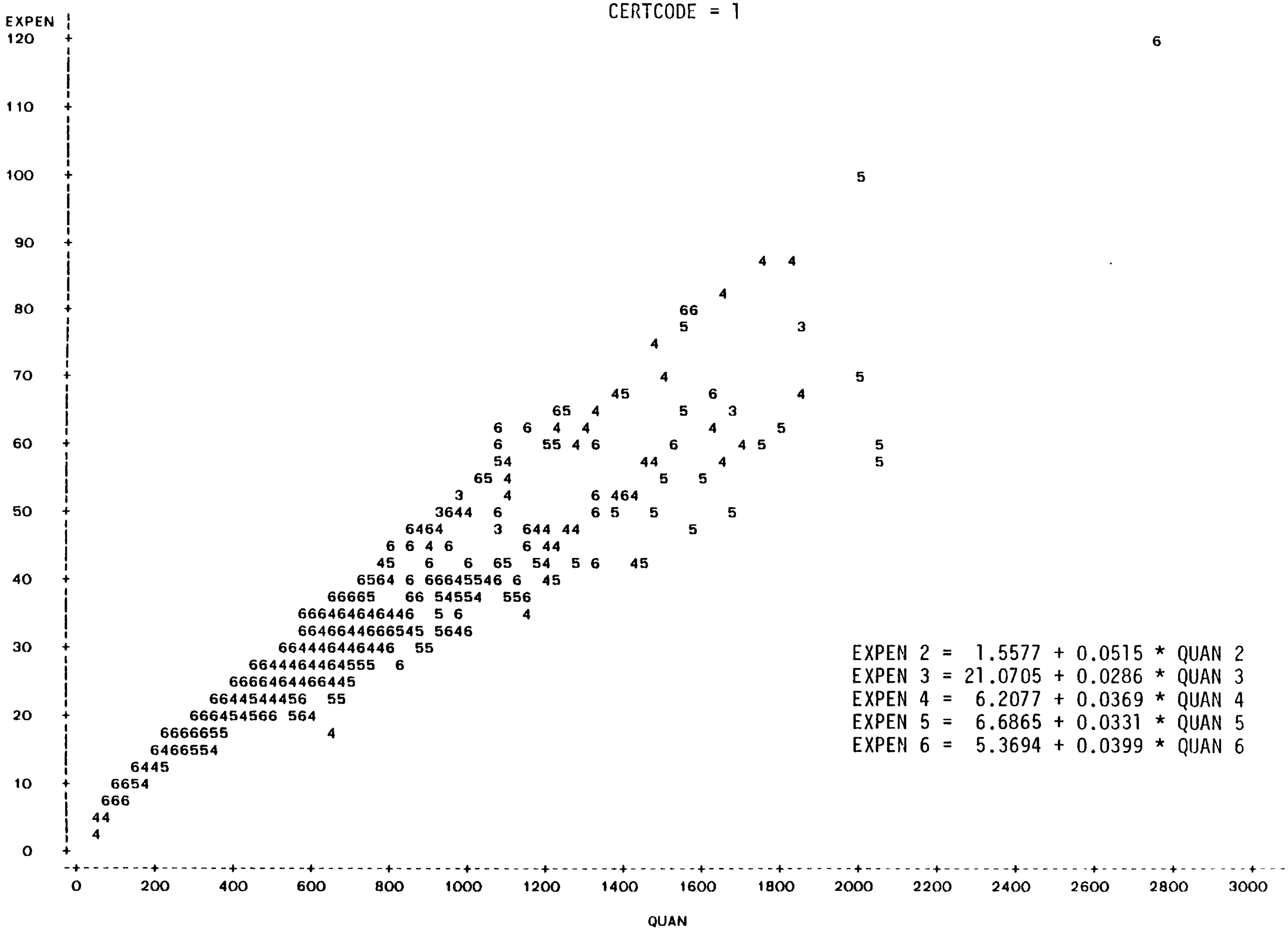


Figure 15

3 - all electric home during the off-season

4 - not all electric home in the winter

5 - not all electric home in the summer

6 - not all electric home during the off-season

All electric homes are households which have and use an electric space heating system. Winter is defined to be billing periods which begin or end in January 1978. Summer is defined to be billing periods which begin or end in July. The off-season is defined as any billing period which begins in April, October, or September or ends in April, October, or November. The resultant partition closely matches the pattern exhibited by a significant majority of utilities during 1978. In the seven figures we use the symbols 1, 2, 3, 4, 5, and 6 to indicate the observation of a quantity-expenditure pair in a particular cell. It is possible for some cells to be empty (notably all electric homes in some primary sampling units) so that not all points will be found in each figure. Finally in each cell, we have fitted a two-part tariff using least squares. Formal grouping tests are not presented as Figures 9-15 are intended to illustrate the qualitative variety of rate schedules in the NIECS data and to suggest appropriate regression strategies for the estimation of marginal price.

In Figures 9 and 10 we see little evidence of seasonal structure. However Figure 9 indicates the possibility that a winter rate may be distinguished from the rest of the season. (If one checks the national electric rate book for Newark, New Jersey 1978 this supposition is verified.) In Figure 11 we note that estimates of marginal price for all electric households do not differ significantly from those of the non-electric homes. Furthermore, seasonality in rates is not exhibited on the basis of the slope estimates.

Figure 12 provides a striking illustration of multiple rate schedules. As we pass the 1400 KWh range households in cells 5 and 6 (non-all electric; summer and off-season) appear to fall on two distinct lines. Also the slope estimates indicate a lower marginal price for all electric homes as is illustrated by the households in cell 1 which tend to cluster below all other households. (Consultation of the rate books indicates multiple rates for small and large users of electricity in the Christian, Illinois cluster.) Figure 13 yields an imprecise picture for all-electric homes due perhaps to their few numbers. The price estimates for groups 4, 5, and 6 do not appear to be significantly different. Figure 14 indicates some clustering of all-electric homes in cell 1 and the possibility of an all-electric rate. The winter rate for not all electric homes is lower than the estimated rates in cells 5 and 6 which does not indicate a winter peaking rate. Finally, Figure 15 shows a definite split in cluster 5 households while the number of all electric homes is too few to make an unbiased qualitative statement.

In summary, we see that the two-part approximation to the rate schedule provides an interesting qualitative tool to help determine the presence of seasonal and differential rate schedules. Furthermore when large numbers of observations are present the loss of efficiency from grouping observations into plausible rate cells is compensated by avoiding basic specification bias.

V. Summary and Conclusions

This chapter has reviewed the theory and estimation of price specification in the demand for electricity. We have demonstrated that (1) measured average price and measured marginal price are statistically endogenous, (2) the statistical contribution of the rate structure premium adjustment is negligible, (3) consumer behavior follows the marginal rather than the average price specification, and (4) estimated price elasticities are not significantly different using the tail-end price in place of the measured marginal rate. Finally, we have used the two-part tariff approximation to the rate schedule to provide a means of determining the presence of seasonal and all electric rate schedules.

Footnotes

1. Another source of bias not discussed in this chapter arises from the endogeneity of appliance ownership dummies. Generally, unobserved factors which influence the choice of a durable will also influence its use. For a complete discussion of this problem and evidence of resulting coefficient bias see Dubin and McFadden (1979).

2. This result is further true when p is correlated with Z. However, it is not in general possible to determine the magnitude of the bias when several explanatory variables are correlated with the error term.

3. A maintained hypothesis is that appliance dummies are exogenous. Dubin and McFadden (1979) find evidence that this leads to under estimates (in magnitude) of the true price effects. This point will be reconsidered in Chapter IV.

4. The rate schedule in Houthakker's study consisted of a connect charge and a fixed marginal price. The marginal price elasticity estimated by Houthakker is not tainted by simultaneity bias.

5. Studies by Acton, Mitchell, and Mowill (1976) and Taylor, Blattenberger, and Verleger (1977), find short-run price elasticities from -.08 to -.35 with endogenous marginal price specifications.

6. The bias for the average price specification is not as large at approximately 5%.

7. We have rejected the null hypothesis that demand for electricity follows the average price specification. This, of course, is not identical to accepting the marginal price specification. However, given the sign change on the coefficient of (RSP/Q) and its standard error we cannot reject the marginal price specification.

8. This result is likely to remain true for the NIECS survey of 1978 given the trend toward less complicated rate schedules.

CHAPTER III

Estimation of Nested Logit Model for Appliance Holdings

In this chapter we describe the estimation of a discrete choice model for room air conditioning, central air conditioning, space heating, and water heating. The data used in this study is from the recent National Interim Energy Consumption Survey of 1978. Appendix I describes references to the data set as well as extensive discussion of procedures used to prepare the data for econometric analysis.

Related discrete choice models are Dubin and McFadden (1979), Goett (1979) and McFadden, Puig, and Kirschner (1977). The model estimated here may be embedded in a larger micro-simulation system such as the Residential End-Use Energy Policy System (REEPS) for the purposes of policy forecasting.

Section II discusses the nested logit model of appliance choice and describes the particular tree extreme value form used in our analysis. Section III discusses the utility maximization problem when utility is a function of ambient temperature and the implications for components of indirect utility. Section IV, V, and VI describe the estimation of the room air conditioner, water heat, and space heat choice models. Section VII estimates the full tree structure and discusses central air conditioning choice.

II. Nested Logit Model of Appliance Choice

This section describes the tree extreme value choice model of alternative appliance portfolio combinations estimated for the NIECS data. From the onset we desired to include as many of the major household appliances in the choice system as possible. We have concentrated on the potential choices of nineteen alternative space heating and air-conditioning packages, three water heat fuel types and the choice of room air-conditioning. The possible combinations of appliance portfolios and the possible number of tree structures which might explain the observed choices are essentially limitless.

The empirical searches for nested logit forms which would produce sensible results concentrated on a subset of the nineteen alternative space heating systems. These alternatives form the trunk of the tree structure. In all, we investigated perhaps 200 logit models for space heating choice. The results of this research elicit two important ingredients in the choice process: (1) the importance of eliminating gas heating system alternatives from the choice model when gas was not available, and (2) the treatment of dominated alternatives (i.e. an alternative in which there exists another alternative which is less expensive in operating and capital costs).

Whether a household has availability to natural gas is clearly an important aspect in the decision to install a gas HVAC. Further, inclusion of gas alternatives which appear economically attractive with respect to the choice set is sure to lead to bias when households are observed to choose systems other than gas because it was not available.

Measures of gas availability were not available within the NIECS data base. To construct a measure of gas availability we followed two distinct procedures. First, a measure of gas availability existed for the Washington Center for Metropolitan studies cross-sectional data. Given our ability to link locational information (at the level of primary sampling units) from one survey to the other, we were able to match the gas availability data from WCMS to NIECS. Unfortunately, gas availability is likely to be determined at the level of city blocks or in regions which correspond to secondary sampling units (see Cowing, Dubin, and McFadden (1981a) which imparts a coarseness to a variable which is to be used at the individual level. A second problem with this procedure was that the survey year for (WCMS) was 1975 while the NIECS survey corresponds to 1978. This gap in time would tend to effect our information about households making choices post 1975.

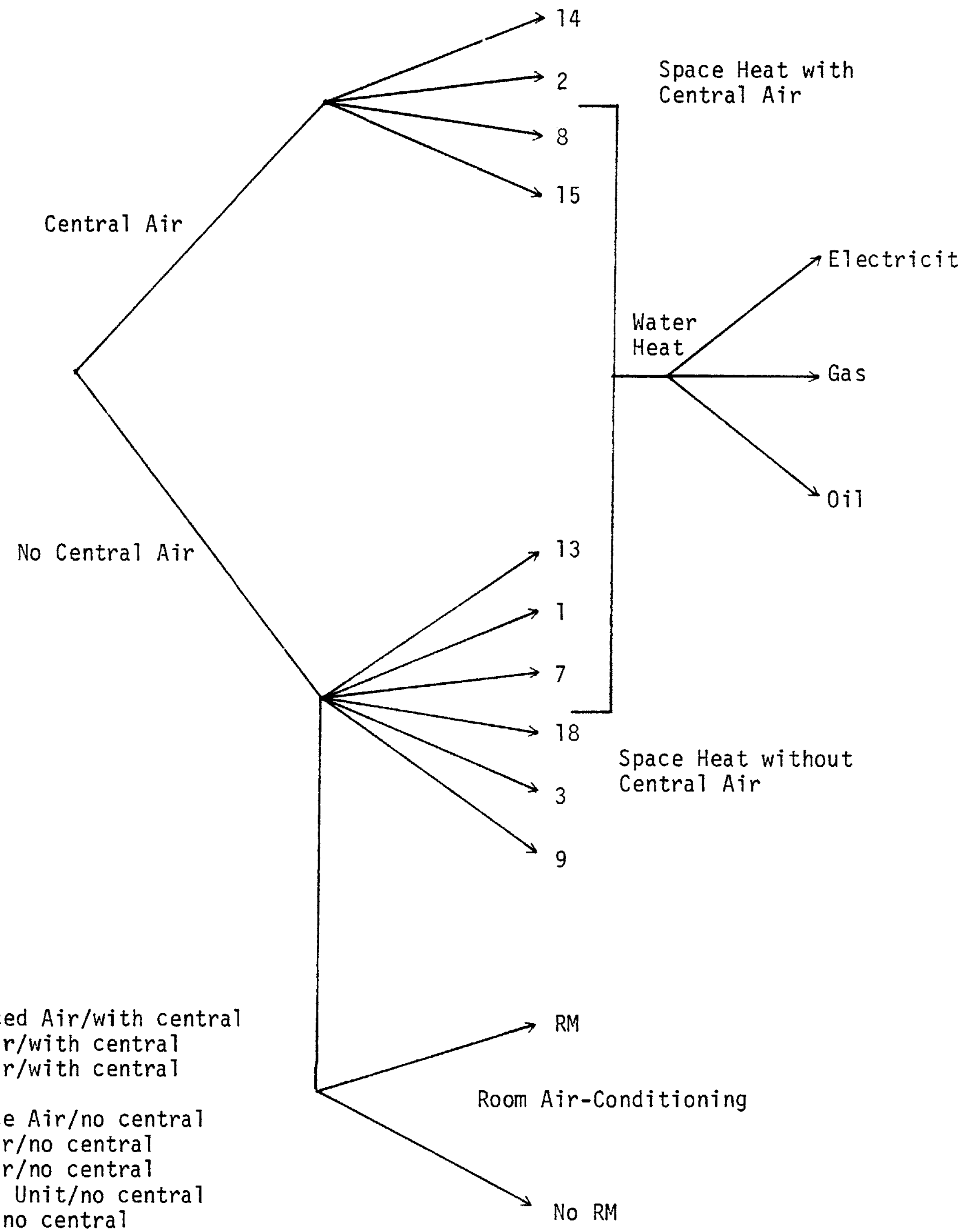
Our second procedure used natural gas related information in two NIECS variables. The first variable indicated whether the household had any gas appliances and was an index of their cumulative consumptions. The second variable indicated if the household used natural gas for any purposes. We computed the percentage of households in each secondary sampling unit which either had a positive gas index or had positive usage. Gas availability was accordingly assigned to each household in the relevant secondary sampling unit. The inherent weakness of this procedure is that it provides information on households in 1978 rather than the decision date which takes place at the point of construction.

The availability of gas is an essentially discrete phenomenon. When gas is available, gas HVAC systems are in the choice set. When gas is not available, the chosen alternative is presumed selected from alternatives which exclude gas systems. To improve our measure of gas availability we made two modifications. The first change assumes that gas is available (irrespective of our previous assignment) if a particular household chooses gas. Our second modification works in quite the opposite direction and imposes the condition of non gas availability whenever a household chooses an alternative which is dominated by a gas alternative.

In early attempts to puzzle through the tree structure of appliance choice, we located a few cases in which a household would choose an oil heating system or an electric heating system when, in fact, an all gas system would have been less expensive in terms of both operating and capital costs. For households in which we had previously assumed the availability of gas this posed an interesting problem: Why do households choose dominated alternatives? The answer might be explicable through variations in tastes across individuals yet it was most often the case that gas was the dominating non-chosen alternative and not other fuels. We resolved this issue by assuming that our discrete indicator of gas availability was incorrect for the household in question.

It was discovered quite early that alternatives which included central air-conditioning behaved quite distinctly from the set of HVAC alternatives which did not. Figure 1 illustrates the nested logit model of four space heating systems with central air-conditioning, six space heating systems without central air, water heat fuel choice, and room air-conditioning. The postulated structure assumes that water heat choice is made conditional on the choice of space heat system, that room air-conditioning is selected as an alternative to central air-conditioning (i.e. room air-conditioning is chosen only when central air is not chosen), and finally that space heat choice is made conditional on the choice of central versus no central air-conditioning.

Figure 1



14 - Electric Forced Air/with central
2 - Gas Forced Air/with central
8 - Oil Forced Air/with central
15 - Heat Pump
13 - Electric Force Air/no central
1 - Gas Forced Air/no central
7 - Oil Forced Air/no central
18 - Electric Wall Unit/no central
3 - Gas Hydronic/no central
9 - Oil Hydronic/no central

To derive a nested logit model for Figure 1 let Y_{wrsc} denote a positive measure of the desirability of alternatives indexed by wrsc where w denotes water heat choice, r indicates room air-conditioning choice, s indicates space heat choice, and c indicates central air choice. We use the notation of Appendix II and specify a probability generating function $G[<Y_{wrsc}>]$ as the composition of four generating functions to reflect the levels of the tree in Figure 1:

(1) $G[<Y_{wrsc}>] = G^c[<G^s[<G^w[<G^r[<Y_{wrsc}>]>]>]>]$.

We take logistic generating forms for G^c, G^s, G^w, and G^r so that:

(2) $G^r[<Y_{rc}>] = [\sum_r Y_{rc}^{\frac{1}{1-\phi}}]^{1-\phi}$

(3) $G^w[<Y_{wsc}>] = [\sum_w Y_{wsc}^{1/1-\sigma}]^{1-\sigma}$

(4) $G^s[<Y_{sc}>] = [\sum_s Y_{sc}^{1/1-\delta_c}]^{1-\delta_c}$

(5) $G^c[<Y_c>] = \sum_c Y_c$

From Theorem 1 of Chapter 1 it follows that:

$$P_{wrsc} = [\partial \ell n G^c/\partial \ell n G^s] \cdot [\partial \ell n G^s/\partial \ell n G^w] \cdot [\partial \ell n G^w/\partial \ell n G^r] \cdot [\partial \ell n G^r/\partial \ell n Y_{wrsc}]$$
$$= P_c \cdot P_{s|c} \cdot P_{w|sc} \cdot P_{r|wsc} = [\partial \ell n G/\partial \ell n Y_{wrsc}]$$

where P_{wrsc} denotes the probability of choosing portfolio combination wrsc and $P_{j|k}$ denotes the conditional probability of choosing alternative j given that alternative k has been chosen. To derive the structure in Figure 1 we assume that the probability of having room air conditioning conditional on HVAC choice is independent of heating system choice. Furthermore, we assume that the probability of water heat fuel choice is independent of room air-conditioning choice. To impose this structure on the probability generating function G, we let $Y_{wrsc} = Y_{wsc} \cdot Y_{rc} \cdot Y_{sc} \cdot Y_c$. This model is consistent with the assumption that households maximize utility:

(6) $U_{wrsc} = V_{wrsc} + \varepsilon_{wrsc}$

where: $V_{wrsc} = \ell n Y_{wrsc}$ denotes the strict utility of alternative wrsc and $<\varepsilon_{wrsc}>$ have a joint generalized extreme value distribution. Note that the assumption $Y_{wrsc} = Y_{wsc} \cdot Y_{rc} \cdot Y_{sc} \cdot Y_c$ implies that strict utility may be written as $\ell n Y_{wsc} + \ell n Y_{rc} + \ell n Y_{sc} + \ell n Y_c = V_{wsc} + V_{rc} + V_{sc} + V_c$ which exhibits the decomposition of the components of indirect utility. The generating function under the conditional independence assumption has the form:

(7) $G[Y_{wrsc}] = G^c[<Y_c G^s[<Y_{sc} G^w[<Y_{wsc}>]] \cdot G^r[<Y_{rc}>]>]$.

It is possible to show that:

(8) $P_{r|c} = e^{V_{rc}/1-\phi} \Big/ \sum_r e^{V_{rc}/1-\phi} \equiv P_{r|wsc}$

(9) $P_{w|sc} = e^{V_{wsc}/1-\sigma} \Big/ \sum_w e^{V_{wsc}/1-\sigma}$

(10) $P_{s|c} = e^{(V_{sc} + J_{sc}(1-\sigma))/(1-\delta_c)} \Big/ \sum_s e^{(V_{sc} + J_{sc}(1-\sigma))/(1-\delta_c)}$

(11) $P_c = e^{(J_c^s(1-\delta_c) + V_c + J_c^r(1-\phi))} \Big/ \sum_c e^{(J_c^s(1-\delta_c) + V_c + J_c^r(1-\phi))}$

where:

(12) $J_{sc} \equiv \ell n[\sum_w e^{V_{wsc}/1-\sigma}]$

(13) $J_c^s \equiv \ell n[\sum_s e^{(V_{sc} + J_{sc}(1-\sigma))/(1-\delta_c)}]$

and

(13) $J_c^r \equiv \ell n[\sum_r e^{V_{rc}/1-\phi}]$

The terms J_c^s, J_c^r, and J_{sc} are respectively the inclusive values of space heat choice given central air choice, room air choice given central air choice, and water heat choice given space heat and central air choice. Furthermore, $(1-\phi)$, $(1-\delta_c)$, and $(1-\sigma)$ are the corresponding inclusive value coefficients. We have allowed the inclusive value coefficient $(1-\delta_c)$ to be different depending on central air choice to reflect a possible dissimilarity in the degree of association in the space heat choice branches. Estimation of the central air-conditioning choice model should identify the coefficients δ_c.

III. Residential Heating and Comfort

Let u[t,Z] denote the utility derived from consumption of a vector of goods Z in an environment with ambient temperature t. It is reasonable to assume that utility is increasing in t up to a temperature T* which provides bliss comfort. Below T* occupants feel too cool and above T* feel too hot. If heating were a free good consumers would set their thermostats at T*. However as heating to an interior temperature T* requires a costly energy input there exists a trade-off between the comfort of the ambient space and the price of obtaining this comfort.

Follwing Brownstone (1980) and Hausman (1979) assume that the utility function u[t,Z] is separable in comfort and goods consumption and suppose that u[t], the utility derived from ambient temperature t, takes the linear form $u[t] = -\alpha[T^*-t]$ for $\alpha > 0$ and $t \leq T^*$. Let F[t] denote the cummulative distribution for the number of days during the heating season in which the daily mean temperature is less than or equal to t. Utility during the heating season from thermostat setting τ is:

$$u[\tau] = \int_{-\infty}^{\tau} -\alpha(T^*-\tau)\, F'(t)dt + \int_{\tau}^{T^*} -\alpha(T^*-t)\, F'(t)dt \qquad (15)$$

The first integral assumes that comfort is constant at the level $(T^*-\tau)$ degrees per hour when outside temperature is below the thermostat level τ. The second integral assumes that comfort increases proportionally to increases in temperture below the bliss temperature point. It is straightforward to demonstrate that equation (15) has an interpretation measured in degree days of heating. From equation (15):

$$u[\tau] = -\alpha\left[(T^*-\tau)F(\tau) + T^*(F(T^*) - F(\tau)) - \int_{\tau}^{T^*} tF'(t)dt\right]$$

$$= -\alpha\left[T^*F[T^*] - \tau F[\tau] - \int_{\tau}^{T^*} tF'(t)dt\right]$$

$$= -\alpha\left[(T^*F[T^*] - \int_{-\infty}^{T^*} tF'(t)dt) - (\tau F(\tau) - \int_{-\infty}^{\tau} tF'(t)dt)\right]$$

$= \alpha[H(\tau) - H(T^*)]$ where $H(t_0)$ denotes total heating degree days measured at base t_0, i.e.

$$H[t_0] = \int_{-\infty}^{t_0} (t_0-t)F'(t)dt = t_0F(t_0) - \int_{-\infty}^{t_0} tF'(t)dt$$

Suppose that the BTUH heating required to maintain an interior temperature τ when exterior temperature t is given by the function $Q(\tau-t)$. Let $B(\tau)$ denote the seasonal heating load resulting from thermostat setting τ. Then:

$$B[\tau] = \int_{-\infty}^{\tau} \text{MAX}[Q[\tau-t],0]\ F'(t)dt \qquad (16)$$

We now consider the optimization problem of maximizing the utility function $U[\tau,Z]$ subject to a budget constraint which takes the heating load $B[\tau]$ into account.

The consumer's choice problem is to maximize utility subject to the budget constraint which allocates wealth W between expenditures on goods Z and on fuel $(P_i/e_i)B(\tau)$ where P_i is the price of fuel i and e_i is the efficiency of the heating system using fuel i. We write:

$\underset{\tau,\ Z}{\text{maximize}}\ U[\tau,Z]$ subject to $(P_i/e_i)\ B[\tau] + Z \leqq W$ for which the

Lagrangian (with multiplier ξ) is:

$L = U[\tau,Z] + \xi[W - Z - (P_i/e_i)\ B(\tau)]$. The first order conditions are:

$L_\tau = U_\tau - \xi(P_i/e_i)\ B'(\tau) = 0$ and

$L_Z = U_Z - \xi = 0$ so that:

$$\frac{U_\tau}{U_Z} = (P_i/e_i)B'(\tau) \tag{17}$$

We see from (17) that the price of comfort depends on the level of comfort. It is possible to re-formulate the optimization problem by using an appropriately defined rate structure premium. Let τ^* denote the solution to (17) so that $(P_i/e_i)[B(\tau^*)-B'(\tau^*)\tau^*]$ is the rate structure premium adjustment which standardizes the optimization problem. The equivalent standarized problem is then:

$$\underset{\tau,\ Z}{\text{Maximize}} \quad U[\tau,Z] \text{ subject to } [(P_i/e_i)\ B'(\tau^*)\tau] + Z \leq$$

$$W - (P_i/e_i)[B(\tau^*) - B'(\tau^*)\tau^*] \tag{18}$$

The indirect utility associated with equation (18) is a function of W and the price of comfort $(P_i/e_i)B'(\tau^*)$. The thermal model discussed in McFadden ad Dubin (1982) was used to estimate the price of comfort for alternative HVAC systems. The procedure approximates the derivative $B'(\tau^*)$ by calculating the change in seasonal utilization associated with a one degree change in the thermostat setting. In our empirical work we ignore the RSP adjustment to W of equation (18).

IV. Room Air Conditioner Choice Model

This section describes the estimation of the choice model for room air conditioning. The analysis considers only the choice of room air conditioning as a cooling alternative to central air conditioning and does not consider either the choice of the number of room air conditioning units or their efficiencies. For detials concerning these latter aspects of the choice process see Brownstone (1980) and Hausman (1979). In the NIECS data set we are provided with information about the number of room air conditioners owned by the household and the number of rooms air conditioned but no information is available on individual room air conditioner efficiency.

The thermal model of McFadden and Dubin (1982) may be used to provide estimates of air conditioning design capacity. Design capacity measures the thousands of BTU's per hour required to maintain a given household at summer design temperatures. Our allocation of capital costs to central air conditioning units assumes that households purchase units of design capacity. We follow the same procedure for room air conditioners and assume that room air conditioners are purchased to meet design cooling loads.

More precisely we have assumed that the total cooling load in the residence is distributed equally among the number of rooms in the residence and have then determined the capital costs (materials and installation) for providing one room air conditioning unit per room. Casual empiricism suggests this is a departure from average behavior yet the assumption allows us to determine total capital costs in a manner which recognizes substantial returns to scale in purchasing larger air conditioning units. For additional details concerning the construction of room air-conditioning costs the reader is referred to Cowing, Dubin, and McFadden (1981e).

Consistent with our determination of room air-conditioning capital costs we have assumed that operating costs for room units distributing the total load are identical to those for a central air-conditioning system. This assumes (perhaps unrealistically) that room air conditioners operated in parallel are as efficient as central systems.

Table 1 presents the mean values of variables used in the discrete choice model.

Table 1

Variable	Description	Mean[a]
RMOPCST	Operating Cost for Room Air-Conditioning (1967$)	71.07
RMCPCST	Capital Cost for Room Air-Conditioning (1967$)	997.60
RMOPCST1	RMOPCST/(Base Load Usage)	0.00819
RMCPCST1	RMCPCST/(Base Load Usage)	0.2737
CDD78	Cooling Degree Days in 1978	1110
RINCOME	Income $(1967\$)/10^3$	10.38
NHSLDMEM	Number of Household Members	3.3

[a]Sample size 770 households corresponds to the set of single family detached owner occupied dwelling built since 1955 which do not have central air-conditining. 591 of these homes appear in the nested logit model of HVAC system choice.

Following the discussion in Section III, we would expect, other things equal, that the probability of choosing room air-conditioning given that the household does not have central air-conditioning should increase with income and decrease as operating and capital costs increase. We have attempted an empirical specification in which these variables are interacted with the "purchase" alternative. In the "no purchase" alternative we enter the number of household members and cooling degree days with the latter a measure of the discomfort the household suffers in not having any air-conditioning. The results are presented in Table 2. RINC1, CDD2, and PERS2 are RINCOME, CDD78, and NHSLDMEM interacted with alternative specific dummies for alternative one, alternative two, and alternative two respectively. A1 is the alternative one specific dummy.

Table 2

Binary Logit Model of Room Air-Conditioning Choice

Given No Central Air-Conditioning[a]

Alternative 1	-	Purchase Room Air-Conditioning	45.06 percent
Alternative 2	-	Do Not Purchase Room Air-Conditioning	54.94 percent

Variable Name	Logit Estimate	Standard Error	T-Statistic
RMOPCST	.2683E-02	.3615E-02	.7421
RMCPCST	.2121E-04	.3286E-03	.6453E-01
RINC1	.3619E-01	.1453E-01	2.490
CDD2	-.9832E-03	.1828E-03	-5.379
PERS2	.3047E-01	.4930E-01	.6180
A1	-1.759	.3434	-5.121

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-471.8	-533.7
Percent Correctly Predicted[b]	70.00	50.00

[a]Estimation is by maximum likelihood using the QUAIL (Qualitive, Intermittent, and Limited Dependent Variable Statistical Program) developed by Daniel McFadden and Hugh Wills.

[b]A case is taken as being correctly predicted when the chosen alternative is forecast to have the highest probability of being chosen.

The insignificance of the operating and capital cost coefficients in Table 2 follows the pattern of results obtained by Goett (1979). It is possible to offer a few possible reasons for this result: 1) measurement error would tend to bias these coefficients to zero and is likely given the assumptions made in assigning operating and capital costs, 2) the desirability of room air-conditioning is likely to be greatest when the cooling load is greatest introducing a spurious correlation between operating and capital costs and room air-conditioning purchases, and 3) operating and capital costs really are not significant determinants of the choice of room air-conditioning given that the household has chosen not to purchase central air-conditioning and income and cooling degree days adequately model the true choice process. It is likely that the insignificance appears due to all three effects. It is possible however to investigate the second effect in more detail.

In Table 3 we present the room air-conditioning choice model where we have normalized the operating and capital costs variables by the scale variable of expected base load usage (ACUEC). Note that the operating cost variable is now signficant but of the unexpected sign while the normalized capital cost variable remains insignificant. The significance of the normalized operating cost variable may be attributable to a regional effect in which the largest average costs of room air-conditioning are associated with regions in which there is a summer peaking marginal electricity price. The summer peak rate is again associated with high average loads per customer due to the presence of very high ambient temperatures and a large percentage of homes using air-conditioning.

Table 3

Binary Logit Model of Room Air-Conditioning Choice

Given No Central Air-Conditioning

Normalized Operating and Capital Costs

Alternative 1 - Purchase

Alternative 2 - Do Not Purchase

Variable Name	Logit Estimate	Standard Error	T-Statistic
RMOPCST1	108.8	33.51	3.247
RMCPCST1	.5335E-01	.5774E-01	.9240
RINC1	.3824E-01	.1435E-01	2.664
CDD2	-.1134E-02	.1245E-03	-9.110
PERS2	.9395E-02	.4889E-01	.1922
A1	-2.713	.4050	-6.699

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-467.0	-533.7
Percent Correctly Predicted	68.83	50.00

Given the essentially unchanged log likelihood and percentage correctly predicted we adopt the cleaned specification presented in Table 4 for use in the the estimation of the HVAC choice tree. Corresponding to the parameter estimates in Table 4 we have constructed the inclusive value of room air conditioning choice for our sample of 911 households. The mean value of RMINCV [room air-conditioning inclusive value] is -.5041 with standard deviation 0.4023.

Table 4

Binary Logit Model of Room Air-Conditioning Choice

Given No Central Air-Conditioning

No Operating or Capital Costs

Alternative 1 - Purchase

Alternative 2 - Do Not Purchase

Variable Name	Logit Estimate	Standard Error	T-Statistic
RINC1	.3765E-01	.1380E-01	2.729
CDD2	-.1104E-02	.1190E-03	-9.281
A1	-1.796	.2322	-7.732

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-472.6	-533.7
Percent Correctly Predicted	70.26	50.00

V. Water Heat Choice Model

This section describes the water heat fuel choice model conditional on choice of space heating system fuel type. Related studies are Dubin and McFadden (1979) and Goett (1979). We begin with a review of the construction of operating and capital costs.

1. Water Heat Operating Costs

We define the end-use service of water heating to be a gallon of heated water. To determine energy service ratios (ESR) we used the March 1978 Consumer Report which reviewed eleven electric and twelve gas water heaters. Consumer Reports determined annual consumption in KWH per year and therms per year for electric and gas units respectively based on 100 gallons of hot water consumption per day. We use the mean value of annual consumption accross models to calculate ESR by fuel type. For electric water heaters the energy-to-service ratio is:

$$\left(10434.55 \ \frac{\text{KWH}}{\text{Yr.}}\right)\left(\frac{1 \text{ Yr.}}{365 \text{ days}}\right)\left(\frac{1 \text{ day}}{100 \text{ gal.}}\right) = 0.28588 \text{ KWH/gal.}$$

and for gas water heaters the energy service ratio is:

$$\left(502.33 \ \frac{\text{Therms}}{\text{gas}}\right)\left(\frac{1 \text{ Yr.}}{365 \text{ days}}\right)\left(\frac{1 \text{ day}}{100 \text{ gal.}}\right) = 0.01376 \text{ Therms/gal.}$$

Following Dubin and McFadden (1979) we assume that oil water heaters are 74 percent as efficient as electric water heaters. Conversion to units of thousand of BTU's per gallon heated implies energy service ratios: 1.376-gas, 0.97542-elec., and 1.318-oil. To determine expected usage we use the relation:

Average usage in KWH for hot water heating = (2819. + 360. * (NHSLDMEM-2)
360. * (If NHSLDMEM equals 1))
+ 365. * 3.98 * HELDISHW

This relationship is discussed in Dubin and McFadden (1979). Note that NHSLDMEM and HELDISHW are NIECS variables which are the number of household members and a dummy variable indicating that the household has a dishwasher, respectively.

Finally, operating costs by fuel type are the product of (1) expected annual usage, (2) the ratio of the ESR of the fuel under consideration to the ESR of the electric water heater, and (3) the price of the fuel in real year built dollars.

2. Water Heat Capital Costs

Construction of water heating capital costs requires a relationship between assumed capacity and structural characteristics of the dwelling and family. We follow the recommended practice ("Handbook of Buying 1978," Consumer Research Magazine) of relating capacity utilization to the number of bathrooms and the number of bedrooms (a proxy for number of persons). This relationship includes allowance for recovery rate differential which occurs between fuel types. Materials and installation costs for different capacity water heaters are obtained from MEANS (1981). These estimates do not include the costs of vent for gas and oil water heaters. To obtain vent costs for each water heater, we consulted the National Construction Estimator (Craftsman Book Co., Solano Beach, CA 1978) and determined that in 1981 dollars material costs would be $18 while installation costs would be $26. The National Construction Estimator also indicated electrical contracting charges of $145 and $161 for water heaters with capacity on either side of 40 gallons. These costs were included in the installation costs obtained from MEANS (1981). Finally, we have included all cost components which are conditional on the type of space heating system installed. When space heating type is gas or electric, the costs for material and installation of an oil tank are included with the costs of oil water heating. When space heating type is gas or oil an additional charge of $112 is added to the labor costs of the electric water heater due to the installation of increased amp service. (National Construction Estimator, 1978). Other charges for all systems are assumed reflected in the cost of the heating systems.

3. Estimation of Water Heat Choice Model

In Table 5 we present the mean values of variables used in the choice model as well as their descriptions.

Estimation is based on a sample of 1158 households who live in single family owner occupied dwellings built since 1955 and who choose either electric, gas, or oil water heaters. As discussed above the gas alternative is removed from the choice set whenever natural gas is unavailable to the household.

We attempted two basic specifications. The first specification included water heat operating and capital costs as well as space heat fuel type dummies interacted with the alternatives. This specification provided generally wrong signs on variables and was difficult to interpret. Our preferred specification used the operating and capital cost variables in normalized form (i.e. divided by expected utilization). We present the results of the normalized model in Table 6. Note that normalized operating and capital costs may be interpreted directly as service prices (price per gallon of hot water heated) and capital cost per unit of service.

All variables other than income appear highly significant. In Table 7 we present the identical choice model without the income variables. The ratio of the capital to operating cost coefficients implies a discount

TABLE 5

Mean Values of Variables in Water Heat
Choice Model (1967 Dollars)

Variables		Description	Mean
WHOPCST	(1)	water heat operating costs	111.30
WHOPCST	(2)	(by alternative)	27.78
WHOPCST	(3)		16.40
WHOPCST1	(1)	water heat operating cost divided	0.02766
WHOPCST1	(2)	by usage (by alternative)	0.006428
WHOPCST1	(3)		0.00404
WHCPCST	(1)	water heat capital cost (by	193.20
WHCPCST	(2)	alternative)	129.00
WHCPCST	(3)		582.30
WHCPCST1	(1)	water heat capital cost divided	0.04941
WHCPCST1	(2)	by usage (by alternative)	0.03343
WHCPCST1	(3)		0.1509
SHE	(1)	(space heat fuel electricity)*(ALT1)	0.1649
SHG	(2)	(space heat fuel gas)*(ALT2)	0.4931
SHO	(3)	(space heat fuel oil)*(ALT3)	0.1589
RINCOME			11.52

TABLE 6

Three Alternative Multinomial Logit Model of Water Heat Fuel Choice Given Space Heat Fuel Choice[a]

Alternative	Frequency Label	Percent of Cases	Frequency Chosen	Percent Chosen[b]
1.000	1158.	100.0	451.0	38.95
2.000	834.0	72.02	652.0	78.18
3.000	1158.	100.0	55.00	4.750

Variable Name	Logit Estimate	Standard Error	T-Statistic
WHOPCST1	-83.32	13.09	-6.365
WHCPCST1	-19.79	7.208	-2.745
RINCOME1	-.1739E-02	.2571E-01	-.6762E-01
RINCOME2	.5122E-02	.2754E-01	.1860
A1	3.791	.6618	5.728
A2	1.891	.7264	2.604
SHE	1.458	.4046	3.602
SHG	2.182	.2398	9.102
SHO	1.593	.4323	3.685

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-413.3	-1141.
Percent Correctly Predicted	84.80	37.99

[a]Note that the natural gas alternative appears in approximately 72 percent of the cases. The remaining 28 percent cases are binary choices between the electric and oil water heat alternatives as gas is unavailable.

[b]Percentage of chosen cases for included alternatives.

TABLE 7

Three Alternative Multinomial Logit Model of Water Heat
Fuel Choice Given Space Heat Fuel Choice - Normalized Costs
Without Income Variables

Variable Name	Logit Estimate	Standard Error	T-Statistic
WHOPCST1	-83.54	13.04	-6.406
WHCPCST1	-19.87	7.060	-2.814
A1	3.775	.5785	6.525
A2	1.938	.6239	3.106
SHE	1.440	.4018	3.584
SHG	2.198	.2365	9.295
SHO	1.592	.4313	3.692

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-413.4	-1141.
Percent Correctly Predicted	84.37	37.99

factor of 23.8 percent. We use the choice model in Table 7 in the estimation of the HVAC choice tree. Table 8 gives the mean and standard deviation of the inclusive values of water heat choice conditioned on space heat fuel type for the sample of 911 households. The calculation of the inclusive values correctly accounts for the availability of natural gas. Thus, when gas is not available the inclusive value corresponds to the electric and oil alternatives only.

TABLE 8

Inclusive Values of Water Heat Choice Given Space Heat Fuel Choice

Variable	Water Heat Inclusive Values Given	Mean	Standard Deviation
WHINCVE	Electricity	2.308	0.5928
WHINCVG	Natural Gas	1.177	0.5230
WHINCVO	Oil	1.318	0.5207

VI. Space Heat System Choice

In McFadden and Dubin (1982) and Cowing, Dubin, and McFadden (1981e) nineteen alternative heating ventilating air-conditioning systems are considered which provide combinations of heating and cooling capacity matched to design temperature conditions. We list the nineteen alternative HVAC systems in Table 9.

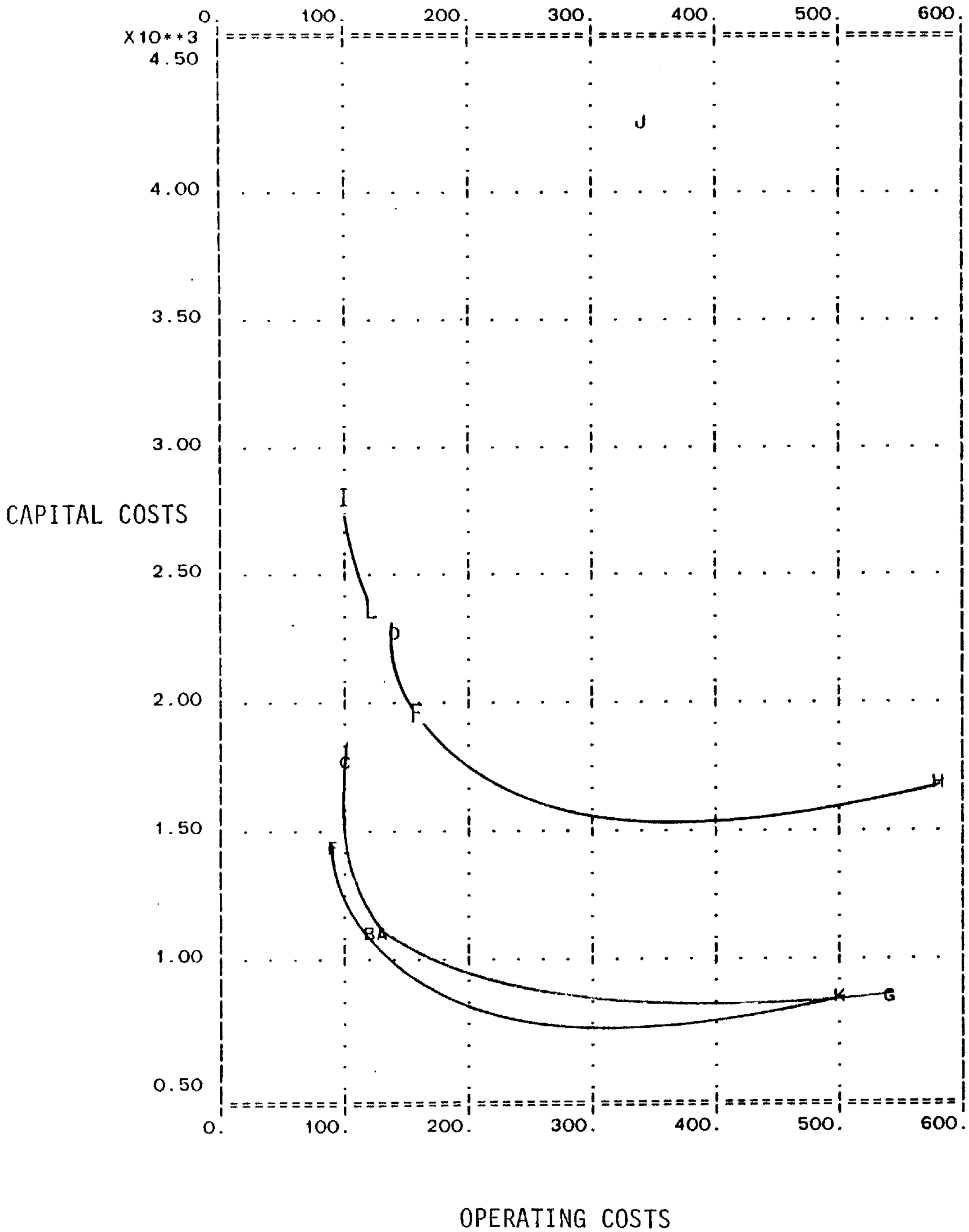
Seven of the nineteen HVAC have very small sample frequencies and are not considered further (4, 6, 10, 12, 16, 17, 19). We illustrate the capital operating cost trade-offs represented by HVAC systems in Figure 2. Prices are converted to 1967 dollars by cost indices from the actual year built costs (see McFadden and Dubin (1982)). During the post 1955 period, operating costs for oil systems were less expensive in real terms than operating costs for gas systems. This situation changed dramatically in the post 1972 period as illustrated in Figure 3.

From Figure 2 we see that baseboard and wall unit systems tend to be dominant in the sense that they have both lower operating and capital costs than other systems. However, wall units (especially gas and oil) are relatively infrequently selected. One explanation is that non-pecuniary aspects of these systems make them unattractive for installation. It is more reasonable to assume, however, that our assignment of costs to the non-central systems are mismatched due to survey ambiguities. Based on these considerations and various attempts with specifications of choice models which included these alternatives, we have opted to eliminate gas and oil wall units from the analysis. The remaining set of ten HVAC systems represent the choices of 911 single-family detached owner occupied households built since 1955. Four of the ten alternatives include central air-conditioning and the sample is selected so that households choosing central air-conditioning use

TABLE 9

HVAC System	Frequency[a]	Description
1	0.2676	Gas Forced Air / No Central Air
2	0.1234	Gas Forced Air / Central Air
3	0.0639	Gas Hot Water / No Central Air
4	0.00496	Gas Hot Water / Central Air
5	0.1214	Gas Wall Unit / No Central Air
6	0.00396	Gas Wall Unit / Central Air
7	0.09118	Oil Forced Air / No Central Air
8	0.02725	Oil Forced Air / Central Air
9	0.06838	Oil Hot Water / No Central Air
10	0.00396	Oil Hot Water / Central Air
11	0.01933	Oil Wall Unit / No Central Air
12	0.00050	Oil Wall Unit / Central Air
13	0.01288	Elec. Forced Air / No Central Air
14	0.03023	Elec. Forced Air / Central Air
15	0.01685	Electric Heat Pump
16	0.00149	Elec. Hot Water / No Central Air
17	0	Elec. Hot Water / Central Air
18	0.05401	Elec. Baseboard / No Central Air
19	0.00694	Elec. Baseboard / Central Air

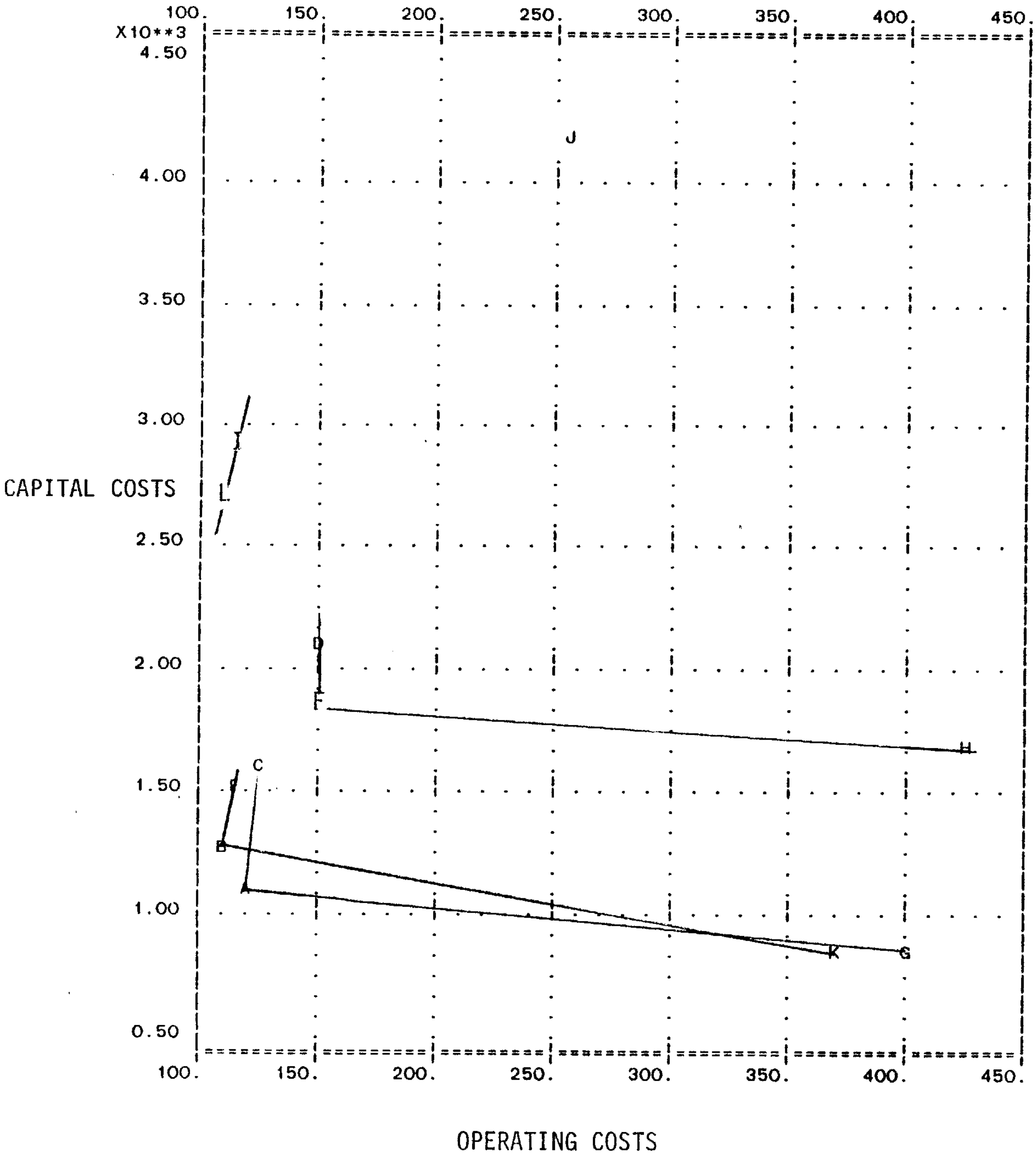
[a]Based on the sample of 2018 owner occupied single-family detached dwelling built since 1955.



HVAC #	
J	15
I	9
L	3
D	8
F	2
H	14
C	7
A	1
G	13
E	11
B	5
K	18

Figure 2

Capital versus operating costs for alternative HVAC systems - sample mean values in the post 1955 period.



HVAC #	
J	15
I	9
L	3
D	8
F	2
H	14
C	7
A	1
G	13
E	11
B	5
K	18

Figure 3

Capital versus operating costs for alternative - HVAC systems - sample mean values in the post 1972 period.

electricity as the primary fuel (a small fraction of homes used gas central air-conditioning). The two branches of the space heat choice model are illustrated in Figure 1 of Section II.

Table 10 presents the mean values of variables used in the choice models. The variables SHOPCST and SHCPCST are calculated using annual predictions of usage and capacity developed in the thermal model. Operating and capital costs for alternatives which include air-conditioning reflect additional costs associated with the central air conditioner and any economies that result from shared costs. For details the reader is referred to Cowing, Dubin, and McFadden (1981e). The variables SHOPCST1 and SHOPCST2 are SHOPCST divided by two scaling factors: expected usage (SHUECE) and the operating cost of HVAC 18. The empirical analysis determined that either method of scaling provided adequate results. Furthermore, the scaled variables have strong intuitive appeal. Consider the operating cost of system j:

$SHOPCST_j = (SHUECE)(D_j)(1/COP_j) \cdot P_j$ where

$SHOPCST_j$ = operating cost of system j

SHUECE = base load usage estimate (delivered BTU's)

D_j = adjustment factor for delivery system losses

COP_j = coefficient of performance for system j

P_j = price of fuel used by system j

The normalization rules imply:

$$SHOPCST1_j = (D_j)(1/COP_j)P_j$$

$$SHOPCST2_j = (D_j)(1/COP_j)(P_j/P_e)$$

TABLE 10

Mean Values of Space Heat
Operating and Capital Costs (by alternative)

Alternative	SHOPCST	SHCPCST	SHOPCST1	SHCPCST1	SHOPCST2	SHCPCST2
1	583.2	882.2	0.00890	0.0179	1.096	2.481
2	134.3	1081.	0.00226	0.0218	0.3090	3.015
3	109.0	1724.	0.00169	0.0364	0.2388	4.999
4	536.1	874.7	0.00813	0.0163	1.000	2.256
5	124.4	2375.	0.00208	0.0461	0.2835	6.477
6	100.8	2839.	0.00156	0.0570	0.2191	7.965
7	656.9	1694.	0.01072	0.0328	1.328	4.521
8	206.4	1921.	0.00408	0.0397	0.5410	5.447
9	182.7	2294.	0.00352	0.0485	0.4707	6.638
10	401.2	4355.	0.00678	0.0780	0.8273	10.60

1	Elec. Forced Air / No Central Air	HVAC #13
2	Gas Forced Air / No Central Air	HVAC #1
3	Oil Forced Air / No Central Air	HVAC #7
4	Elec. Baseboard / No Central Air	HVAC #18
5	Gas Hot Water / No Central Air	HVAC #3
6	Oil Hot Water / No Central Air	HVAC #9
7	Elec. Forced Air / Central Air	HVAC #14
8	Gas Forced Air / Central Air	HVAC #2
9	Oil Forced Air / Central Air	HVAC #8
10	Electric Heat Pump	HVAC #15

Note that HVAC 18 has a coefficient of performance equal to one, has delivery factor one, and uses electricity so that the operating cost of HVAC 18 is (SHUECE*P_e).

The first normalization method replaces operating cost by an efficiency adjusted price, while the second method further scales all costs by the price of electricity. The efficiency adjusted price $SHOPCST1_j$ is related to the price of comfort since the latter is $SHOPCST1_j$ multiplied by the marginal increase in usage required to change the thermostat setting one degree. For a given household this quantity is constant across alternatives and would change all normalized operating costs in a proportional manner. Empirical results obtained using the calculated price of comfort rather than normalized operating costs were very similar yet more difficult to interpret for quick checks of the discount rate.

The normalized variables made sense on econometric grounds since the unobserved component of utility would tend to be otherwise heteroscedastic. Furthermore, the normalization seems valid on psychometric grounds since it is reasonable to assume that households view costs relative to the costs of some standard system.

Table 11 presents the results of estimating subsets of the ten alternative systems. The water heat inclusive value is not included in these specifications. Income, while included, has not been presented based on its insignificance across the various specifications. The results of the estimation are quite sensible both in terms of significance and sign. Furthermore, without extensive specification testing it is hard to detect any rejection of the independence of irrelevant alternatives assumption. Future work will explore departures from this assumption in the preferred specification using the methods of Hausman and McFadden (1981).

TABLE 11

Estimation of Space Heat Choice Model -
(Without Water Heat Inclusive Value) - Alternative Specifications[a]

	Alternative Label	Frequency	Percent of Cases	Frequency Chosen	Percent Chosen
Specifications 1 and 2:	1.000	591.0	100.0	21.00	3.553
	2.000	424.0	71.74	294.0	69.34
	3.000	591.0	100.0	99.00	16.75
	4.000	591.0	100.0	78.00	13.20
	5.000	424.0	71.74	57.00	13.44
	6.000	591.0	100.0	42.00	7.107
Specifications 3 and 4:	1.000	414.0	100.0	21.00	5.072
	2.000	334.0	80.68	294.0	88.02
	3.000	414.0	100.0	99.00	23.91
Specifications 5 and 6:	4.000	177.0	100.0	78.00	44.07
	5.000	90.00	50.85	57.00	63.33
	6.000	177.0	100.0	42.00	23.73
Specifications 7 and 8:	7.000	289.0	100.0	60.00	20.76
	8.000	223.0	77.16	186.0	83.41
	9.000	289.0	100.0	43.00	14.88
Specifications 9 and 10:	7.000	320.0	100.0	60.00	18.75
	8.000	231.0	72.19	186.0	80.52
	9.000	320.0	100.0	43.00	13.44
	10.00	320.0	100.0	31.00	9.688

[a]Total cases 911.

TABLE 11, cont.

	Alternatives 1	2	3	4	5	6
Variable[c]	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3	1 2 3	4 5 6	4 5 6
SHOPCST1	-700.9 (73.51)[b]	-	-934.2 (136.7)	-	-817.4 (129.3)	-
SHCPCST1	-24.47 (9.945)	-	-39.96 (19.19)	-	-36.35 (20.06)	-
SHOPCST2	-	-6.689 (1.036)	-	-8.771 (1.480)	-	-6.359 (2.546)
SHCPCST2	-	-0.3460 (0.0699)	-	-0.4113 (.1358)	-	-0.5922 (0.1295)
A1	2.341 (.827)	2.994 (1.181)	1.592 (0.827)	3.288 (1.352)	-	-
A2	2.795 (.533)	2.311 (.511)	1.953 (0.499)	1.944 (0.486)	-	-
A3	0.7230 (.443)	0.3882 (.437)	-	-	-	-
A4	3.675 (.731)	3.779 (1.052)	-	-	4.530 (1.332)	2.500 (2.428)
A5	1.141 (.514)	0.9629 (.514)	-	-	1.337 (.599)	0.9454 (0.621)
A7	-	-	-	-	-	-
A8	-	-	-	-	-	-
A9	-	-	-	-	-	-
Log Likelihood	-567.3	-579.3	-135.3	-137.3	-90.49	-98.34
Percent Correctly Predicted	65.14	64.97	88.65	88.89	78.53	76.27

[b]Standard errors in parenthesis.

[c]Coefficients of income interacted with reported alternative specific dummies not reported. All coefficients insignificant.

TABLE 11, cont.

	Alternatives 7	8	9	10
Variable[c]	7 8 9	7 8 9	7 8 9 10	7 8 9 10
SHOPCST1	-509.6 (89.81)[b]	-	-471.1 (76.38)	-
SHCPCST1	-34.86 (10.85)	-	-19.08 (6.138)	-
SHOPCST2	-	-8.863 (1.562)	-	-5.095 (.8705)
SHCPCST2	-	-.2325 (0.0773)	-	-.1068 (0.0371)
A1	-	-	-	-
A2	-	-	-	-
A3	-	-	-	-
A4	-	-	-	-
A5	-	-	-	-
A7	2.424 (.780)	6.578 (1.444)	1.251 (.654)	2.541 (.733)
A8	2.886 (.526)	3.252 (.5709)	1.473 (.586)	1.407 (.616)
A9	-	-	-1.654 (.592)	-1.809 (.605)
Log Likelihood	-141.9	-137.7	-228.5	-234.2
Percent Correctly Predicted	80.28	79.58	72.50	70.94

[b]Standard Errors in parenthesis.

[c]Coefficients of income interacted with reported alternative specific dummies not reported. All coefficients insignificant.

Estimation of discount factors appear robust across specifications. (For a discussion of the discount factor and its interpretation see Dubin and McFadden (1979)). We present the point estimates in Table 12. The discount rates, which range from 2.1 percent to 9.3 percent, may be interpreted as real rather than nominal factors which annualize capital costs. These values are quite low compared to estimates obtained by Dubin and McFadden (1979) and Hausman (1979).

Table 13 presents the results of estimating subsets of the HVAC alternatives where we have included the water heat choice inclusive value. The variable income is not included in this estimation. Point estimates of discount factors are given in Table 14. The general pattern for the inclusive value coefficient appears to be significant with the incorrect sign under the first normalization procedure and insignificant with the correct sign under the second normalization procedure. Given the small differential between the means of the inclusive value variable across fuel types, it is likely that there is significant interaction between the inclusive value variable and the alternative specific dummies. This is further confirmed by the fact that the model continues to robustly estimate the coefficients of operating and capital costs.

To further explore the interaction hypothesis we have estimated specifications 11, 12, 13, and 14 in Table 13. These models eliminate the alternative specific variable for the oil alternatives. The estimate of the inclusive value coefficient for water heat choice conditional on choosing a space heat system without air-conditioning varies from significance with the wrong sign to insignificance as before. However, the estimate of the coefficient conditional on choice of HVAC within the air-conditioning branch cannot be rejected from equaling one under either normalization procedure. There is no a priori reason to expect that the

TABLE 12

Discount Rates from Space Heat Choice Model without Inclusive Value for Water Heat Choice

Specification	Discount Factor (Percent)
1	3.49
2	5.17
3	4.28
4	4.69
5	4.45
6	9.31
7	6.84
8	2.62
9	4.05
10	2.10

TABLE 13

Space Heat Choice Models with Water Heat Inclusive Value

	Alternatives 1	2	3	4	5	6
Variables:	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3	1 2 3	4 5 6	4 5 6
SHOPCST1	-755.3 (83.29)	-	-1032 (147.1)	-	-910.6 (154.9)	-
SHCPCST1	-24.42 (9.615)	-	-35.0 (17.54)	-	-34.30 (18.57)	-
SHOPCST2	-	-6.432 (1.077)	-	-8.771 (1.514)	-	-5.802 (2.587)
SHCPCST2	-	-0.3261 (0.0666)	-	-0.3643 (0.1251)	-	-0.5795 (0.1263)
A1	4.463 (1.41)	2.617 (1.496)	7.732 (2.75)	4.973 (2.37)	-	-
A2	2.748 (0.341)	2.323 (0.319)	1.781 (.297)	1.993 (0.2822)	-	-
A3	0.5449 (0.241)	0.2645 (0.228)	-	-	-	-
A4	5.342 (1.374)	3.150 (1.40)	-	-	7.175 (2.450)	1.852 (2.852)
A5	1.495 (0.247)	1.366 (0.251)	-	-	2.054 (0.376)	1.403 (0.407)
A7	-	-	-	-	-	-
A8	-	-	-	-	-	-
A9	-	-	-	-	-	-
WHINCV	-1.212 (0.832)	0.4886 (0.6714)	-3.892 (1.852)	-0.6431 (1.371)	-1.674 (1.308)	0.4514 (0.991)
Log Likelihood	-568.0	-580.9	-133.7	-138.0	-90.44	-98.81
Percent Correctly Predicted	66.16	64.81	89.61	88.89	76.84	76.84

Table 13, cont.

	Alternatives 7	8	9	10	11	12
Variables:	7 8 9	7 8 9	7 8 9 10	7 8 9 10	1 2 3 4 5 6	1 2 3 4 5 6
SHOPCST1	-519.3 (94.51)	-	-441.4 (80.27)	-	-749.8 (82.98)	-
SHCPCST1	-27.68 (8.97)	-	-17.76 (5.636)	-	-39.62 (7.696)	-
SHOPCST2	-	-8.798 (1.642)	-	-4.475 (0.9035)	-	-6.189 (1.049)
SHCPCST2	-	-0.1715 (0.0605)	-	-0.1005 (0.0351)	-	-0.3723 (0.0557)
A1	-	-	-	-	3.697 (1.367)	2.140 (1.436)
A2	-	-	-	-	2.117 (0.204)	2.034 (0.200)
A3	-	-	-	-	-	-
A4	-	-	-	-	4.541 (1.325)	2.677 (1.341)
A5	-	-	-	-	1.169 (0.194)	1.186 (0.194)
A7	4.302 (1.580)	7.717 (2.03)	1.193 (0.448)	2.182 (0.575)	-	-
A8	2.457 (0.256)	2.692 (0.274)	1.476 (1.151)	2.284 (1.030)	-	-
A9	-	-	-1.028 (1.199)	-0.1847 (1.067)	-	-
WHINCV	-0.9738 (1.00)	-0.3788 (0.86)	0.4499 (0.86)	1.176 (0.74)	-1.186 (0.84)	0.4465 (0.673)
Log Likelihood	-143.8	-140.7	-229.9	-234.6	-570.6	-581.6
Percent Correctly Predicted	81.66	80.62	72.81	70.94	65.82	64.97

TABLE 13, cont.

	Alternatives	
	13	14
Variables:	7 8 9 10	7 8 9 10
SHOPCST1	-408.5 (70.82	-
SHCPCST1	-16.77 (5.467)	-
SHOPCST2	-	-4.395 (0.776)
SHCPCST2	-	-0.09885 (0.03376)
A1	-	-
A2	-	-
A3	-	-
A4	-	-
A5	-	-
A7	1.154 (0.446)	2.158 (0.557)
A8	2.453 (0.212)	2.459 (0.205)
A9	-	-
WHINCV	1.148 (0.294)	1.293 (0.306)
Log Likelihood	-230.3	-234.6
Percent Correctly Predicted	72.19	70.94

TABLE 14

Discount Rates for Space Heat Choice Model
with Inclusive Value of Water Heat Choice

Specification	Discount Factor
1	0.0323
2	0.0507
3	0.0339
4	0.0415
5	0.0377
6	0.0999
7	0.0533
8	0.0195
9	0.0402
10	0.0225
11	0.0528
12	0.0602
13	0.0411
14	0.0225

inclusive value coefficients should differ in the two branches so that any differences between the two estimates would be explicable only by differences in the degree of inter-correlations in each space heat choice cluster. The sequential estimation procedure cannot impose the constraint that the estimates of the inclusive value coefficients be equal.

We have adopted the strategy of not including the water heat choice inclusive value in the space heat choice estimation. We argue that the differences in the inclusive values are small and will be adequately captured in the alternative specific dummies. Further work will be required to determine the correct specification of water heat choice in the full nested logit model.

In Table 15 we present means of the space heat inclusive values constructed conditional on choice of central air-conditioning. Note that the difference in the size of the mean values corresponds to including central air-conditioning operating and capital costs in the space heat costs for the central air branch. This point will be taken up again in the next section.

VII. Central Air-Conditioning Choice

This section presents the results from estimation of the central air-conditioning choice model. From equation (11) of Section II, we see that the probability of air-conditioning choice depends on the inclusive value of room air-conditioning (when central air is not chosen), the inclusive values of space heat choice given air-conditioning choice, and on other attributes of the utility of purchasing an air-conditioning system. We follow the formulation of indirect utility discussed in Section IV on room air conditioner choice and use income and cooling degree days interacted with the first and second alternatives

TABLE 15

Means of Space Heat Inclusive Values
Conditioned on Central Air Choice

Variable	Mean
SHINCVNC (inclusive value given no central air-conditioning)	-0.6149
SHINCVC (inclusive value given central air-conditioning)	-2.389

(central vs. non-central) as determinants of the utility associated with either alternative. The inclusive value of room air-conditioning choice appears interacted with the second alternative as does the inclusive value of space heat choice given no central air-conditioning. The inclusive value of space heat choice given central air-conditioning is interacted with alternative one.

The results of the estimation are presented in Table 16. While real income and cooling degree days are significant and have the expected sign the coefficients of the inclusive value terms are all insignificant. To pursue the central air specification, we relax the assumption that the co-efficients of operating and capital costs are identical for both components of costs in the space heat given central branch of the tree structure. To do this, we remove the "pure" operating and capital costs for central air-conditioning from the total operating and capital space heating costs in alternatives 7, 8, 9, and 10. This cannot effect the space heat choice estimation as the operating and capital costs for central air-conditioning (excluding joint cost components) are constant across alternatives.

In terms of the indirect utility notation of Section II, we note that the utility of alternatives 7, 8, 9, and 10 may be written:

$$V_{s|c=1} = V'_{s|c=1} + V_{c=1} \qquad (1)$$

where:

$V_{s|c=1}$ = indirect utility of space heat choice s given central air-conditioning (c=1)

$V'_{s|c=1}$ = indirect utility of space heat choice s given central air-conditioning (c=1) which varies by alternative s

$V_{c=1}$ = indirect utility of central air-conditioning (c=1) (this does not vary with alternative s)

TABLE 16

Binary Logit Model of Central Air-Conditioning Choice -
Central Air-Conditioning Costs Included in
Space Heat Inclusive Value

	Alternative Label	Frequency	Percent of Cases	Frequency Chosen	Percent Chosen
Central AC	1.000	911.0	100.0	320.0	35.13
No Central AC	2.000	911.0	100.0	591.0	64.87

Variable Name	Logit Estimate	Standard Error	T-Statistic
SHINCVC	.1230	.8037E-01	1.531
SHINCVNC	-.7985E-01	.9647E-01	-.8278
RMINCV	.8684	1.085	.8003
RINCOME1	.9201E-01	.2456E-01	3.747
CDD2	-.1600E-02	.5936E-03	-2.696
$\Lambda 2$	3.767	.3840	9.811

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-451.0	-631.5
Percent Correctly Predicted	77.72	50.00

The operating and capital cost components of $V_{c=1}$ are respectively CACOPC and CACCST. The mean values of these variables are $73.77 and $888.30 respectively. When these costs are removed from the corresponding costs in $V_{s|c=1}$, the mean value of space heat inclusive value changes from -2.389 to -0.9980.

We present in Table 17 the re-estimated central air-conditioning choice model. In this specification we include the separate operating and capital costs CACOPC and CACCST interacted with the air conditioning choice alternative. Table 18 presents the estimated central air-conditioning choice model in which CACOPC and CACCST are normalized by predicted air-conditioning usage. It is interesting to note that the inclusive value coefficients given in Table 17 are consistent with the hypothesis of utility maximization (see McFadden (1981)) although the coefficients of CACOPC and CACCST are insignificant and of the wrong sign respectively.

The results in Table 18 using normalized operating and capital cost yield insignificant coefficients for two of the three inclusive values. This result may be due to spurious correlations among the variables in non-normalized and normalized forms. Future research will be needed to elicit the correct normalization rule. For the present we argue that a basic model may be used as a very good predictor of the choice of central air-conditioning and should perform adequately in the construction of instrumental variables used in the estimation of the utilization equations. The basic model (without air-conditioning operating and capital costs or inclusive values) is presented in Table 19.

TABLE 17

Binary Logit Model of Central Air-Conditioning Choice - Central Air-Conditioning Costs Without Normalization

Variable Name	Logit Estimate	Standard Error	T-Statistic
SHINCVC	.7453	.3059	2.437
SHINCVNC	.2819	.2126	1.326
RMINCV	2.588	1.006	2.574
RINCOME1	.1108	.2492E-01	4.447
CDD2	-.6440E-03	.5321E-03	-1.210
A2	4.158	.4215	9.866
CACOPC	-.2753E-02	.2984E-02	-.9227
CACCST	.8750E-03	.3673E-03	2.382

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-446.4	-631.5
Percent Correctly Predicted	78.38	50.00

TABLE 18

Binary Logit Model of Central Air-Conditioning Choice - Central Air-Conditioning Costs Normalized by Base Load Usage

Variable Name	Logit Estimate	Standard Error	T-Statistic
SHINCVC	-.1184	.3510	-.3373
SINCVNC	-.5294	.2621	-2.020
RMINCV	.8699	1.098	.7920
RINCOME1	.9005E-01	.2579E-01	3.491
CDD2	-.1708E-02	.5705E-03	-2.994
A2	2.051	.5446	3.765
CACOPC	-247.5	55.51	-4.459
CACCST	.1341	.1483	.9037

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-439.0	-631.5
Percent Correctly Predicted	78.81	50.00

TABLE 19

Binary Logit Model of Central Air-Conditioning Choice -
No Central Air-Conditioning Costs or Inclusive Values

Variable Name	Logit Estimate	Standard Error	T-Statistic
RINCOME1	.7869E-01	.1273E-01	6.181
CDD2	-.1632E-02	.1329E-03	-12.28
A2	3.477	.2550	13.64

Auxiliary Statistics	At Convergence	At Zero
Log Likelihood	-460.1	-631.5
Percent Correctly Predicted	77.39	50.00

VIII. The Effect of the ASHRAE 90-75 Building Standards on the Saturation of Alternative HVAC Systems

This section calculates the mean predicted probabilities of HVAC system choice under two alternative levels of building thermal characteristics. The first alternative is an uninsulated dwelling without storm windows or double glazing. The second alternative is the ASHRAE 90-75 voluntary thermal standard for new construction. Under this standard, all windows are stormed or double glazed, walls and ceiling are insulated, heating and cooling system capacities are reduced, and tight construction is used to reduce infiltration. The AHSRAE standards vary by region as discussed in McFadden and Dubin (1982).

For the purposes of calculating mean predicted probabilities we use the HVAC choice model illustrated in Figure 1. Coefficient estimates for the six alternative space heat choice model given no central air-conditioning and for the four alternative choice model given central air-conditioning are presented in Table 11.

Table 20 presents the mean predicted probabilities under the two alternative levels of building thermal characteristics as well as the probabilities for the observed level of building thermal integrity. The availability of natural gas is assumed to remain constant under each scenario. Predicted probabilities do not appear to shift significantly between the observed thermal integrity and the no insulation cases yet there is some predicted movement into oil systems from the electric systems.

Under the proposed ASHRAE standards there is a marked shift into electric baseboard and heat pump systems and away from other HVAC's.

The proposed ASHRAE standards would thus appear to increase the shares of energy efficient heating and cooling systems. These results should be viewed tentatively given that they include vintage as well as new construc-

TABLE 20

Mean Predicted Probabilities of HVAC System Choice Under Alternative Thermal Integrities

HVAC	NOBS	Base Case	No Insulation	ASHRAE 90-75
1	911	0.0368	0.03996	0.02238
2	655	0.7000	0.6586	0.6884
3	911	0.1598	0.1896	0.1184
4	911	0.1432	0.09514	0.2321
5	655	0.1283	0.1552	0.1153
6	911	0.0646	0.09007	0.04918
7	911	0.1737	0.1759	0.1522
8	655	0.7932	0.7881	0.7881
9	911	0.1390	0.1426	0.1288
10	911	0.1170	0.1148	0.1523

tion and do not take into account the costs of additional insulation. Future analysis will consider these effects and a broader scope of policy scenarios.

IX. Summary and Conclusions

This chapter has estimated a nested logit model of the choice of room air-conditioning, water heat fuel, space heat and central air-conditioning systems. The models estimated predict very well and may be used recursively to determine probabilities of alternative portfolios. It was found that the operating and capital cost components of utility in the room air-conditioning choice model were insignificant. Operating and capital costs were significant determinants in water heat fuel choice and space heat system choice after normalization for scale effects. Evidence remains inconclusive as to whether water heat choice given space heat choice is consistent with utility maximization, but evidence appears more conclusive that space heat choice given the choice of central air-conditioning is consistent with utility maximization. Estimates of discount rates are determined to be much larger for water heat choice given the choice of space heat system than for space heat system choice given the decision to install central air-conditioning. The latter discount rates are about an order of magnitude smaller than estimates given in Dubin and McFadden (1979).

Finally, we have used the space heat choice model to calculate changes in the predicted shares of HVAC systems which would result under the proposed ASHRAE 90-75 standards.

Footnotes

1. These preliminary investigations are essentially data analysis done in the absence of a good classical procedure for selecting the correct tree structure. Standard errors should be interpreted with this process in mind.

CHAPTER IV

Estimation of the Demand for Electricity and Natural Gas Using the NIECS Billing Data

The purpose of this chapter is to estimate the demand for electricity and natural gas using the NIECS monthly billing data. A sample of 911 households is selected to correspond to the HVAC nested logit model of Chapter III so that simultaneity between appliance choice and usage may be explored. Estimation utilizes the econometric forms suggested in Chapter I for joint continuous-discrete systems. A complete discussion of the NIECS billing data is given in Appendix I.

Section II presents the electricity demand model estimated by ordinary least squares. Section III considers the natural gas demand estimation. Consistent estimation procedures applied to both demand equations are presented in Section IV.

II. Demand for Electricity by Aggregated Billing Period

In this section we estimate the demand for electricity conditioned on appliance holdings using the monthly billing data from NIECS. A discussion of the procedures used to process the billing data is given in Appendix I. The form of the estimated equation is given by:

$$X_t^e - QEBASE = \sum_j^J UEC_{jt}\delta_{jt}(\alpha_j^1 + P_{jt}\alpha_j^2 + Y\alpha_j^3) + \varepsilon_t^e \quad (1)$$

where:

X_t^e = demand for electricity in period t

QEBASE = base usage of electricity in the presence of electric refrigerators, ovens, ranges, microwave ovens, freezers, washers, and clothes dryers

UEC_{jt} = predicted usage of appliance j in period t

δ_{jt} = indicator of appliance j in period t

Y = income

ε_t^e = error term for electricity equation

α_j^1, α_j^2, α_j^3 = parameters

J = number of appliance portfolios

The decomposition of residual (total-base) usage into component demands has been discussed in Chapter I. The procedure has also been applied in the works of Dubin and McFadden (1979), Goett, McFadden, and Earl (1980) and Parti and Parti (1980). For the purposes of our study we limit attention to the usage of electricity by space heating, air-conditioning, room air-conditioning, and water heating. This selection of appliances corresponds to the choice model of Chapter III and should account for the greater sources of electricity demand in residences.

Table 1 presents the mean values of the variables UEC_{jt} and P_{jt} where j includes the HVAC systems 2, 8, 13, 14, 15, 18, room air-conditioning, and water heating. When an HVAC system includes both heating and air-conditioning we distinguish the predicted unit energy consumptions by the letters S and A. Thus, UEC14S and UEC14A denote the predicted usage of HVAC system 14 for space heating and air-conditioning respectively. The UEC determination across appliances utilizes the predicted thermal variable SHUEC with adjustments for delivery system, efficiency, and the length of the billing period. Further details may be found in Appendix I.

The variable P_{jt} (denoted by P2A, P8A, etc.) represents the service price for appliance j in period t. We have used the predicted thermal variable DSHUEC which measures the marginal increase in usage resulting from a one degree change in the thermostat, and the price of electricity to calculate the marginal service price. Further details concerning the construction of unit energy consumptions and service prices may be found in Appendix I. The time index t refers to the three temperature aggregated billing periods: Winter, Off-Season, and Summer. The marginal price of electricity is allowed to vary seasonally as discussed in Appendix I.

Table 2 presents the definitions of variables used in the electricity demand model. The product of UEC_{jt} and α_{jt} is denoted by the neumonic SU followed by an HVAC system number. Thus, SU18 is the product of a dummy variable for HVAC system 18 and UEC18. Table 3 gives the mean values for these variables by aggregated billing period.

TABLE 1

Mean Values of UEC's and Service Prices by Time Period

Variable	Winter	Off-Season	Summer	Units
P18	38.64	24.89	3.844	$/1°
P13	41.99	27.32	4.206	$/1°
P14S	41.99	27.32	4.206	$/1°
P15S	24.61	13.60	2.279	$/1°
P14A	.2402E-01	2.033	5.456	$/1°
P15A	.2402E-01	2.033	5.456	$/1°
P2A	.2402E-01	2.033	5.456	$/1°
P8A	.2402E-01	2.033	5.456	$/1°
PRMAC	.2402E-01	2.033	5.456	$/1°
PWH	.1128E-01	.1116E-01	.1186E-01	$/gallon
UEC18	.2948E-05	8538.	391.2	KWH
UEC13	.3203E+05	9365.	428.0	KWH
UEC14S	.3203E+05	9365.	428.0	KWH
UEC15S	.1868E+05	4473.	234.9	KWH
UEC15A	3.427	466.8	1953.	KWH
UEC14A	3.427	466.8	1953.	KWH
UEC8A	3.427	466.8	1953.	KWH
UEC2A	3.427	466.8	1953.	KWH
UECRMAC	3.427	466.8	1953.	KWH
UECWH	2059.	1655.	1492.	KWH
NOBS	777	845	802	

WH - Electric water heating
RMAC - Room air-conditioning

TABLE 2

Variable Definitions

Variable	Description
SU18	(HVAC 18 dummy)(UEC18)
SU13	(HVAC 13 dummy)(UEC13)
SU14S	(HVAC 14 dummy)(UEC14S)
SU15S	(HVAC 15 dummy)(UEC15S)
SU14A	(HVAC 14 dummy)(UEC14A)
SU15A	(HVAC 15 dummy)(UEC15A)
SUWHE	(Water heat electric dummy)(UECWH)
SURMAC	(Room air conditioner dummy)(UECRMAC)

SU18P, SU13P, SU14SP, SU15SP, SU14AP, SU15AP, SUWHEP, and SURMACP are variables multiplied by service prices.

SU18Y, SU13Y, SU14SY, SU15SY, SU14AY, SU15AY, SUWHEY, and SURMACY are variables multiplied by income.

MPE	Marginal price of electricity ($/KWH)
EDAYS	Number of days in aggregated period
NHSLDMEM	Number of household members
NETEQUAN	Net electricity usage (KWH)

TABLE 3

Mean Values of Variables Appearing in Electricity Demand Model

Variable	Winter	Off-Season	Summer
SU18	1962.	770.2	29.56
SU13	518.0	374.2	6.656
SU14S	964.3	606.4	32.46
SU15S	521.0	162.4	7.404
SU18P	.6659E+05	.1818E+05	110.8
SU13P	.1375E+05	8551.	49.58
SU14SP	.3257E+05	.1976E+05	218.8
SU15SP	.1972E+05	2947.	22.90
SU18Y	.4146E+05	.1662E+05	560.3
SU13Y	.1099E+05	9004.	162.5
SU14SY	.3517E+05	.1792E+05	917.3
SU15SY	.1361E+05	4713.	184.6
SU14A	.2227	40.15	301.1
SU15A	.1600	19.79	100.3
SU2A	.7845	109.4	521.0
SU8A	.3657	22.83	116.1
SU14AP	.2310E-01	117.8	3850.
SU15AP	.9249E-01	65.84	1149.
SU2AP	.1633	514.0	4042.
SU8AP	.2571	85.21	1090.
SU14AY	5.583	1119.	7652.
SU15AY	3.836	622.5	2644.
SU2AY	19.01	3033.	.1326E+05
SU8AY	11.89	724.7	3576.
SUWHE	654.8	644.2	604.3
SUWHEP	6.611	5.763	6.478
SUWHEY	.1589E+05	.1530E+05	.1420E+05
INCOME	22.97	23.00	22.88
MPE	.3946E-01	.3904E-01	.4149E-01
EDAYS	182.4	145.8	134.1
NHSLDMEM	3.264	3.243	3.287
NETEQUAN	4663.	3034.	4275.
NOBS	777	845	802

We have selected the sample to correspond to the 911 households represented in the discrete choice models of Chapter III. Given three billing periods per household, we would have 2733 potential observations. From Table 1 we note that 2424 (=777 + 845 + 802) of the 2733 had available electricity billing data.

The dependent variable for equation (1) is denoted, NETEQUAN, and is the difference in total usage EQUAN and base usage for excluded appliances QEBASE. The construction of QEBASE uses UEC values (in KWH/day units) for electric refrigerators, ovens, ranges, microwave ovens, freezers, washers, and clothes dryers. These UEC values are combined with ownership dummies and then multiplied by the number of days in the billing period EDAYS. The UEC values were obtained from Cambridge Systematics/West (1981). The results of least squares regression of the electricity demand model are given in Tables 4 and 5. Note that in the winter period we have excluded variables related to air-conditioning. The least squares estimates for the summer period did not produce sensible results and are omitted. A pattern of summer consumption dependence on cooling systems would have been expected. That this was not the case suggests the need for further analysis into the precise nature of billing cycle variations.

The instrumental variable estimation is facilated computationally when we adopt a restricted form of equation (1) in which the coefficients of the variables interacted with price are constrained to be equal. We similarly restrict the coefficients of UEC and UEC interacted with income.

Table 6 presents the means and definitions of the constrained variables. Note that coefficients are permitted to differ between heating and cooling systems. The constrained demand models are presented in Tables 7 and 8 for

TABLE 4

Electricity Demand Model Estimated by Ordinary
Least Squares: Winter Period

Dependent Variable NETEQUAN

Variable Name	Estimated Coefficient	T-Statistic
ONE	-468.4	-.7346
MPE	-.1838E+05	-1.149
EDAYS	8.889	5.240
NHSLDMEM	444.8	5.758
SU18	.8788	16.00
SU18P	-.3589E-02	-6.483
SU18Y	-.1127E-01	-6.391
SU13	.5544	7.332
SU13P	.3464E-02	2.795
SU13Y	-.7797E-02	-2.417
SU14S	.4339	5.438
SU14SP	-.9289E-02	-7.811
SU14SY	.6361E-02	4.656
SU15S	1.010	12.69
SU15SP	-.1249E-01	-7.939
SU15SY	-.1034E-03	-.2770E-01
SUWHE	1.113	1.703
SUWHEP	-64.68	-1.185
SUWHEY	.3764E-01	3.691

R-Squared = .7930
Number of Observations = 777
Sum of Squared Residuals = .7127E+10
Standard Error of the Regression = 3066.

TABLE 5

Electricity Demand Model Estimated by Ordinary Least Squares: Off-Season

Dependent Variable NETEQUAN

Variable Name	Estimated Coefficient	T-Statistic
ONE	-895.1	-1.722
MPE	4619.	.4308
EDAYS	4.568	4.514
NHSLDMEM	240.7	4.710
SU18	.5102	7.978
SU18P	-.1009E-01	-4.072
SU18Y	.1334E-01	4.143
SU13	.6330	5.372
SU13P	-.1904E-01	-4.517
SU13Y	.1497E-01	4.370
SU14S	.6056	4.196
SU14SP	-.4928E-02	-1.497
SU14SY	.5804E-02	1.069
SU15S	.5119	.9158
SU15SP	-3666E-01	-2.370
SU15SY	.4825E-01	2.712
SU14A	-6.094	-2.844
SU14AP	1.066	2.294
SU14AY	.6123E-01	.7387
SU15A	-2.696	-.5397
SU15AP	.9560	1.370
SU15AY	-.3755E-01	-.2907
SURMAC	.8205	.9575
SURMACP	-.5073E-01	-.4783
SURMACY	.7381E-03	.2927E-01
SU8A	-2.643	-1.421
SU8AP	.6076	2.827
SU8AY	.4970E-01	1.296
SU2A	-.3693	-.6714
SU2AP	-.9563E-01	-2.616
SU2AY	.7291E-01	4.300
SUWHE	2.692	8.177
SUWHEP	-101.1	-4.268
SUWHEY	-.1512E-01	-1.650

R-Squared	=	.8483
Number of Observations	=	845
Sum of Squared Residuals	=	.3755E+10
Standard Error of the Regression	=	2152.

TABLE 6

Mean Values for Variables in Constrained Demand Model

SUSHE = SU18 + SU13 + SU14S + SU15S
SUSHEP = SU18P + SU13P + SU14SP + SU15SP
SUSHEY = SU18Y + SU13Y + SU14SY + SU15SY

SUCAC = SU14A + SU15A + SU2A + SU8A
SUCACP = SU14AP + SU15AP + SU2AP + SU8AP
SUCACY = SU14AY + SU15AY + SU2AY + SU8AY

Variable	Winter	Off-Season	Summer
SUSHE	3965.	1913.	76.09
SUSHEP	.1326E+06	.4943E+05	402.1
SUSHEY	.1012E+06	.4825E+05	1825.
SUCAC	1.533	192.2	1038.
SUCACP	.5360	782.8	.1013E+05
SUCACY	40.32	5499.	.2714E+05
NOBS	777	845	802

TABLE 7

Ordinary Least Squares Regression of Constrained Electricity Demand Model: Winter

Dependent Variable is NETEQUAN

Variable Name	Estimated Coefficient	T-Statistic
SUWHE	1.345	1.961
SUWHEP	1.049	.1791E-01
SUSHE	.7317	18.54
SUSHEP	-.4439E-02	-8.775
SUSHEY	-.3808E-02	-4.663
ONE	-1391.	-1.930
INCOME	44.39	3.981
MPE	-.1820E+05	-1.015
EDAYS	9.096	4.817
NHSLDMEM	392.0	4.450

R-Squared	=	.7351
Number of Observations	=	777
Sum of Squared Residuals	=	.9121E+10
Standard Error of the Regression	=	3449.

TABLE 8

Ordinary Least Squared Regression of Constrained Electricity Demand Model: Off-Season

Dependent Variable is NETEQUAN

Variable Name	Estimated Coefficient	T-Statistic
SUWHE	2.855	11.83
SUWHEP	-140.8	-6.233
SUCAC	-.4496	-.8933
SUCACP	-.1394	-4.325
SUCACY	.9954E-01	6.648
SURMAC	.8073	2.437
SUSHE	.4559	9.374
SUSHEP	-.6016E-02	-4.679
SUSHEY	.8308E-02	5.673
ONE	-1077.	-2.025
INCOME	-14.14	-1.778
MPE	.1113E+05	1.002
EDAYS	4.828	4.582
NHSLDMEM	264.6	4.870

R-Squared	=	.8241
Number of Observations	=	845
Sum of Squared Residuals	=	.4353E+10
Standard Error of the Regression	=	2289.

the winter and off-season periods respectively.

We have excluded the variable SUWHEY from the constrained model in Table 7 due to its high colinearity with other income variables. The constrained model in Table 8 further excludes the price and income variables combined with SURMAC. These excluded variables were not significant under any of the test specifications. The coefficients in Table 7 are reasonably well determined and of the expected sign with the exception of the space heat income term. The estimates in Table 8 do confirm negative price and positive income effects both for heating and air-conditioning.

We conclude this section with the calculation of price and income elasticities conditional on the choice of HVAC system. The elasticities are evaluated at the mean values of variables by billing period and presented in Table 9.

III. Demand for Natural Gas by Aggregated Billing Period

This section presents the estimation of the demand for natural gas using the NIECS aggregated billing data. We follow the general procedures of Section II and attempt a decomposition of residual natural gas usage into component appliance demands.

Mean values of unit energy consumptions are given in Table 10 for HVAC systems 1, 2, and 3 and gas water heating. Table 10 further includes the corresponding service prices and their mean values. The choice of systems again corresponds to the nested logit model of Chapter III and the resulting sample of 1380 (= 459 + 476 + 445) observations corresponds to available billing data on 655 households for which gas was available.

The dependent variable for the natural gas demand equation is denoted NETGQUAN and is the difference between total usage GQUAN and base usage

TABLE 9

Income and Price Elasticities Conditional on
HVAC System Choice For Constrained Electricity Demand Model

Partial Elasticity of Net Usage with respect to:	Winter	Off-Season
MPE	-0.154*	+0.143*
INCOME	+0.219	-0.107*
PWH	+0.005*	-0.891
Space Heat Service Price		
SYSTEM 18	-1.084	-0.421
SYSTEM 13	-1.280	-0.507
SYSTEM 14	-1.280	-0.507
SYSTEM 15	-0.438	-0.121
Space Heat Income Effect		
SYSTEM 18	-0.553	+0.538
SYSTEM 13	-0.601	+0.590
SYSTEM 14	-0.601	+0.590
SYSTEM 15	-0.350	+0.282
Central Air-Conditioning Service Price	-	-0.044
Central Air-Conditioning Income Effect	-	+0.352

*Coefficient not significantly different from zero at 5% level.

QGBASE. QGBASE was calculated in an analogous manner to the elctricity variable QEBASE and includes the base usage of clothes drying, ovens, and ranges. Unit energy consumptions (measured in therms/day) were obtained from Werth (1978).

Tables 11 and 12 give the definitions and means of the variables used in the natural gas demand model. The results of the least squares regression of the gas demand model are given in Tables 13 and 14. We ignore the residual demand for natural gas in the summer period.

We follow the approach of Section II and consider a constrained version of the gas demand model for which price, income, and UEC variable coefficients are assumed equal across the three HVAC systems. The mean values and definitions of the constrained variables are given in Table 15. Least squares estimation of the constrained gas demand model is presented in Tables 16 and 17 for the winter and off season periods. Note that we have excluded the income effect for water heating and allow its effect to be captured in an independent income term. The price and income elasticities for the constrained model are given in Table 18.

IV. Consistent Estimation of the Demand for Electricity and Natural Gas

Econometric studies of unit energy consumptions have assumed, implicitly or explicitly, statistical independence of appliance choice and the additive equation error and have proceeded with least squares estimation. In practice some correlation of unobserved variables is likely. For an appliance such as a water heater, unobserved factors which increase intensity of use (e.g. tastes for hot water clothes washing) are likely to decrease the probability of choosing the operating to capital cost intensive electric system. Least squares estimation of the UEC equation induces a classical bias due to

TABLE 10

Mean Values of UEC's and Service Prices by Time Period

Variables	Winter	Off-Season	Summer	Units
P1	8.290	5.913	.8150	$/1°
P2	8.290	5.913	.8150	$/1°
P3	7.690	5.424	.7516	$/1°
PWH	.3142E-02	.3121E-02	.3114E-02	$/gallon
UEC1	1167.	347.6	16.15	Therms
UEC2	1167.	347.6	16.15	Therms
UEC3	1083.	319.5	14.93	Therms
UECWH	104.6	84.35	68.68	Therms

TABLE 11

Variable Definitions

Variable	Description
SU1	(HVAC 1 dummy)(UEC1)
SU2	(HVAC 2 dummy)(UEC2)
SU3	(HVAC 3 dummy)(UEC3)
SUWHG	(Water heat gas dummy)(UECWH)

SU1P, SU2P, SU3P, and SUWHGP are variables multiplied by service prices.

SU1Y, SU2Y, SU3Y, and SUWHGY are variables multiplied by income.

MPG	Marginal price of natural gas ($/Therms
GDAYS	Number of days in aggregated period
NHSLDMEM	Number of household members
NETGQUAN	Net natural gas usage (Therms)

TABLE 12

Mean Values of Variables Appearing in
Natural Gas Demand Model

Variable	Winter	Off-Season	Summer
SU1	568.6	208.4	8.474
SU1P	5738.	2686.	12.09
SU1Y	.1272E+05	5227.	188.8
SU2	422.5	108.1	5.519
SU2P	4565.	1159.	7.229
SU2Y	.1199E+05	3010.	149.7
SU3	130.1	21.84	1.480
SU3P	1758.	134.5	1.920
SU3Y	3555.	545.6	42.31
SUWHG	97.56	79.81	63.92
SUWHGP	.3029	.2445	.1966
SUWHGY	2368.	2029.	1500.
MPG	.2284	.2268	.2263
GDAYS	193.6	154.0	127.1
NHSLDMEM	3.218	3.214	3.254
NETGQUAN	1437.	501.9	126.0
INCOME	22.95	23.22	22.87
NOBS	459	476	445

TABLE 13

Natural Gas Demand Model Estimated by Ordinary Least Squares: Winter

Dependent Variable NETGQUAN

Variable Name	Estimated Coefficient	T-Statistic
SU1	.7109	6.543
SU1P	-.2422E-01	-7.208
SU1Y	.5919E-02	1.662
SU2	.8339	7.791
SU2P	-.2913E-01	-6.123
SU2Y	.7749E-02	2.316
SU3	.3041	2.406
SU3P	.1106E-02	.1574
SU3Y	.1873E-01	3.394
SUWHG	-5.283	-3.012
SUWHGP	1614.	3.589
SUWHGY	-.1241E-01	-.3071
ONE	366.0	2.158
MPG	-2378.	-3.347
GDAYS	4.122	9.874
NHSLDMEM	34.51	2.390

R-Squared	=	.7514
Number of Observations	=	459
Sum of Squared Residuals	=	.6962E+08
Standard Error of the Regression	=	396.4

TABLE 14

Natural Gas Demand Model Estimated by Ordinary Least Squares: Off-Season

Variable Name	Estimated Coefficient	T-Statistics
SU1	.5002	3.317
SU1P	-.1607E-01	-3.913
SU1Y	.1077E-01	2.210
SU2	.1434	1.079
SU2P	.4070E-03	.8390E-01
SU2Y	.2499E-01	6.072
SU3	.2669E-01	.5846E-01
SU3P	.4687E-02	.1210
SU3Y	.3522E-01	2.478
SUWHG	.8981	.6459
SUWHGP	-41.15	-.9945E-01
SUWHGY	-.5378E-01	-2.400
ONE	31.68	.3490
MPG	-744.8	-2.031
GDAYS	2.437	7.576
NHSLDMEM	25.62	3.082

R-Squared	=	.7424
Number of Observations	=	476
Sum of Squared Residuals	=	.2722E+08
Standard Error of the Regression	=	243.3

TABLE 15

Mean Values for Variables in Constrained Demand Model

SUSHG = SU1 +SU2 + SU3
SUSHGP = SU1P + SU2P + SU3P
SUSHGY = SU1Y + SU2Y + SU3Y

SUSHG	1121.	338.4	15.47
SUSHGP	.1206E+05	3979.	21.24
SUSHGY	.2826E+05	8782.	380.7

TABLE 16

Ordinary Least Squares Regression of Constrained
Natural Gas Demand Model: Winter

Dependent Variable is NETGQUAN

Variable Name	Estimated Coefficient	T-Statistic
SUSHG	.4309	4.523
SUSHGP	-.1737E-01	-5.353
SUSHGY	.1581E-01	6.585
SUWHG	-6.711	-3.726
SUWHGP	2079.	4.200
ONE	809.8	4.066
MPG	-3123.	-3.973
INCOME	-10.52	-2.908
GDAYS	4.022	8.753
NHSLDMEM	33.52	2.054

R-Squared	=	.6815
Number of Observations	=	459
Sum of Squared Residuals	=	.8917E+08
Standard Error of the Regression	=	445.7

TABLE 17

Ordinary Least Squares Regression of Constrained
Natural Gas Demand Model: Off-Season

Dependent Variable is NETGQUAN

Variable Name	Estimated Coefficient	T-Statistic
SUSHG	.5541	5.039
SUSHGP	-.9647E-02	-2.626
SUSHGY	.1399E-01	5.141
SUWHG	-2.757	-2.019
SUWHGP	640.0	1.509
ONE	249.0	2.495
MPG	-1042.	-2.694
INCOME	-3.645	-2.500
GDAYS	1.818	5.430
NHSLDMEM	26.56	2.978

R-Squared	=	.7029
Number of Observations	=	476
Sum of Squared Residuals	=	.3139E+08
Standard Error of the Regression	=	259.5

TABLE 18

Price and Income Elasticities Conditional on HVAC
System Choice for Constrained Natural Gas Demand Model

Partial Elasticity of Net Usage With Respect to:	Winter	Off-Season
MPG	-0.496	-0.471
INCOME	-0.168	-0.169
PWH	+0.475	+0.338*
Space Heat Service Price		
SYSTEM 1	-0.117	-0.040
SYSTEM 2	-0.117	-0.040
SYSTEM 3	-0.101	-0.033
Space Heat Income Effect		
SYSTEM 1	+0.295	+0.275
SYSTEM 2	+0.295	+0.275
SYSTEM 3	+0.273	+0.207

*Coefficient not significant from zero at 5% level.

correlation of an explanatory variable and the equation disturbance.

Dubin and McFadden (1979) consider several alternative consistent procedures for estimation of the parameters of the UEC equation. In Appendix II these methods are outlined and an argument is made for the asymptotic efficiency and simplicity of a simple instrumental variable method. The IV method uses consistent estimates of choice probabilities (interacted with the explanatory variables) as instruments. The consistency of this procedure has been noted by McFadden, Kirschner, and Puig (1977) and by Heckman (1979). Using the choice probabilities as instruments yields an estimator distinct from two-stage least squares in which choice dummies are replaced by consistent estimates of their expected values. This latter method is termed a reduced form estimator and is discussed in Appendix II.

We have estimated the constrained electricity and natural gas demand models of Sections II and III by instrumental variables. The estimated choice probabilities are obtained from the nested logit model of Chapter III and care must be taken to calculate unconditional probabilities using the appropriate form of Bayes' Rule. The probability of choosing electric water heat, for example, is the sum of the conditional probabilities of choosing electric water heat given space heat fuel multiplied by the unconditional probability of each fuel type.

Attempts to estimate the unconstrained demand models by instrumental variable methods were unsuccessful given the number of endogenous right hand side variables and the effective inter-correlations among the calculated instruments. We thus follow the simpler procedure of estimating the constrained models and allow the instrument list to include variables in Tables 2 and 11 for which choice dummies are replaced by consistent estimates of their expectations. Tables 19, 20, 21, and 22 present the IV estimates of the constrained models by fuel and aggregated billing period. The parameter estimates are qualitatively similar to their least squares counterparts. To formally test for significant differences in the estimated parameters we have employed a test due to Hausman (1978). The test requires that each suspected endogenous variable be regressed against the instrument list. Fitted values of these variables are then included in the model as additional explanatory variables. A test of the joint significance of the included fitted values is then equivalent to a specification test of correlation between the structural explanatory variables and the equation error. The models in Tables 19, 20, 21, and 22 estimated by instrumental variables in comparison with their least squares analogues in Tables 7, 8, 16, and 17 yield chi-squared statistics of 3.08, 8.17, 1.84, and 3.44 with degrees of freedom of 6, 6, 5, and 5 respectively. Under standard levels of significance we cannot reject the hypothesis of independence between the choice dummies and the unobserved UEC equation errors. This result must be viewed as provisional and highly dependent on the structure of the constrained demand model. Further study will explore the underlying UEC specifications and attempt alternative consistent estimation methods.

V. Summary and Conclusions

This chapter reports estimates of the demand for electricity and natural gas using the NIECS monthly billing data. The procedure attempted to decompose the energy consumption for each household into component demands attributable to type of HVAC system, water heating, and room air-conditioning. The sample of households was selected to correspond to the discrete choice modeling of Chapter III. In this way, we were able to consider simultaneity in the

TABLE 19

Instrumental Variable Estimation of Constrained
Electricity Demand Model: Winter

Dependent Variable NETEQUAN

Variable Name	Estimated Coefficient	T-Statistic
SUWHE	1.173	.8191
SUWHEP	23.77	.2040
SUSHE	.7206	6.380
SUSHEP	-.2403E-02	-1.480
SUSHEY	-.5031E-02	-2.643
ONE	-1393.	-1.654
INCOME	46.10	3.209
MPE	-.1230E+05	-.5627
EDAYS	7.006	2.971
NSHLDMEM	383.1	4.091

R-Squared	=	.7190
Number of Observations	=	777
Sum of Squared Residuals	=	.9675E+10
Standard Error of the Regression	=	3552.

TABLE 20

Instrumental Variable Estimation of Constrained Electricity Demand Model: Off-Season

Dependent Variable is NETEQUAN

Variable Name	Estimated Coefficient	T-Statistic
SUWHE	3.649	9.431
SUWHEP	-176.5	-4.914
SUCAC	-.7168E-02	-.8546E-02
SUCACP	-.1638	-3.635
SUCACY	.9554E-01	4.022
SURMAC	.9903	1.089
SUSHE	.1941	2.256
SUSHEP	-.5005E-02	-2.234
SUSHEY	.1272E-01	5.192
ONE	-902.3	-1.554
INCOME	-22.81	-2.635
MPE	.1037E+05	.8563
EDAYS	4.510	3.949
NHSLDMEM	255.2	4.454

R-Squared	=	.8147
Number of Observations	=	845
Sum of Squared Residuals	=	.4586E+10
Standard Error of the Regression	=	2349.

TABLE 21

Instrumental Variable Estimation of Constrained
Natural Gas Demand Model: Winter

Dependent Variable is NETGQUAN

Variable Name	Estimated Coefficient	T-Statistic
SUSHG	.2709	2.411
SUSHGP	-.1386E-01	-3.651
SUSHGY	.1636E-01	6.106
SUWHG	-2.834	-1.066
SUWHGP	1557.	1.886
ONE	777.9	2.731
MPG	-2519.	-2.176
INCOME	-11.42	-2.854
GDAYS	3.409	4.579
NHSLDMEM	13.22	.6292

R-Squared	=	.6728
Number of Observations	=	459
Sum of Squared Residuals	=	.9163E+08
Standard Error of the Regression	=	451.8

TABLE 22

Instrumental Variable Estimation of Constrained
Natural Gas Demand Model: Off-Season

Dependent Variable is NETGQUAN

Variable Name	Estimated Coefficient	T-Statistic
SUSHG	.6018	4.992
SUSHGP	-.1084E-01	-2.791
SUSHGY	.1371E-01	4.792
SUWHG	-1.060	-.6208
SUWHGP	73.88	.1467
ONE	145.3	1.277
MPG	-579.2	-1.335
INCOME	-3.524	-2.378
GDAYS	1.766	3.050
NHSLDMEM	25.96	2.359

R-Squared	=	.7013
Number of Observations	=	476
Sum of Squared Residuals	=	.3156E-08
Standard Error of the Regression	=	260.2

demand system and test the hypothesis that unobserved characteristics which affect the choice of HVAC system are related to unobserved characteristics influencing the demand for energy given system choice. The large number of potentially endogenous explanatory variables reduced the effectiveness of the instrumental variable method used to achieve consistent parameter estimates. We thus adopted a strategy of estimating a constrained demand system and tested for simultaneity using the methods of Hausman (1978). Preliminary evidence does not detect endogeneity of appliance holdings in the constrained system. Further research will explore the system specifications and apply general simultaneous equation methods.

Appendix I. A Review of the Appended NIECS Data Base and the Monthly Billing Data

This appendix reviews the National Interim Energy Consumption Survey (NIECS) data bank developed at the Massachusetts Institute of Technology during the summer and fall of 1981 by Thomas C. Cowing, Jeffrey A. Dubin, and Daniel L. McFadden. Although the NIECS data contain a great deal of detailed information on the residential energy demand characteristics of individual households, it does not contain all of the information required to model household appliance choice and utilization. Substantial amounts of additional data are required, much of it in the form of thermal performance and price information.

A significant determinant of appliance choice, for example, will be related to the capital cost (appliance cost plus installation costs) and expected operating cost of alternative appliance portfolios facing the household at the time the decision is made. Consider in turn the components of expected operating costs and capital costs for alternaivee heating-ventilation-air-conditioning (HVAC) systems. Expected operating costs are related to energy utilization which varies seasonally and with the thermal integrity of the housing shell. Energy utilization may be predicted using a thermal network model of the home but requires detailed information on daily temperature distribution, amount and placement of insulation, etc. Expected operating costs are further related to the various coefficients of performance in each HVAC system and to expectations of the course of energy prices. The use of expected fuel prices in a life-cycle intertemporal utility maximization model requires an extensive time-series of data (e.g. by fuel type and state) since expectations are presumably based in large part on past prices.

Capital costs for alternative HVAC systems are related to capacity load requirements which may be calculated using a thermal model under design conditions. For heating systems this requires knowledge of winter design temperatures. For cooling systems it is necessary to collect the summer daily temperature range as well as the summer design temperature. In addition, capital costs given capacity are expected to vary cross-sectionally given the variability of the labor component of installation costs in a national cluster sample. Finally the determination of fuel utilization conditional on choice of HVAC system reqquires explicit construction of HVAC service prices. This calculation requires that <u>marginal</u> prices be determined which correspond to the period in which energy consumption is observed.

The purpose of this appendix is to detail the components of the NIECS data base in its appended form. Section one outlines the documentation and evaluation of the NIECS data base given principally in a series of technical reports by Cowing, Dubin, and McFadden. We go on further to describe the source and description of additonal raw variables matched to each NIECS household. Section two considers the NIECS billing data and reviews procedures used to reprocess the data in a form suitable for econometric research. Section three examines the use of the monthly billing data in the construction marginal prices and section four considers a case study of a particular NIECS household as an illustration of the data structure and as a detailed internal consistancy check. A final section includes several fortran programs described in the text with associated output.

I. The Appended NIECS Data Base

1. Review of Documentation of the NIECS Data Base

The National Interim Energy Consumption Survey (NIECS) contains detailed energy demand information at the household level of 4081 households over the period April 1978 to March 1979. Among the data included are information on the structural characteristics of the housing unit, demographic characteristics of the household, fuel usage, appliance characteristics and actual energy consumption over the 12-month period. The NIECS annual file coded 59 separate variables to report these items. In Table 1 we provide a list of the NIECS information in summary form.

The preparation of a data bank to organize and classify a subset of the NIECS annual file was undertaken by Tom Cowing, Jeff Dubin, and Dan McFadden in the Summer of 1981. At this point an evaluation of the data set was made to determine its usefulness for a demand for energy study. For substantive details concerning this evaluation the reader may consult Cowing, Dubin, and McFadden, "Residential Energy Demand Modeling and the NIECS Data Base: An Evaluation" (1982). In their report, Cowing, Dubin, and McFadden review the NIECS data and consider an assessment of measurement error, sample design, imputation, and other data problems. Related source documents are [101], [108], [112], [107]. [109], [110], [111], and [105].

A collateral evaluation of the NIECS data was conducted by Carl Blumstein, Carl York, and William Kemp [19]. This report has been reviewed and evaluated by Cowing, Dubin, and McFadden (1981b). Independent reports on the weather information contained in the NIECS data set and on procedures used to locate state locations for NIECS households are given in Cowing, Dubin, and McFadden (1981c) and Cowing Dubin, and McFadden (1981d).

Table 1. NIECS Information - A Summary[1]

Housing characteristics
Housing type
Year house built
Number of floors
Floor area
Number of rooms
Number and type of windows
Number and type of storm windows
Number and type of outside doors
Number of storm doors
Presence, type, amount of attic insulation
Wall insulation

Retrofit/conservation efforts[2]
Storm windows
Weatherstripping
Clock thermostat
Attic insulation
Wall insulation
Floor insulation
Hot water pipe insulation
Hot water heater insulation
Other insulation
Caulking
Plastic coverings on windows or doors

Heating/cooling equipment
Main heating system type and fuel
Secondary heating system type and fuel
Type of air conditioning equipment
Number of rooms air conditioned

Household appliances
Fuel used for water heating
Number and type of refrigerators
Number and type of cooking equipment
Use of other household appliances

Demographic characteristics
Number age, sex, and employment status of household members
Marital status of respondent
Race of respondent
Eduction of respondent and spouse
Total household income for 1977
Housing tenure (own or rent)

Energy use and consumption[3]
Use of electrictiy, natural gas LPG, and fuel oil
- for different functions
- paid by household
- consumption, and expenditure

Other information
Geographic location
Heating degree days
Cooling degree days
Type of community

[1]Questions were also asked about ownership and use of motor vehicles, but this information was not relevent to this project.

[2]Refers to conservation actions taken between January 1977 and the date of the interview, fall 1978.

[3]Data on monthly household fuel consumption and expenditures by type of fuel were obtained from fuel suppliers. The data cover the one-year period from April 1978 through March 1979.

2. Additional Variables

In addition to the data items provided directly within NIECS, additional variables were collected and matched to the data base most frequently at the level of the primary sampling unit. Table 2 lists these variables and gives their descriptions.

TABLE 2

Variable Name	Description
AVEPYB	Average electricity price year house built
AVEP78	Average electricity price 1978
AVGPYB	Average gas price year house built
AVGP78	Average gas price 1978
AVOPYB	Average oil price year house built
AVOP78	Average oil price 1978
CDD4170	Cooling degree days 10 yr. normals
CDD78	Cooling degree day 1978
CERTCODE	Certainty code of location match
ELEVAWS	Elevation of ASHRAE Weather Station
ELEVDDWS	Elevation of degree day weather station
ELEVPSU	Elevation (ft.) of PSU Location
HDD4170	Heating degree days 30 yr. normals
HDD7879	Heating degree days in 1978-1979
IINDEX	City cost index for installation (mech. goods)
LATAWS	Latitude of Ashrae weather station
LATDDWS	Latitude of degree-day weather station
LATPSU	Latitude of PSU location
MINDEX	City cost index for materials (mech. goods)
SDDB	Summer design dry bulb
SODR	Summer outdoor daily temperture range
WCMSINDX	Index of matched WCMS PSU
WMAET	Winter median of annual extreme temperatures
W99T	Winter 99 percent temperature

II. Reprocessing the Monthly Billing Data

In this section we discuss the monthly billing data matched to the (NIECS) National Interim Energy Consumption Survey. Following a brief review of the data collection procedure we describe our strategy to re-process the raw billing data into a form useful for econometric analysis. Summary information based on the re-processed data provides a measure of data quality for empirical studies.

NIECS is a four stage area probability sample consisting of 103 primary sampling units. The NIECS sample was drawn from the contiguous United States and the District of Columbia. In final form the sample represents individually specific information on 4081 households. In 3842 cases demographic and structural attributes were obtained by personal interview. In the remaining 239 cases data were obtained by mailed questionnaire and the contractor, Response Analysis Corporation, found it necessary to impute a substantive number of the missing responses. At the completion of each interview, households were asked to sign a Department of Energy waiver allowing Response Analysis to collect data on fuel utilization directly from the appropriate fuel supplier. Utilities responded in varying degrees of completeness. Table 3 summarizes the data collection response rates for 4080 households who used electricity. Referring to Table 3, we see that in approximately three-fourths of the sample at least eleven months of billing data were collected. This is a strikingly high percentage of the cases. In an additional twelve percent of the sample, fewer than ten months of billing data were collected. For these households, the contractor provided imputed annual information using various "hot-deck" and regression estimates. The usefulness of the imputed annual figures for econometric analysis seems questionable so that it would seem best to concentrate empirical efforts on the first group with nearly complete data.

For each household a maximum of twenty billing periods were recorded with an average length of 30 days per billing period. In each billing period the following information was recorded: the expenditure in dollars for the fuel, the quantity in kilowatt-hours for electricity consumed, the beginning year, month, and date, and the ending year, month, and date. Also recorded were a code for whether or not the beginning and ending dates were known or imputed, whether the end of each billing period was an actual or estimated meter reading, and the total number of heating and cooling degree days for the billing period computed to fourteen separate bases.

In all cases the month in which the billing period took place was known with certainty. Documentation provided by Response Analysis Corporation indicates that there were two major categories of billing date completeness.

The first category consists of the majority of dates unknown for all billing periods. In this case, billing periods were assumed to begin on the fifteenth of the month and end of the fifteenth of the following month with the beginning and ending date codes set to indicate that this assumption had been made. The second category consists of households in which specific dates were unknown for only a few periods at the beginning of the billing record. In this case the initial months were assigned a billing date equal to the first known billing date. It is only possible to determine the exact duration of a billing period for those cases in which the beginning and ending dates are known with certainty.

TABLE 3

Energy Consumption Records and Missing Data for Survey Households Using Electricity

	Electricity no. of households	Percent
Total households using fuel	4080	100.0
Data received from fuel supplier	3509	86.0
11 months or more	3023	74.1
5-10 months	340	8.3
Less than 5 months	146	3.6
Household pays directly to supplier - no data available	334	8.2
Household not identified in company records	128	3.1
Company refused to participate	0	-
Company unknown or not located	0	-
Authorization Form not signed	206	5.1
Fuel used included in rent or paid in other way	237	5.8

Source: NIECS: Report on Methodology, Part 1. Household and Utility Company Surveys, Response Analysis Corporation, Princeton, N.J; Feb., 1981, Section 5.

Table 4 exhibits the actual data from the NIECS billing tape for the 90th household. From Table 4 we see that 14 billing periods were coded. Columns C and D indicate whether the beginning and ending dates are known or unknown. The code for this variable is 0 known and 1 unknown. As columns C and D consist of all zeroes, we know that all dates for observation 90 were known with certainty. Reading across the top row of Table 4 we see that the starting date was January 19, 1978 (columns E, F, G), and that the ending date was February 23, 1978 (columns H, I, J), which corresponds to 35 elapsed days (column K). Quantity, expenditure, heating and cooling degree days (base 65) are recorded in columns A, B, L, and M respectively.

In the econometric analysis of the demand for electricity we must insure that all observations correspond to the behavior of economic agents. Thus we follow a procedure for reprocessing the raw data which determined quantities and expenditures for periods of time bounded at either end by actual meter readings. The estimated versus actual code is given in column O of Table 4. The codes in this case are 0 for no data, 1 for actual meter reading, 2 for estimated reading, 8 for no information provided from utility on this item, and 9 for fuel not used. Note that these codes refer to the end of the period so that it is impossible to tell whether period one data is ever actual or estimated.

Given the possibility that a code eight corresponds to an actual meter reading rather than an estimated reading we have followed the convention of bounding observations by code ones or code eights and flagging the later cases to indicate their suspect quality. Given that we do not have any information from the utility for the beginning of period one (i.e. the end of period 0) it would seem useful to treat the

TABLE 4

Observation No. 90

													0	8
403	38.28	0	0	78	1	19	78	2	23	35	1455	0	1	1
290	28.76	0	0	78	2	23	78	3	23	28	920	0	2	1
280	28.50	0	0	78	3	23	78	4	20	28	553	0	3	1
341	35.79	0	0	78	4	20	78	5	23	33	398	8	4	2
261	28.95	0	0	78	5	23	78	6	21	29	37	70	5	2
290	31.42	0	0	78	6	21	78	7	21	30	7	196	6	2
232	25.57	0	0	78	7	21	78	8	20	30	0	303	7	1
280	30.25	0	0	78	8	20	78	9	18	29	30	135	8	2
251	27.38	0	0	78	9	18	78	10	17	29	237	5	9	1
340	33.12	0	0	78	10	17	78	11	20	34	485	0	10	1
331	32.54	0	0	78	11	20	78	12	20	30	833	0	11	1
425	39.54	0	0	78	12	20	79	1	18	29	1020	0	12	1
303	29.29	0	0	79	1	18	79	2	24	37	1528	0	13	1
206	20.32	0	0	79	2	24	79	3	22	26	670	0	14	1
0	0	0	0	0	0	0	0	0	0	0	0	0	15	0
0	0	0	0	0	0	0	0	0	0	0	0	0	16	0
0	0	0	0	0	0	0	0	0	0	0	0	0	17	0
0	0	0	0	0	0	0	0	0	0	0	0	0	18	0
0	0	0	0	0	0	0	0	0	0	0	0	0	19	0
0	0	0	0	0	0	0	0	0	0	0	0	0	20	0
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O

A: Quantity KWH
B: Expenditures in $
C: Begin date known
D: End date known
E: Begin year
F: Begin month
G: Begin day
H: End year
I: End month
J: End day
K: Elapsed days
L: Heating degree days - 65°
M: Cooling degree days - 65°
N: Billing period No.
O: End of period
Actual or estimated code

end of period zero as if it had been assigned with a code eight.

Reading down column 0 we see that the end of period zero i.e. the beginning of period one has the assigned eight code. In the next line we see that the end of period one corresponds to an actual meter reading. Thus billing period number 1 provides a tentatively valid observation. Comparing the rows of Table 4 for billing periods 1 and 2 we see that the end of period one (equivalent to the beginning of period two) is an actual meter reading. Also the end of period two is an actual reading so that billing period two is bounded by actual readings.

As we go down further in the table, we see that the beginning of period 4 is an actual reading but that the ends of periods 4, 5, and 6 are estimated. Not until the end of period 7 do we have another actual reading. We thus aggregate the information in periods 4, 5, 6, and 7 to obtain a single observation bounded at each end with actual meter readings. This aggregated period contains 1124 kilowatt hour consumption (341 + 261 + 290 + 232) and corresponds to 122 days or approximately 4 months.

A computer program (reproduced in Section VI) was written which processes the raw billing data and produces the following variables: flag code given in Table 6, start code, end code, expenditure, heating and cooling degree days (base 65 and base 75), and quantity consumed.

A zero value for the flag code indicates no data, a one indicates that the processed observation is bounded by actual meter readings, a two indicates actual meter readings at both ends of the period but at least one imputed date at either end-point, a three indicates that at least one end-point corresponds to the eight code (no information on type of meter reading), and finally a four corresponds to not knowing whether one of the end-points is actual versus estimated and that at least one end-point has an imputed date.

Table 5 illustrates the reprocessed data for observation 90. In our reprocessing we found it adequate to allow space for up to 15 billing periods rather that the twenty records allowed for in the raw data set. Note that while Table 4 reports information on 14 billing periods, the reprocessed information corresponds to 10 observations in Table 5. The start and end codes summarize the seven variables allocated in the raw data set for beginning and ending dates and elapsed days. The start and end codes are defined as the number of days from January 1, 1978. A negative number thus would correspond to the number of days before January 1, 1978. The difference between the start and end codes for any billing perod is then the elapsed number of days. For example, the start code in Table 5 for the first reprocessed observation indicates that the observation begins 18 days past the first of January, while the end code indicates that the observation ends 53 days past the first of January for an elapsed time of 35 days. This number may be cross checked in Table 4.

Note that the first three reprocessed observations in Table 5 are identical to their counterparts in Table 4. The fourth observation in Table 5 corresponds to the aggregation of periods 4, 5, 6, 7 from Table 4. Finally, the flag code in column 1 of Table 5 is appropriately set for each reprocessed observation as can be checked with the aid of Table 6 and Table 4.

As mentioned above, we have provided for up to 15 billing records for each of the households under consideration. Table 7 provides a summary of the processing of 2018 cases for which the certainty code of housing location match was greater than three and for which the household was

TABLE 5

Observation 90 Re-Processed

Flag	Start Code	End Code	Expenditure	HDD	CDD	Quantity
3.00	18.00	53.00	38.28	1455.00	0.0	403.00
1.00	53.00	81.00	28.76	920.00	0.0	290.00
1.00	81.00	109.00	28.50	553.00	0.0	280.00
1.00	109.00	231.00	121.73	442.00	577.0	1124.00
1.00	231.00	289.00	57.63	267.00	140.00	531.00
1.00	289.00	323.00	33.12	485.00	0.0	340.00
1.00	323.00	353.00	32.54	833.00	0.0	331.00
1.00	353.00	382.00	39.54	1020.00	0.0	425.00
1.00	382.00	419.00	29.29	1528.00	0.0	303.00
1.00	419.00	445.00	20.32	670.00	0.0	206.00
0.00	0.00	0.00	0.00	0.00	0.0	0.00
0.00	0.00	0.00	0.00	0.00	0.0	0.00
0.00	0.00	0.00	0.00	0.00	0.0	0.00
0.00	0.00	0.00	0.00	0.00	0.0	0.00
0.00	0.00	0.00	0.00	0.00	0.0	0.00

TABLE 6

Explanation for Variable Flag

Code	Definition
0	No data
1	Actual meter readings; known dates
2	Actual meter readings; at least one imputed date
3	No data on actual vs. estimated; known dates
4	No data on actual vs. estimated; at least one imputed date

owner-occupied and single-family detached. For details on the location match the reader may consult Cowing, Dubin, and McFadden (1981d). Table 8 provides a similar summary for the processing of the natural gas billing data. Tables 7 and 8 indicate that no information was available for 127 households in the electricity data and that no information was available for 874 households in the natural gas data. However 79.87 percent and 88.13 percent of the electricity and natural gas billing data are assigned a flag code of one which indicates a very high quality for the overall processed data sets.

III. Use of Billing Data to Obtain Marginal Prices

This section considers the construction of the marginal price of electricity and the marginal price of natural gas from the monthly billing data. Details concerning the theory of this calculation (as opposed to its implementation are presented in Chapter 2.)

In the process described of going from the raw monthly data to the processed data, we emphasized a need to bound each observation by actual meter readings. These observations correspond to the behavior of the individual. In determining bills, however, it is likely that estimated as well as actual quantities are applied to the rate schedule by the utility. Thus to determine marginal price we recommend the use of the billing data as it appears on the monthly data set.

Under the assumption that the rate schedule can be approximated by a two-part tariff, an appropriate procedure collects all observations from within a primary sampling unit (this roughly corresponds to the area covered by a single utility), and fits a marginal price using ordinary least squares regression of expenditure on a constant term and quantity:

TABLE 7

Summary Statistics for Variable Flag: Electricity Billing Data

Code	Absolute Frequency	Relative Frequency (PCT)	Adjusted Relative Frequency
0	5809	20.48	-
1	18015	63.51	79.87
2	1496	5.27	6.63
3	2635	9.29	11.68
4	410	1.45	1.82

Total: 28,365

127 missing cases
1891 partial cases

TABLE 8

Summary Statistics for Variable Flag: Natural Gas Billing Data

Code	Absolute Frequency	Relative Frequency (PCT)	Adjusted Relative Frequency (PCT)
0	4827	28.13	-
1	10869	63.34	88.13
2	122	0.71	0.99
3	1195	6.96	9.69
4	147	0.86	1.19

Total: 17,160

874 missing cases
1144 partial cases

(1) $E_t = \alpha + \beta Q_t + V_t$ with:

E_t = expenditure by observation t

Q_t = quantity consumed by observation t

V_t = random error term for observation t

α = fixed charge in two-part tariff

β = marginal price

Before public release, a procedure designed to protect confidentiality randomly adjusted the beginning and ending date of each billing period by up to three days. This innoculation procedure was designed to prevent matching of households with the billing data provided by the fuel supplier. Does this innoculation prevent recovery of marginal rates? Suppose we assume that the two-part tariff is an adequate representation of the billing schedule and that a random fraction ξ_{2t} of billing period two data is assigned to billing period one data to produce an observed (expenditure, quantity) observation (E_t^*, Q_t^*). Let (E_{1t}, Q_{1t}) and (E_{2t}, Q_{2t}) be the true expenditure, quantity pairs for two contiguous billing periods determined by relation (1). Then,

(2) $E_t^* = E_{1t} + \xi_{2t}E_{2t}$ and

(3) $Q_t^* = Q_{1t} + \xi_{2t}Q_{2t}$

From equation (1),

(4) $E_{1t} = \alpha + \beta Q_{1t} + V_{1t}$ and

(5) $E_{2t} = \alpha + \beta Q_{2t} + V_{2t}$. Thus:

(6) $E^*_t = \alpha + \beta Q^*_t + V_{1t} + \xi_{2t}V_{2t} + \alpha\xi_{2t}$ so that

(7) $E^*_t = \alpha + \beta Q^*_t + \varepsilon_t$ where

(8) $\varepsilon_t = V_{1t} + \xi_{2t}V_{2t} + \alpha\xi_{2t}$

If ordinary least squares is an appropriate technique for estimation of (1), it should also provide consistent estimates of the parameters in (7). Thus, the innoculation done by Response Analysis Corporation would not appear to invalidate the basic statistical integrity of the procedure used to determine marginal prices although it is expected that the standard error of the least squares regression will be increased due to the noise introduced by the randomization process.

In Section VI we reproduce the Fortran programs which calculate the marginal prices of electricity and natural gas from the NIECS billing data. The fortran program which processes the raw electricity billing data constructs four marginal prices AEMPE78 - marginal price of electricity for all electric homes, SMPE78 - summer marginal price of electricity, WMPE78 - winter marginal price of electricity in 1978 and OSMPE78 - off season marginal price of electricity in 1978. Consistency conditions and internal checks are imposed on the estimated prices so that at least ten observations are used in the regression analysis and so that winter and summer rates are in fact peak rates. For details the reader is referred to the code itself.

The Fortran program for processing natural gas marginal price does not attempt to discern a seasonal effect. Note that the level of aggregation assumed throughout is that of the primary sampling unit (PSU). We therefore assume that all observations within a given PSU are served by one utility.

IV. Adaptation of Annual Thermal Model to Monthly Billing Data

In this section we summarize the heating and cooling energy calculations analyzed in McFadden and Dubin (1982). The calculation considers the dominant modes of heat transfer between interior and exterior in both the design and normal operational modes. For details concerning either the thermal modeling principles or characteristics of single-family dwellings in NIECS used in the calculations the reader should consult McFadden and Dubin (1982).

1. Summary of Winter Heating Calculation

In Table 9 we reproduce a summary of the winter heating calculation. From Table we find that delivered energy per hour on a winter day with mean ambient temperature t and thermostat setting τ is:

$$(9) \qquad Q = [A_w U_w + A_c U_c + A_{win} U_{win}] (\tau - t) + A_c U_f (\tau - t_g)$$

$$+ \theta V[.0103 + .00015 (\tau - t)](\tau - t) - \text{INTERNAL}$$

The notation is

A_w, A_c, A_{win}	wall, ceiling, and window areas
U_w, U_c, U_{win}, U_f	conductivities of wall, ceiling, window (average), and floor
θ	window infiltration loss factor
V	volume
t_g	ground temperature, assumed constant thoughout the winter
INTERNAL	internal load from occupants and appliances.

We may rewrite (9) in the form

$$(10) \qquad Q = w_3 + w_1 (\tau - t) + w_2 (\tau - t)^2 \qquad \text{with:}$$

$$w_0 = A_c U_f (\tau - t_g)$$

$$w_1 = A_w U_w + A_c U_c + A_{win} U_{win} + .0103 \theta V$$

$$w_2 = .00015 \theta V$$

$$w_3 = w_0 - \text{INTERNAL}$$

TABLE 9

Summary of Winter Heating Capacity Calculation

Design Btuh is the sum of the following components.

1. Wall losses:

$$\begin{bmatrix}\text{Exterior wall area} \\ \text{surrounding heated} \\ \text{space, excluding} \\ \text{windows}\end{bmatrix} \left[\frac{0.9394 + 0.0138\ I_w}{2.85 + I_w}\right] \cdot [75 - t_e]$$

2. Ceiling losses:

$$[\text{Ceiling area}] \cdot [3.834 + 0.943\ I_c]^{-1} \cdot [75 - t_e]$$

3. Floor losses:

$$[\text{Ceiling area}] \cdot [75 - (36 - 0.3\ t_e))/10.05$$

4. Window losses:

$$\left[\frac{A_{ws}}{2.78} + \frac{A_{wn}}{0.98} + \frac{A_{sds}}{1.32} + \frac{A_{sdn}}{0.88}\right] \cdot (75 - t_e)$$

5. Infiltration losses:

$$\left[1.14 - \left[\frac{0.28\ (A_{ws} + A_{sds})}{A_{ws} + A_{wn} + A_{sds} + A_{sdn}}\right]\right] \begin{matrix}(0.25 + 0.02165(15) + 0.00833(75-t_e)) \\ (0.018) \cdot V \cdot (75 - t_e)\end{matrix}$$

Notation:

I_w	R-value of wall insulation (minimum of 0.95 for air gap if no insulation).
I_c	R-value of ceiling insulation
$t_i=75$	interior design temperature (°F)
t_e	exterior winter design temperature (°F)
A_{ws}	area of stormed windows (ft^2)
A_{wn}	area of non-stormed windows (ft^2)
A_{sds}	area of stormed sliding glass doors (ft^2)
A_{sdn}	area of non-stormed sliding glass doors (ft^2)
V	volume of conditioned space (ft^2)

Source: McFadden and Dubin (1982).

The mean and standard deviation of the thermal coefficients w_0, w_1, w_2, w_3 are given in Table 10.

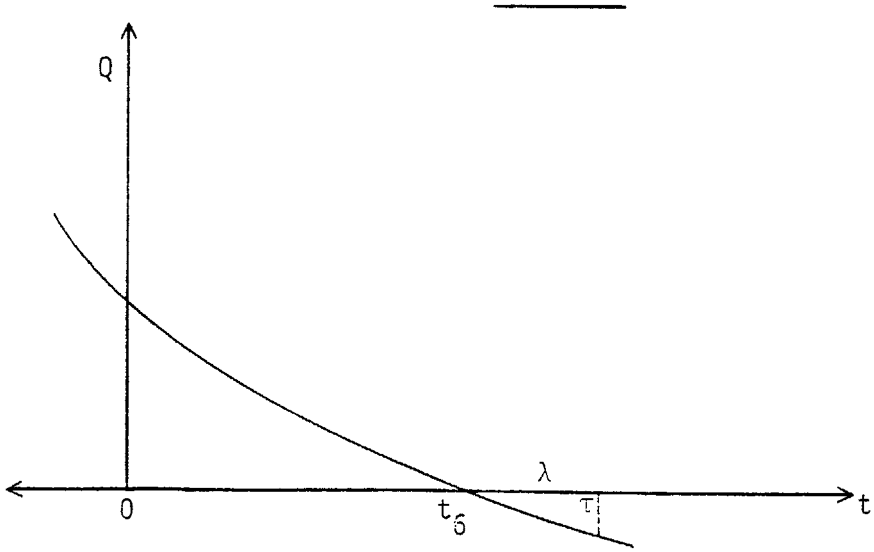
TABLE 10

Variable	Mean[a]	Standard Deviation
w_0	-1604	814.9
w_1	618.5	258.5
w_2	1.666	.7101
w_3	-4050.	1077.

[a]Based on the sub-sample of 2018 households from NIECS in which household is single-family detached, household is owner-occupied, and the certainty code of the location match is one or two (see Cowing, Dubin, and McFadden (1981d) for details.)

We illustrate the heat function in Figure 1.

FIGURE 1



Inspection of equation (10) shows the heat function falls as the daily mean temperature increases and has a slope which is increasing. Furthermore, the effect of INTERNAL causes the heat function to go negative beyond a critical temperature t_6. To maintain the thermostat setting τ it would in fact be necessary to "crack a window" and let some of the internal heat dissapate. The critical temperature t_6 is defined relative to the thermostat setting τ and only the difference $(\tau - t_6)$ is uniquely determined. Note that equation (10) implies:

(11) $$\lambda = (\tau - t_6) = -(w_1/2w_2) * [1 - ((1-4w_2w_3)/w_1^2)^{1/2}]$$

2. <u>Summary of Summer Cooling Calculation</u>

In Table 11 we present a summary of the summer cooling calculation. This calculation considers three types of net heat flows: (1) radiation heat gain during daylight hours, (2) conduction through walls and ceiling, and (3) conduction through windows and infiltration in the presence of the daily cycle of radiation, temperatures, and flywheel effects.

Let q_o denote peak radiation heat gain (before adjustment for latent heat). From Table 11,

$$q_o = 13.6\, A_w U_w + 13.77\, A_c U_c + 37.5\, A_{win} U_{win},$$

where A_w, A_c, A_{win} are wall, ceiling, and window areas, and U_w, U_c, U_{win} are corresponding conductivities. The radiation at hour h (with h=0 at noon) is approximately:

$$Q_R(h) = q_o \max(0, \cos \frac{\pi h}{12}) \qquad |h| \leq 12$$

TABLE 11

Summary of Summer Cooling Capacity Calculation

Design Btuh is the sum of the following components:

1. Wall gains:

$$\begin{bmatrix}\text{Exterior wall} \\ \text{area surrounding} \\ \text{conditioned space,} \\ \text{excluding windows}\end{bmatrix} \cdot (13.6 + t_e - 75 - 0.5t_r) \cdot \left[\frac{(0.9394 + 0.0138\ I_w)}{2.85 + I_w}\right]$$

2. Ceiling gains:

$$\begin{bmatrix}\text{Ceiling} \\ \text{Area}\end{bmatrix} \cdot \left[\frac{(0.9276 + 0.0165\ I_c)}{(1.916 + 0.608\ I_c)}\right] \cdot [13.77 - 0.202\ t_r + 0.592\ (t_e - 75)]$$

3. Window gains (assuming storms removed):

$$(A_{ws} + A_{wn} + A_{sds} + A_{sdn})\ (0.8\ t_e - 30)$$

4. Internal load: (INTERNAL)

1200 + 400 (number of occupants)

5. Infiltration gains:

$$0.018 \cdot V \cdot (t_e - 75) \cdot [0.25 + 0.02165(7.5) + 0.00833\ (t_e - 75)]$$

The sum of 1-5 is increased by 30 percent to account for latent heat load (dehumidification)

Notation:

t_e	summer design maximum temperature (°F)
t_r	summer design temperature range (°F)
I_w	R-value of wall insulation
I_c	R-value of ceiling insulation
$A_{ws} + A_{wn} + A_{sds} + A_{sdn}$	total area of windows and sliding glass doors
V	volume of conditioned space

Source: McFadden and Dubin (1982)

Conduction through walls and ceiling, internal load, and average window conduction is assumed uniform over the day due to flywheel effects and equals:

$$Q_A(t) = A_w U_w (t-\tau) + A_c U_c (.592)(t-\tau) - A_{win} U_{win} (t-\tau) +$$

$$(.0074)V(t-\tau) + \text{INTERNAL} \quad = \quad q_1 + q_2\,(t-\tau)$$

Finally, net heat gain which varies with the tempreature cycle, due to infiltration, attic ventilation, and cyclic window conduction is given by:

$$Q_v(h) = [2(.094)A_c U_c + A_{win} U_{win} + (.0074)V]\frac{t_r}{2}\cos(\frac{\pi h}{12})$$

$$= q_3 \cos(\frac{\pi h}{12})$$

where t_r = summer outdoor temperature range.

Combining these sources, net energy gain at hour h is:

$$Q = Q_r(h) + Q_A(t) + Q_v(h)$$

$$= q_o \max(0, \cos(\frac{\pi h}{12})) + q_1 + q_2(t-\tau) + q_3 \cos(\frac{\pi h}{12})$$

The following approximation is derived in McFadden and Dubin (1982) to determine the average BTU's per hour extracted by the air conditioner during a twenty-four hour period:

$$
(12) \qquad Q = \begin{cases} 0 & \text{for } t < t_1 \\ ((t-t_1)/(t_2-t_1)) \cdot q_4 + (t-t_1)(t_2-t) \cdot q_8 & \text{for } t_1 \leq t < t_2 \\ q_4 + ((t-t_2)/t_3-t_2)).(q_5-q_4)+(t-t_2)(t_3-t).q_9 & \text{for } t_2 \leq t < t_3 \\ 1.3((q_0/\pi) + q_1 + q_2\ (t-\tau)) & \text{for } t \geq t_3 \end{cases}
$$

where:

$t_1 \quad = \tau - (q_0 + q_1 + q_3)/q_2$

$t_2 \quad = \tau - q_1/q_2$

$t_3 \quad = \tau - (q_1-q_3)/q_2$

$Q(t_1) \quad = 0$

$Q(t_2) = q_4 = 1.3(q_0+q_3)/\pi$

$Q(t_3) = q_5 = 1.3(q_3+q_0/\pi)$

$t_4 \quad = \tau - (q_1 + (q_0+q_3)/\sqrt{2})/q_2$

$t_5 \quad = \tau - (q_1 - q_3/\sqrt{2})/q_2$

$Q(t_4) = q_6 = 1.3(q_0 + q_3)(1/\pi - 1/4)/\sqrt{2}$

$Q(t_5) = q_7 = 1.3(q_0/\pi + q_3(1/\pi + 3/4)/\sqrt{2})$

$q_8 \quad = [q_6 - ((t_4-t_1)/(t_2-t_1))q_4]/(t_4-t_1).(t_2-t_4)$

$q_9 \quad = [q_7 - q_4 - ((t_5-t_2)/(t_3-t_2)).(q_5 - q_4)]/(t_5-t_2).(t_3-t_5)$

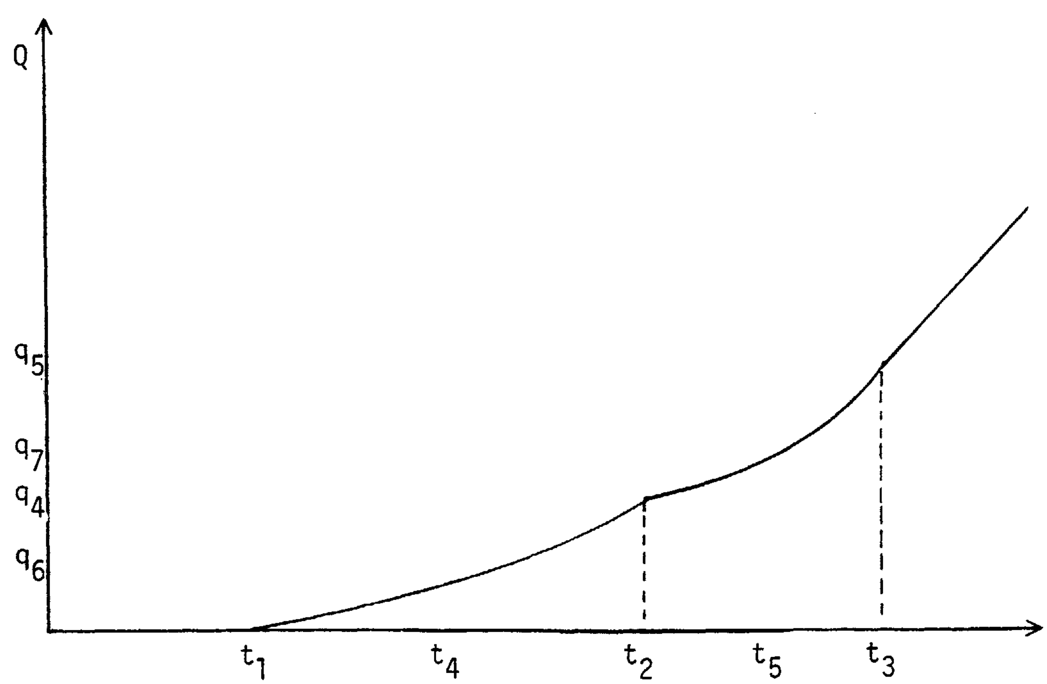
Means and standard deviations of the thermal coefficients are given in Table 12.

TABLE 12

Variable	Mean	Standard Deviation
Q_0	12150.	5327.
Q_1	2446.	664.6
Q_2	562.7	231.1
Q_3	2717.	1354.
Q_4	6151.	2685.
Q_5	8560.	3791.
Q_6	933.3	407.9
Q_7	7696.	3386.
Q_8	-6.191	3.056
Q_9	-37.22	26.16

We illustrate the cooling function in Figure 2 .

FIGURE 2



The temperatures t_1, t_2, t_3 define distinct cooling ranges. Below temperature t_1 there is no predicted cooling. In the range t_1 to t_2 there is daytime cooling only and the cooling function has been approximated by a quadratic which is increasing in the daily mean temperature at an increasing rate (reflecting the sign of the average value of q_8). In the range t_2 to t_3 there is continuous cooling which is again approximated by the convex shaped quadratic. Beyond temperature t_3 cooling is again continuous however the cooling function is now linear relfecting a range of daily mean temperatures which exceed the thermostat setting τ.

3. Determination of Energy Consumption Levels for the NIECS Billing Data

Following the approach of McFadden and Dubin (1982) let $F(t) = (1 + e^{-b(t-\mu)})^{-1}$ denote a logistic approximation to the cumulative distribution of daily mean temperatures for a given billing period. To determine total energy consumption for heating we integrate the heat function in equation (10) for all temperatures below the critical temperature t_6. Total delivered heat per hour averaged over the billing period is then:

$$\int_{-\infty}^{\min[\tau,t_6]} [w_3 + w_1(\tau-t) + w_2(\tau-t)^2]F'(t)dt$$

When $\tau \leq t_6$, the integral (13) may be evaluated by:

$$w_3P_\tau - w_1/b \cdot \ln[1-P_\tau] + 2w_2/b^2 \cdot \gamma\,[b(\tau-\mu)]$$

where $\gamma(\lambda) = \int_{-\infty}^{\lambda} \ln[1+e^{s}]\, ds$. In the case $\tau > t_6$ note that:

$$\int_{-\infty}^{t_6} [w_3 + w_1(\tau-t) + w_2(\tau-t)^2]\, F'(t)\, dt$$

$$= \int_{-\infty}^{t_6} [w_1 + 2w_2(\tau-t_6)](t_6-t) + w_2(t_6-t)^2] F'(t) dt \qquad \text{as}$$

$$w_3 + w_1(\tau-t) + w_2(\tau-t)^2$$

$$= w_3 + w_1(t_6-t+\tau-t_6) + w_2(t_6-t+\tau-t_6)^2$$

$$= [w_1+2(\tau-t_6)](t_6-t) + w_2(t_6-t)^2 + [w_3 + w_1(\tau-t_6) + w_2(\tau-t_6^2)]$$

$$= [w_1+2(\tau-t_6)](t_6-t) + w_2(t_6-t)^2$$

since $[w_3+w_1(\tau-t_6) + w_2(\tau-t_6)^2] = 0.$

Evaluation of the integral (13) yields:

$$-[w_1+2w_2(\tau-t_6)]/b \,.\, \ln[1-P_{t_6}] + 2w_2/b^2 \,.\, \gamma[b(t_6-\mu)]$$

The calculation of cooling per hour averaged over the distribution of daily mean temperatures similarly requires the integration of the cooling function (12) from the critical temperature t_1 up to the upper limit of the temperature distribution. To facilitate the integration of the cooling function the following moments are derived:

$$\mu_1(t') = \int_{-\infty}^{t'} (t'-t)F'(t)dt = \frac{1}{b}\ln[1+e^{b(t'-\mu)}] = \frac{-1}{b}\ln[1-P_{t'}]$$

$$= (t'-\mu) + \frac{1}{b}\ln[1 + e^{-b(t'-\mu)}]$$

$$\mu_2(t') = \int_{-\infty}^{t'} (t'-t)^2F'(t)dt = \frac{2}{b^2}\ \gamma[b(t'-\mu)] \qquad \text{so that}$$

$$\xi_1(t',t'') = \int_{t'}^{t''} (t''-t)F'(t)dt = (t''-t')[1 - P_{t'}] + \frac{1}{b}\ln[\frac{P_{t'}}{P_{t''}}] \quad \text{and}$$

$$\xi_2(t',t'') = \int_{t'}^{t''} (t''-t)(t-t')F'(t)dt = (t''-t')[\mu_1(t')+\mu_1(t'')]$$

$$+ \mu_2(t') - \mu_2(t'')$$

Finally, integration of the cooling function (12) yields:

$$q_4/(t_2-t_1) \cdot \xi_1(t_2,t_1) + q_8 \cdot \xi_2(t_1,t_2) +$$

$$q_4 \cdot (P_{t_3}-P_{t_2}) + ((q_5-q_4)/(t_3-t_2)) \cdot \xi_1(t_3,t_2) + q_9 \cdot \xi_2(t_2,t_3) +$$

$$q_5 \cdot (1 - P_{t_3}) - 1.3(q_2/b)\ln[P_{t_3}]$$

Application of the formulae for $\xi_1(t'',t')$ and $\xi_2(t',t'')$ require modification to allow for numerical indeterminancies occurring at high temperatures. Consider first the formula for $\xi_1(t_B,t_A)$ with $t_B > t_A$:

$$\xi_1(t_B,t_A) = \mu_1(t_A) - \mu_1(t_B) - (t_A-t_B)F(t_B)$$

$$= -1/b \ln[1-P_{t_A}] + 1/b \ln[1-P_{t_B}] - (t_A-t_B) P_{t_B}$$

$$= 1/b \ln[(1-P_{t_B})/(1-P_{t_A})] - (t_A-t_B) P_{t_B}$$

$$= 1/b [\ln(1 + e^{b(t_A-\mu)}) - \ln(1 + e^{b(t_B-\mu)})] - (t_A-t_B)P_{t_B}$$

When $b(t_B-\mu)$ is sufficiently large so that P_{t_B} is approximately equal to one, we have:

$$\xi_1(t_{high},t_A) \doteq \frac{1}{b} \ln(1+e^{b(t_A-\mu)}) - (t_B-\mu) - (t_A-t_B)$$

$$= -(t_A-\mu) + \frac{1}{b} \ln(1 + e^{b(t_A-\mu)}) = \frac{-1}{b} \ln [P_{t_A}]$$

In the calculation of $\xi_2 (t_A,t_B)$ note that:

$$\xi_2(t_A,t_B) = [t_B-t_A][\mu_1(t_A)+\mu_1(t_B)] + \mu_2(t_A)-\mu_2(t_B)$$

Since $\mu_1(t_A) + \mu_1(t_B) = \frac{-1}{b} \ln[1-P_{t_A}] - \frac{1}{b} \ln[1-P_{t_B}]$

$$= \frac{-1}{b} \ln[(1-P_{t_A})(1-P_{t_B})]$$ and

$(1 - P_{t_A}) = [1+e^{b(t_A-\mu)}]^{-1}$ we have:

$$\mu_1(t_A) + \mu_1(t_B) = \frac{1}{b} \ln(1+e^{b(t_A-\mu)}) + \frac{1}{b} \ln(1+e^{b(t_B-\mu)}).$$

In the case in which $b(t_B-\mu)$ is large we have;

$$\mu_1(t_A) + \mu_1(t_{high}) = (t_A-\mu) - \frac{1}{b} \ln P_{t_a} + (t_{high}-\mu).$$

Finally, the calculation of $\mu_2(t_B)$ when $b(t_B-\mu)$ is large follows from:

$$\mu_2(t_{high}) = \int_{-\infty}^{t_{high}} (t_{high}-t)^2 F'(t)dt$$

$$\doteq \int_{-\infty}^{\infty} (t_{high}-t)^2 F'(t)dt = VAR(t) + (\mu-t_{high})^2$$

$$= \pi^2/3b^2 + (t_{high}-\mu)^2$$

Since $\gamma[b(t_{high}-\mu)] = \frac{b^2}{2} \mu_2(t_{high})$ we have

$$\gamma[b(t_{high}-\mu)] \doteq \frac{\pi^2}{6} + \frac{1}{2}[b(t_{high}-\mu)]^2$$

The empirical determination of the paramters b and μ from observations of heating and cooling degree days measured at similar and dissimilar bases is discussed in McFadden and Dubin (1982).

The logistic distribution provides reasonably stable temperature profiles provided the number of heating or cooling degree days per day during a billing period is not "too small." In the exceptional cases the temperature distribution is taken to be a unit mass at the mean temperature.

This completes the summary of the heating and cooling calculations analyzed in McFadden and Dubin (1982). In Section VI we include a listing of the Fortran program which performs the billing period analysis. Inputs to the program are the processed billing period data as described in Section II, cooling coefficients q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7, and the heating coefficients XXLAM, W1A, W1, W2, W3A, W3 where:

XXLAM $= -\lambda = (t_6 - \tau)$

W1A $= w_1$ when $\tau \leq t_6$ and $(w_1+2w_2(\tau - t_6))$ when $\tau > t_6$

W2 $= w_2$

W3A $= w_3$ when $\tau \leq t_6$ and 0 when $\tau > t_6$

W3 $= w_3$

Note that the heating and cooling coefficients remain constant over different billing periods for a given household. Outputs of the program are predicted usage in thousands of BTU's for heating and cooling when winter thermostat setting is 70 degrees and summer thermostat setting is 75 degrees as well as the predicted changes in these consumption levels for a one degree change in the thermostat setting. The latter estimates are used in the computation of the marginal price of comfort. Finally, the critical temperatures t_1, t_2, t_3, and t_6 as well as an estimate of mean temperature are provided for each billing period.

4. Standardization of Billing Period Data

To prepare the processed billing data for analysis we have aggregated the fifteen or fewer observations per household into three distinguishable cases. The aggregation takes place however by temperature rather than time. The first case collects all observations for which the daily mean temperature is less than the critical temperature t_1. This corresponds to a period in which there is no cooling and in which there is likely to be continuous heating. The second case collects observations for each household in which the daily mean temperature exceeds critical temperature t_1 but is lower than the critical temperature t_6. In this situation households are likely to be experiencing positive heating and cooling degree days and will utilize both heating and cooling modes. The last case collects observations for which the daily mean temperature exceeds critical temperature t_6. This case corresponds to temperatures for which heating is unnecessary. Tables 13 and 14 give mean values for the aggregated billing data by fuel type and thermal mode. SHUEC refers to predicted heating usage in thousands of BTU's. ACUEC refers to predicted cooling usage in thousands of BTU's. The variables DSHUEC and DACUEC give the marginal increase in energy utilization for a one degree change in thermostat setting sustained for the period in question. In the heating mode this corresponds to raising ambient temperature from 70 to 71 degrees while in the cooling mode this corresponds to a change in temperature from 75 to 74 degrees. Note that usage has not been adjusted to reflect the coefficient of HVAC performance and that mean values are presented for all available observations independent of their chosen system type.

Table 13

Mean Values of Aggregated Billing Data by Thermal Mode - Electricity

	No Cooling	Heat and Cooling	No Heating
DAYS	183	149	137
HDD65	6783	1921	154
CDD65	.8954	75.96	1049
QUAN(KWH)	6719	4884	4917
EXPEN($)	257	264	255
SHUEC	103200	29120	1312
DSHUEC	3250	2319	309
ACUEC	55.57	6434	24180
DACUEC	9.05	722	1630

Table 14

Mean Values for Aggregated Billing Data by Thermal Mode - Natural Gas

	No Cooling	Heat and Cooling	No Heating
DAYS	189	157	131
HDD65	7064	2071	177
CDD65	4.362	94.83	959.6
QUAN(KWH)	1479	544	172
EXPEN($)	376	140	55
SHUEC	107100	32160	1508
DSHUEC	3326	2480	342
ACUEC	176	7091	22930
DACUEC	22	777	1524

V. Case Study of Household Number 1271

This section illustrates the processing of data from a selected household in the NIECS data file. The household was selected on the basis of three criteria: the household resides in Boston, Masschusetts (a location in which additional weather related information was readily available), the household had available electricity and natural gas billing data, and the household selected one of nineteen alternative HVAC systems of particular interest to our study. The household selected is identified by a unique Department of Energy identification number which in this case is 1271.

Table 15 and Table 16 present the re-processed billing data for household 1271. The electricity billing data cover a period of 462 days while the gas billing data are for a period of length 394 days. Table 17 and Table 18 present the thermal model output for electricity and natural gas respectively. Table 19 presents the actual values of selected variables for household 1271. To compare the processed billing data with the annual information (including the thermal model output based on the annual data) we have selected a subset of the observations which correspond to a period of approximately one year. These subsets lie within the dotted lines in Tables 15, 16, 17, and 18. Tables 20 and 21 present the results of adding together the billing data for the year. Note that ACUEC, DACUEC, SHUECG, and DSHUECG in Tables 20 and 21 have not been adjusted to reflect system coefficient of performance, while similar numbers in Table 19 do reflect COP adjustments. As may be seen by inspection, the estimates in Tables 20 and 21 compare very favorably with each other and with those of the annual file (in Table 19). Furthermore, the thermal model aggregates very well across time and gives values which track the temperature profile quite well.

Tables 22 and 23 presents the aggregated billing data by thermal mode and fuel type as described in Section IV.4. Table 22 implies unit electricity consumptions (UEC) of 168.8 KWH/day in the heating season and 11.99 KWH/day in the cooling season for electric resistance heating and air-conditioning respectively.

Tables 24 and 25 present the thermal model coefficients and critical temperatures for household 1271. Figure 3 displays the heating function (MBTUH) and Figure 4 displays the cooling function (MBTUH). The horizontal axis is daily mean temperatures. Over the range in which the thermal mode is utilized, the relationships are quite linear. Note, however, that these functions embody the attributes of a particular structure with given insulation levels and may well shift remarkably from household to household.

Table 26 presents the operating and capital costs for ten alternative HVAC systems facing household 1271 in the year of house contruction 1962. Costs have been normalized to 1967 dollars. Details on capacity estimation and allocation of capital costs are given in McFadden and Dubin (1982) and Cowing, Dubin, and McFadden (1981e). In Figure 5, we plot captial against operating costs. Given gas availability and conditional on not choosing air-conditioning it is interesting to note that household 1271 chooses the gas hydronic system 3 which appears dominated by the gas space heating system 1. The challenge of the discrete choice model is to adequately describe the choice process in the presence of unobserved cost components.

Table 15

Electricity Billing Data - Household 1271

HHIDNO	FLAG	Start Code	End Code	QUAN	EXPEN	HDD65	CDD65	HDD75	CDD75
1271	3	-16	19	1199	61.95	1273	0	1633	0
1271	1	19	80	2026	97.24	2442	0	3072	0
1271	1	80	109	913	45.79	695	0	995	0
1271	1	109	139	719	37.98	470	0	780	0
1271	1	139	170	721	37.59	99	14	405	0
1271	1	170	201	832	41.23	22	119	225	2
1271	1	201	232	930	44.78	2	216	122	16
1271	1	232	261	760	37.83	100	35	365	0
1271	1	261	292	820	41.22	390	0	710	0
1271	1	292	321	830	40.91	510	0	810	0
1271	1	321	353	968	47.80	974	0	1304	0
1271	1	353	382	876	40.79	1062	0	1362	0
1271	1	382	410	815	35.08	1201	0	1491	0
1271	1	410	446	1059	44.65	1176	0	1546	0
0	0	0	0	0	0	0	0	0	0

Table 16

Natural Gas Billing Data - Household 1271

HHIDNO	FLAG	Start Code	End Code	QUAN	EXPEN	HDD65	CDD65	HDD75	CDD75
1271	3	87	119	171.20	68.79	706	0	1036	0
1271	1	119	178	128.80	59.73	391	37	964	0
1271	1	178	239	103.90	45.87	34	318	362	16
1271	1	239	300	175.20	75.92	593	23	1200	0
1271	1	300	361	271.80	111.62	1654	0	2284	0
1271	1	361	481	663.20	266.57	3914	0	5154	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Table 17

Thermal Model Output - Household 1271 (Electricity)

SHUEC	DSHUEC	ACUEC	DACUEC	T1	T2	T3	T6	TEMP
20800.66	669.35	0	0	42.84	70.15	75.73	63.66	28.34
40604.20	1189.67	0	0	42.84	70.15	75.73	63.66	24.64
10631.69	517.98	0	0	42.84	70.15	75.73	63.66	40.69
6663.95	510.35	428.57	111.63	42.84	70.15	75.73	63.66	49.00
914.30	315.24	2746.87	240.60	42.84	70.15	75.73	63.66	61.94
202.87	77.13	4443.27	316.51	42.84	70.15	75.73	63.66	67.98
42.98	19.03	5700.77	388.10	42.84	70.15	75.73	63.66	71.76
1072.82	257.14	2787.20	233.02	42.84	70.15	75.73	63.66	62.48
5273.07	517.54	831.19	145.03	42.84	70.15	75.73	63.66	52.10
7397.13	499.06	231.27	90.60	42.84	70.15	75.73	63.66	47.07
15469.52	592.64	0	0	42.84	70.15	75.73	63.66	34.25
17405.55	555.51	0	0	42.84	70.15	75.73	63.66	28.03
20223.63	554.35	0	0	42.84	70.15	75.73	63.66	21.75
18871.11	674.80	0	0	42.84	70.15	75.73	63.66	32.06
0	0	0	0	0	0	0	0	0

Table 18

Thermal Model Output - Household 1271 (Natural Gas)

SHUEC	DSHUEC	ACUEC	DACUEC	T1	T2	T3	T6	TEMP
10634.64	565.23	0	0	42.84	70.15	75.73	63.66	42.63
4798.60	737.95	4179.21	401.15	42.84	70.15	75.73	63.66	58.71
311.75	110.21	9836.77	683.19	42.84	70.15	75.73	63.66	69.57
7862.02	877.43	3185.15	348.32	42.84	70.15	75.73	63.66	55.36
25796.79	1109.09	0	0	42.84	70.15	75.73	63.66	37.58
62916.17	2249.40	0	0	42.84	70.15	75.73	63.66	32.05
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Table19

Selected Variables from NIECS for Household 1271

HDD4170	6848	heating degree days
CDD4170	387	cooling degree days
HDD7879	7057	heating degree days
CDD78	378	cooling degree days

NXELYR	$495
NCELYRP	10214 KWH
NXNGYR	$567
NCNGYRB	1370.10 Therms

WMPE78	.045172	$/KWH
SMPE78	.049483	$/KWH
OSMPE78	.045172	$/KWH
AEMPE78	.045172	$/KWH
AVEP78	.053319	$/KWH
AVGP78	.40778	$/Therm
MPG78	.31540	$/Therm

ACUEC	4082	MBTU
DACUEC	368	MBTU
SHUECE	103580	MBTU
DSHUECE	5534	MBTU
SHUECG	141130	MBTU
DSHUECG	7541	MBTU

Table 20

Aggregated Monthly Billing Data - Electricity

Household 1271

DAYS	363
KWH	10395
EXPEN	513
HDD65	6766
CDD65	384
HDD75	10150
CDD75	18
SHUECE	105678
DSHUECE	5051
ACUEC	17169
DACUEC	1525

Table 21

Aggregated Monthly Billing Data - Natural Gas

Household 1271

DAYS	362
Therms	1343
EXPEN	560
HDD65	6586
CDD65	378
HDD75	9964
CDD75	16
SHUECE	101685
DSHUECE	5084
ACUEC	17201
DACUEC	1433

Table 22

Aggregated Electricity Billing Data - Household 1271

HHIDNO	FLAG[1]	QUAN	EXPEN	HDD65	CDD65	SHUEC	DSHUEC	ACUEC	DACUEC	DAYS
1271	1.29	7856	373.30	8823	0	144006	4754	0	0	250
1271	1.00	3850	195.53	1569	49	21321	2099	7025	821	150
1271	1.00	1762	86.01	24	335	246	96	10144	705	62

[1]Average of aggregated flag values.

Table 23

Aggregated Natural Gas Billing Data - Household 1271

HHIDNO	FLAG[1]	QUAN	EXPEN	HDD65	CDD65	SHUEC	DSHUEC	ACUEC	DACUEC	DAYS
1271	1.67	1106	446.98	6274	0	99348	3924	0	0	213
1271	1.00	304	135.65	984	60	12661	1615	7364	749	120
1271	1.00	104	45.87	34	318	312	110	9837	683	61

[1]Average of aggregated flag values.

Table 24

Thermal Coefficients for Household 1271

Q0	13552.00
Q1	2400.00
Q2	623.54
Q3	3476.30
Q4	7046.50
Q5	10127.00
Q6	1069.30
Q7	9021.80
Q8	-6.4391
Q9	-31.5420
W0	-2245.40
W1	617.21
W1A	648.48
W2	2.1305
W3	-4645.40
W3A	0

Table 25

Critical Temperatures for Household 1271

T1[a]	42.84
T2[a]	70.15
T3[a]	75.73
T4[a]	50.84
T5[a]	75.09
TEMP[b]	47.30
T6[c]	63.66

[a]based on $\tau = 74$
[b]based on annual temperature distribution
[c]based on $\tau = 70$

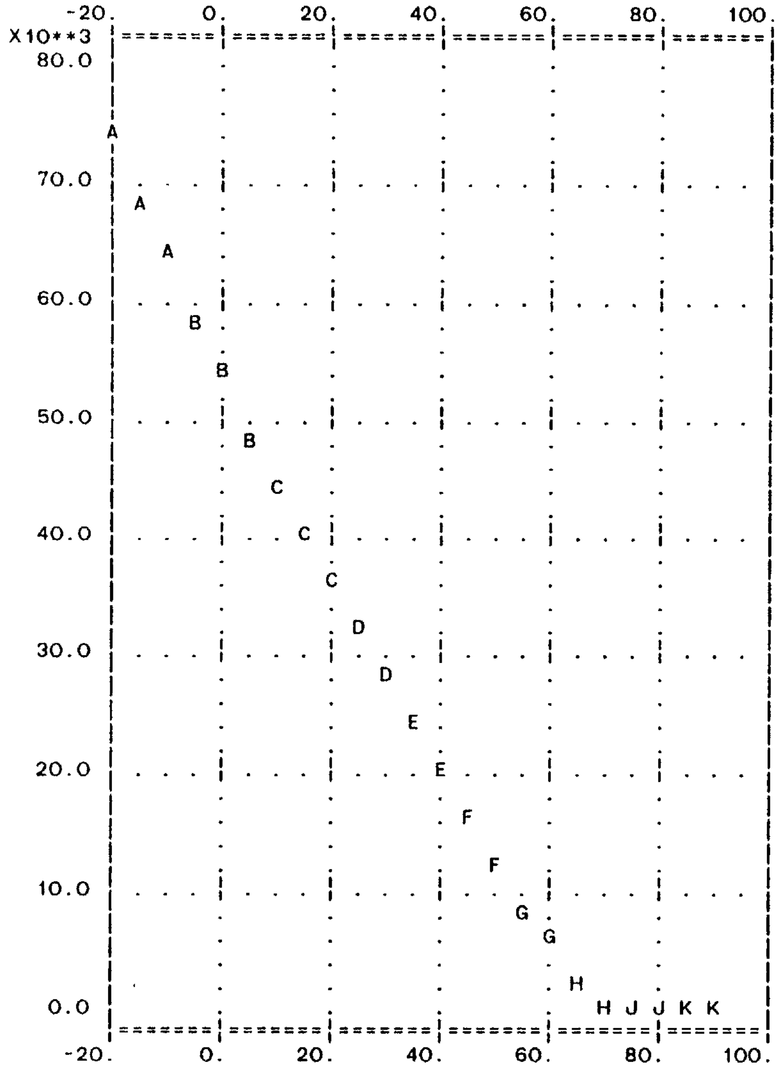


Figure 3

Heat Function for Household 1271

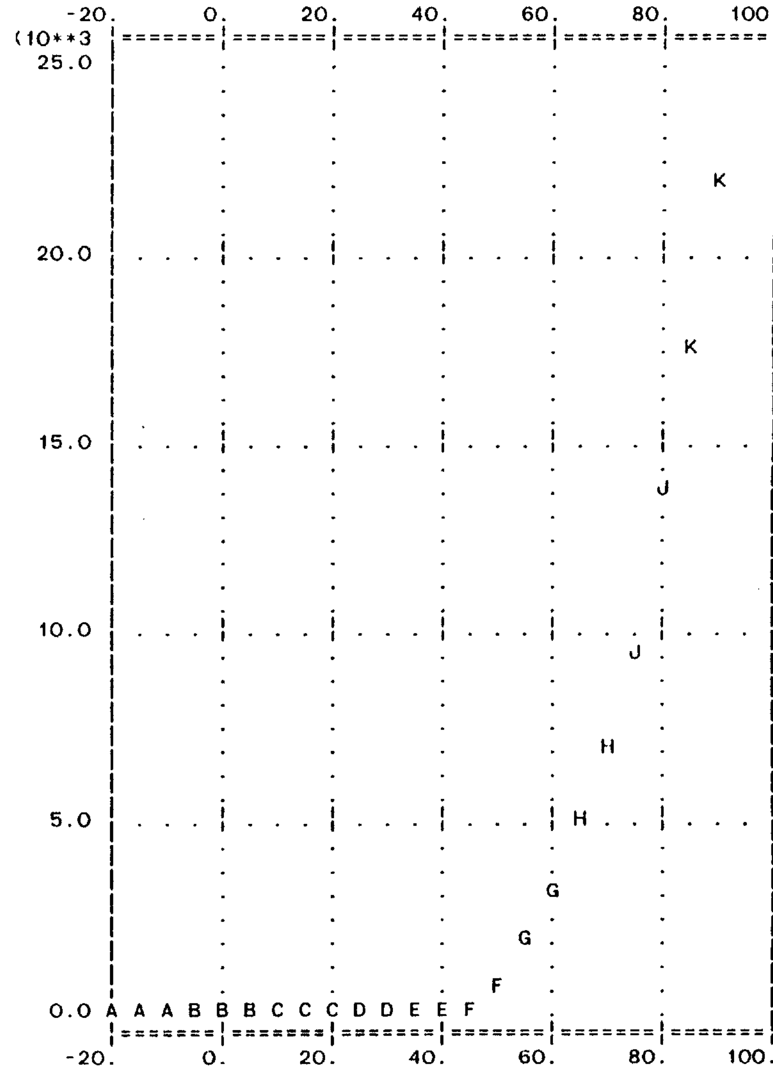


Figure 4

Cooling Function for Household 1271

Table 26

Operating and Capital Costs of Alternative HVAC

in Year House Built - Household 1271

1967 Dollars

OPCST1	341.54
CAPCST1	1201.80
OPCST2	385.02
CAPCST2	2043.30
OPCST3	315.83
CAPCST3	2788.90
OPCST7	139.83
CAPCST7	1834.40
OPCST8	183.31
CAPCST8	2424.20
OPCST9	129.30
CAPCST9	3272.10
OPCST13	1203.20
CAPCST13	982.60
OPCST14	1246.70
CAPCST14	1930.50
OPCST15	630.83
CAPCST15	5084.90
OPCST18	1103.20
CAPCST18	1129.70
ACHEAT	28.731
SHEATN	56.968
SHEATD	62.136
SHEATP	57.458

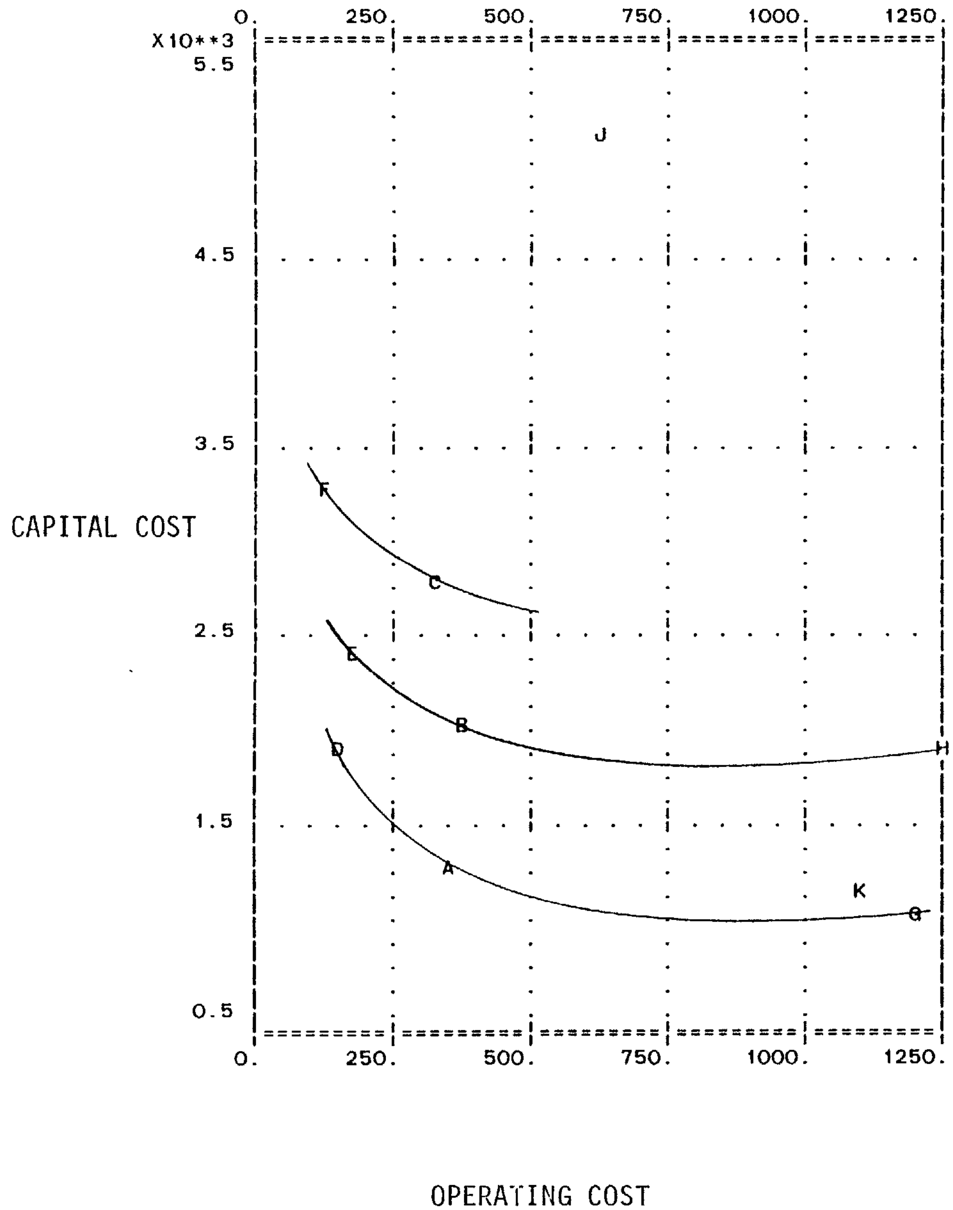


Figure 6

Graph of operating and capital costs for alternative HVAC systems for household 1271.

	HVAC #
A	1
B	2
C	3
D	7
E	8
F	9
G	13
H	14
J	15
K	18

VI. Computer Programs and Selected Output

1. Reprocessing the Raw Electricity Billing Data

For documentation on the billing tape see: "Technical Documentation for the Residential Energy Consumption Survey: National Interim Energy Consumption Survey 1978-1979, Household Monthly Energy Consumption and Expenditure, Public Use Data Tapes, User's Guide, August, 1981 (forthcoming NTIS). Input file one of the program corresponds to the ninth data file on the montly billing tape.

2. Reprocessing the Raw Natural Gas Billing Data

Input file one of the program corresponds to the tenth data file on the monthly billing tape.

3. Determination of Seasonal Marginal Prices for Electricity Billing Data

A) OUTPUT - Marginal prices by primary sampling unit

B) OUTPUT - Record of observations processed

4. Determination of the Marginal Price of Natural Gas

A) OUTPUT - Marginal price by Primary Sampling Unit

B) OUTPUT - Record of observations processed

5. Thermal Load Model for Processed Billing Data

1. Reprocessing the Raw Electricity Billing Data

```
      INTEGER P11,P1,P2
      LOGICAL ACTUAL,ESTIM,KNOW,UNKNOW
      DIMENSION A(20,14),METER(20),B(15,10)
      SUMO=0.0
      SUM1=0.0
      SUM2=0.0
      SUM3=0.0
      SUM4=0.0
      SUMOB=0.0
       DO 5 K=1,3842
      READ(4,15) SAMPLE
15    FORMAT(F3.1)
      IF (SAMPLE.EQ.0.0) GO TO 7
      DO 10 I=1,15
      DO 10 J=1,10
10    B(I,J)=0.0
      READ(1,6) HHIDNO,NBILLS,((A(I,J),J=1,14),I=1,20),(METER(I),
     1 I=1,20)
6     FORMAT(F4.0,6X,I2,20(4X,F7.1,3X,F5.2,2F1.0,6F2.0,63X,2F5.0,
     1 40X,2F5.0,20X),20I1)
      IF(NBILLS.EQ.99) GO TO 100
      P1=0
      P2=0
      NOB=1
20    IF (P2.EQ.NBILLS) GO TO 150
      P2=P2+1
      IF ((METER(P2).NE.1).AND.(METER(P2).NE.8)) GO TO 20
      QUAN=0.0
      EXPEN=0.0
      HDD65=0.0
      CDD65=0.0
      HDD75=0.0
      CDD75=0.0
      P11=P1+1
      DO 30 J=P11,P2
C     ACCUMULATE EXPEN,QUAN,HDD,CDD
      QUAN=QUAN+A(J,1)
      HDD65=HDD65+A(J,11)
      CDD65=CDD65+A(J,12)
      HDD75=HDD75+A(J,13)
      CDD75=CDD75+A(J,14)
      EXPEN=EXPEN+A(J,2)
30    CONTINUE
C     CONVERT BEGINING AND ENDING DATES TO SUMMARY NUMBER
      BY=A(P11,5)+1900.0
      BM=A(P11,6)
      BD=A(P11,7)
      EY=A(P2,8)+1900.0
      EM=A(P2,9)
      ED=A(P2,10)
C     CALCULATION FOR BEGINING PERIOD
      IF (BM.GT.2.0) GO TO 31
      NB=INT(365.25*(BY-1.0))+INT(30.6*(BM+13.0))+INT(BD)-621049
      GO TO 32
31    NB=INT(365.25*BY)+INT(30.6*(BM+1.0))+INT(BD)-621049
```

```
32    CONTINUE
C     CALCULATION FOR ENDING PERIOD
      IF (EM.GT.2.0) GO TO 33
      NE=INT(365.25*(EY-1.0))+INT(30.6*(EM+13.0))+INT(ED)-621049
      GO TO 34
33    NE=INT(365.25*EY)+INT(30.6*(EM+1.0))+INT(ED)-621049
34    CONTINUE
      N1178=101479
      NB=NB-N1178
      NE=NE-N1178
C     CALCULATION OF FLAG CODE
      IF (P1.EQ.0) GO TO 60
      ACTUAL=((METER(P1).EQ.1).AND.(METER(P2).EQ.1))
      ESTIM=.NOT.(ACTUAL)
      GO TO 70
60    ESTIM=.TRUE.
70    CONTINUE
      KNOW=((A(P11,3).EQ.0.0).AND.(A(P2,4).EQ.0.0))
      UNKNOW=.NOT.(KNOW)
      IF (ACTUAL.AND.KNOW) FLAG=1.0
      IF (ACTUAL.AND.UNKNOW) FLAG=2.0
      IF (ESTIM.AND.KNOW) FLAG=3.0
      IF (ESTIM.AND.UNKNOW) FLAG=4.0
C     LOAD DATA FOR CURRENT OBSERVATION
      B(NOB,1)=HHIDNO
      B(NOB,2)=FLAG
      B(NOB,3)=FLOAT(NB)
      B(NOB,4)=FLOAT(NE)
      B(NOB,5)=QUAN
      B(NOB,6)=EXPEN
      B(NOB,7)=HDD65
      B(NOB,8)=CDD65
      B(NOB,9)=HDD75
      B(NOB,10)=CDD75
C     EXIT LOOP FOR CURRENT OBSERVATION
      P1=P2
      NOB=NOB+1
      IF (NOB.EQ.16) GO TO 150
      GO TO 20
100   CONTINUE
      XACTVE=0.
      WRITE(17,101) XACTVE
101   FORMAT(F3.0)
      GO TO 5
150   XNOB=FLOAT((NOB-1))
      DO 120 J=1,15
      IF (B(J,2).NE.0.0) GO TO 121
      SUMO=SUMO+1.
      GO TO 120
121   IF (B(J,2).NE.1.0) GO TO 122
      SUM1=SUM1+1.
      GO TO 120
122   IF (B(J,2).NE.2.0) GO TO 123
      SUM2=SUM2+1.
      GO TO 120
```

```
123   IF (B(J,2).NE.3.0) GO TO 124                                      OBS01110
      SUM3=SUM3+1.                                                      OBS01120
      GO TO 120                                                         OBS01130
124   IF (B(J,2).NE.4.0) GO TO 120                                      OBS01140
      SUM4=SUM4+1.                                                      OBS01150
120   CONTINUE                                                          OBS01160
      DO 130 M1=1,15                                                    OBS01170
      WRITE(2,200) (B(M1,M2),M2=1,10)                                   OBS01180
130   CONTINUE                                                          OBS01190
200   FORMAT(F6.0,1X,F3.0,1X,2(F6.0,1X),2(F10.2,1X),4(F6.0,1X),5X)      OBS01200
      SUMOB=SUMOB+XNOB                                                  OBS01210
      XACTVE=1.0                                                        OBS01220
      WRITE(17,101) XACTVE                                              OBS01230
      WRITE(3,300) XNOB                                                 OBS01240
300   FORMAT(5X,F10.2,65X)                                              OBS01250
      GO TO 5                                                           OBS01260
7     READ(1,6)                                                         OBS01270
5     CONTINUE                                                          OBS01280
      WRITE(5,499)                                                      OBS01290
499   FORMAT(80X)                                                       OBS01300
      WRITE(5,500) SUMO,SUM1,SUM2,SUM3,SUM4,SUMOB                       OBS01310
500   FORMAT(1X,6(F9.0,1X),19X)                                         OBS01320
      STOP                                                              OBS01330
      END                                                               OBS01340
```

2. Reprocessing the Raw Natural Gas Billing Data

```
      INTEGER P11,P1,P2
      LOGICAL ACTUAL,ESTIM,KNOW,UNKNOW
      DIMENSION A(20,14),METER(20),B(15,10)
      SUMO=0.0
      SUM1=0.0
      SUM2=0.0
      SUM3=0.0
      SUM4=0.0
      SUMOB=0.0
      DO 5 K=1,3842
      READ(4,15) SAMPLE
15    FORMAT(F3.1)
C     THIS CODE IS SPECIFIC TO THE GAS VERSION OF OBSER ONLY. IT SHOULD
C     NOT APPEAR IN THE ELEC VERSION. THIS SECTION OF CODE ALLOWS THE
C     PROGRAM TO DISREGARD THE FIRST FIVE OBSERVATIONS IN THE GAS
C     BILLING DATA. THESE OBSERVATIONS APPEAR IN THE ELECTRICITY DATA
C     BUT DO NOT APPEAR IN THE GAS DATA.
      IF ((K.LE.5).AND.(SAMPLE.EQ.1.0)) GO TO 100
      IF ((K.LE.5).AND.(SAMPLE.EQ.0.0)) GO TO 5
      IF ((K.GT.5).AND.(SAMPLE.EQ.0.0)) GO TO 7
C     IN THE ELECTRICITY VERSION OF OBSER FORTRAN THESE LINES ARE
C     REPLACED WITH *** IF (SAMPL.EQ.0.0) GO TO 7 ***
C     END OF SPECIFIC CODE
      DO 10 I=1,15
      DO 10 J=1,10
10    B(I,J)=0.0
      READ(1,6) HHIDNO,NBILLS,((A(I,J),J=1,14),I=1,20),(METER(I),
     1 I=1,20)
6     FORMAT(F4.0,6X,I2,20(4X,F7.1,3X,F5.2,2F1.0,6F2.0,63X,2F5.0,
     1 40X,2F5.0,20X),20I1)
      IF(NBILLS.EQ.99) GO TO 100
      P1=0
      P2=0
      NOB=1
20    IF (P2.EQ.NBILLS) GO TO 150
      P2=P2+1
      IF ((METER(P2).NE.1).AND.(METER(P2).NE.8)) GO TO 20
      QUAN=0.0
      EXPEN=0.0
      HDD65=0.0
      CDD65=0.0
      HDD75=0.0
      CDD75=0.0
      P11=P1+1
      DO 30 J=P11,P2
C     ACCUMULATE EXPEN,QUAN,HDD,CDD
      QUAN=QUAN+A(J,1)
      HDD65=HDD65+A(J,11)
      CDD65=CDD65+A(J,12)
      HDD75=HDD75+A(J,13)
      CDD75=CDD75+A(J,14)
      EXPEN=EXPEN+A(J,2)
30    CONTINUE
C     CONVERT BEGINING AND ENDING DATES TO SUMMARY NUMBER
      BY=A(P11,5)+1900.0
```

```
      BM=A(P11,6)
      BD=A(P11,7)
      EY=A(P2,8)+1900.0
      EM=A(P2,9)
      ED=A(P2,10)
C     CALCULATION FOR BEGINING PERIOD
      IF (BM.GT.2.0) GO TO 31
      NB=INT(365.25*(BY-1.0))+INT(30.6*(BM+13.0))+INT(BD)-621049
      GO TO 32
31    NB=INT(365.25*BY)+INT(30.6*(BM+1.0))+INT(BD)-621049
32    CONTINUE
C     CALCULATION FOR ENDING PERIOD
      IF (EM.GT.2.0) GO TO 33
      NE=INT(365.25*(EY-1.0))+INT(30.6*(EM+13.0))+INT(ED)-621049
      GO TO 34
33    NE=INT(365.25*EY)+INT(30.6*(EM+1.0))+INT(ED)-621049
34    CONTINUE
      N1178=101479
      NB=NB-N1178
      NE=NE-N1178
C     CALCULATION OF FLAG CODE
      IF (P1.EQ.0) GO TO 60
      ACTUAL=((METER(P1).EQ.1).AND.(METER(P2).EQ.1))
      ESTIM=.NOT.(ACTUAL)
      GO TO 70
60    ESTIM=.TRUE.
70    CONTINUE
      KNOW=((A(P11,3).EQ.0.0).AND.(A(P2,4).EQ.0.0))
      UNKNOW=.NOT.(KNOW)
      IF (ACTUAL.AND.KNOW) FLAG=1.0
      IF (ACTUAL.AND.UNKNOW) FLAG=2.0
      IF (ESTIM.AND.KNOW) FLAG=3.0
      IF (ESTIM.AND.UNKNOW) FLAG=4.0
C     LOAD DATA FOR CURRENT OBSERVATION
      B(NOB,1)=HHIDNO
      B(NOB,2)=FLAG
      B(NOB,3)=FLOAT(NB)
      B(NOB,4)=FLOAT(NE)
      B(NOB,5)=QUAN
      B(NOB,6)=EXPEN
      B(NOB,7)=HDD65
      B(NOB,8)=CDD65
      B(NOB,9)=HDD75
      B(NOB,10)=CDD75
C     EXIT LOOP FOR CURRENT OBSERVATION
      P1=P2
      NOB=NOB+1
      IF (NOB.EQ.16) GO TO 150
      GO TO 20
100   CONTINUE
      XACTVE=0.
      WRITE(17,101) XACTVE
101   FORMAT(F3.0)
      GO TO 5
150   XNOB=FLOAT((NOB-1))
```

```
      DO 120 J=1,15                                                      OBS01110
      IF (B(J,2).NE.0.0) GO TO 121                                       OBS01120
      SUMO=SUMO+1.                                                       OBS01130
      GO TO 120                                                          OBS01140
121   IF (B(J,2).NE.1.0) GO TO 122                                       OBS01150
      SUM1=SUM1+1.                                                       OBS01160
      GO TO 120                                                          OBS01170
122   IF (B(J,2).NE.2.0) GO TO 123                                       OBS01180
      SUM2=SUM2+1.                                                       OBS01190
      GO TO 120                                                          OBS01200
123   IF (B(J,2).NE.3.0) GO TO 124                                       OBS01210
      SUM3=SUM3+1.                                                       OBS01220
      GO TO 120                                                          OBS01230
124   IF (B(J,2).NE.4.0) GO TO 120                                       OBS01240
      SUM4=SUM4+1.                                                       OBS01250
120   CONTINUE                                                           OBS01260
      DO 130 M1=1,15                                                     OBS01270
      WRITE(2,200) (B(M1,M2),M2=1,10)                                    OBS01280
130   CONTINUE                                                           OBS01290
200   FORMAT(F6.0,1X,F3.0,1X,2(F6.0,1X),2(F10.2,1X),4(F6.0,1X),5X)       OBS01300
      SUMOB=SUMOB+XNOB                                                   OBS01310
      XACTVE=1.0                                                         OBS01320
      WRITE(17,101) XACTVE                                               OBS01330
      WRITE(3,300) XNOB                                                  OBS01340
300   FORMAT(5X,F10.2,65X)                                               OBS01350
      GO TO 5                                                            OBS01360
7     READ(1,6)                                                          OBS01370
5     CONTINUE                                                           OBS01380
      WRITE(5,499)                                                       OBS01390
499   FORMAT(80X)                                                        OBS01400
      WRITE(5,500) SUMO,SUM1,SUM2,SUM3,SUM4,SUMOB                        OBS01410
500   FORMAT(1X,6(F9.0,1X),19X)                                          OBS01420
      STOP                                                               OBS01430
      END                                                                OBS01440
```

3. Determination of Seasonal Marginal Prices from Electricity Billing Data

```
      LOGICAL ELEC,NELEC,WIN,SUM,OFF
      DIMENSION A(4,20),QUAN(5,2000),EXPEN(5,2000),N(5),XMPR(5)
      DO 1 L=1,103
      READ(1,20) NOBPSU
20    FORMAT(14X,I3,63X)
      DO 5 I=1,5
      DO 4 J=1,20
      QUAN(I,J)=0.0
      EXPEN(I,J)=0.0
4     CONTINUE
5     N(I)=0
      DO 200 KKK=1,NOBPSU
      READ(2,30) AELEC
30    FORMAT(3X,F2.0,75X)
      ELEC=(AELEC.EQ.1.0)
      NELEC=.NOT.(ELEC)
      READ(3,40) NBILLS,((A(M1,M2),M1=1,4),M2=1,20)
40    FORMAT(10X,I2,20(4X,F7.1,3X,F5.2,4X,F2.0,4X,F2.0,145X),20X)
      IF (NBILLS.EQ.99) GO TO 200
      DO 300 J=1,NBILLS
      IF (A(1,J).EQ.0.) GO TO 300
      IF (A(2,J).EQ.0.) GO TO 300
      IF(A(1,J).GE.3000.) GO TO 300
      IF(A(2,J).GE.995.) GO TO 300
      WIN=((A(3,J).EQ.1.).OR.(A(4,J).EQ.1.))
      SUM=((A(3,J).EQ.7.).OR.(A(4,J).EQ.7.))
      OFF=((A(3,J).EQ.4.).OR.(A(3,J).EQ.10.).OR.(A(3,J).EQ.9.).OR.
     1 (A(4,J).EQ.4.).OR.(A(4,J).EQ.10.).OR.(A(4,J).EQ.11.))
      IF (ELEC) GO TO 50
      IF (NELEC.AND.WIN) GO TO 60
      IF (NELEC.AND.SUM) GO TO 70
      IF (NELEC.AND.OFF) GO TO 80
      GO TO 90
50    L1=1
      GO TO 125
60    L1=2
      GO TO 125
70    L1=3
      GO TO 125
80    L1=4
      GO TO 125
90    L1=5
125   N(L1)=N(L1)+1
      QUAN(L1,N(L1))=A(1,J)
      EXPEN(L1,N(L1))=A(2,J)
300   CONTINUE
200   CONTINUE
C     RUN REGRESSIONS AND STORE XMPR
      DO 400 I=1,4
      IF (N(I).LT.10) GO TO 377
      SUMX=0.
      SUMY=0.
      SXY=0.
      SXX=0.
      NNN=N(I)
```

```
      DO 500 J=1,NNN
      SUMX=SUMX+QUAN(I,J)
      SUMY=SUMY+EXPEN(I,J)
500   CONTINUE
      XBAR=SUMX/FLOAT(N(I))
      YBAR=SUMY/FLOAT(N(I))
      DO 510 L2=1,NNN
      SXY=SXY+((QUAN(I,L2)-XBAR)*(EXPEN(I,L2)-YBAR))
      SXX=SXX+((QUAN(I,L2)-XBAR)*(QUAN(I,L2)-XBAR))
510   CONTINUE
      IF (SXX.EQ.O.) GO TO 378
      XMPR(I)=SXY/SXX
      IF (XMPR(I).LE.O) GO TO 375
      IF(YBAR-(XBAR*XMPR(I)).LT.O.) GO TO 376
      GO TO 400
375   CONTINUE
C     NEGATIVE MARGINAL PRICE
      XMPR(I)=-99.
      GO TO 400
376   CONTINUE
C     NEGATIVE INTERCEPT
      XMPR(I)=-99.
      GO TO 400
377   CONTINUE
C     TOO FEW OBSERVATIONS
      XMPR(I)=-99.
      GO TO 400
378   CONTINUE
C     SINGULAR MATRIX
      XMPR(I)=-99.
400   CONTINUE
      SXX=O.
      SUMX=O.
      SXY=O.
      SUMY=O.
      NNN=O.
      DO 411 I=1,5
411   NNN=NNN+N(I)
      IF (NNN.LT.15) GO TO 425
      DO 413 I=1,5
      DO 413 J=1,NNN
      SUMX=SUMX+QUAN(I,J)
      SUMY=SUMY+EXPEN(I,J)
413   CONTINUE
      XBAR=SUMX/NNN
      YBAR=SUMY/NNN
      DO 415 I=1,5
      DO 415 L2=1,NNN
      SXY=SXY+((QUAN(I,L2)-XBAR)*(EXPEN(I,L2)-YBAR))
      SXX=SXX+((QUAN(I,L2)-XBAR)*(QUAN(I,L2)-XBAR))
415   CONTINUE
      IF (SXX.EQ.O.) GO TO 425
      XMPR(5)=SXY/SXX
      IF (XMPR(5).LE.O) GO TO 425
      IF (YBAR-(XBAR*XMPR(5)).LT.O.) GO TO 425
```

```
      GO TO 450
425   XMPR(5)=-99.
450   CONTINUE
C     SET THE MARGINAL PRICES **************************************
C     WE FIRST CHECK TO SEE IF THE MARGINAL PRICE USING ALL
C     OBSERVATIONS HAS BEEN SET.  IF THIS MARGINAL PRICE IS SET WE
C     LEAVE IT ALONE.  IF NOT, THE OVERALL RATE IS SET TO THE FIRST
C     VALID RATE STARTING WITH NON-ELEC. OFF SEASON, THEN NON-ELEC.
C     SUMMER, NON-ELEC. WINTER AND FINALLY THE ALL ELEC. RATE.
C     IN THE NEXT STEP WE RESET THE NON ELEC. OFF SEASON RATE TO
C     THE OVERALL RATE IF THE FORMER IS INVALID.  THE SUMMER AND
C     WINTER NON-ELEC. RATES ARE THEN COMPARED TO THE NON-ELEC.
C     OFF SEASON RATE FOR PEAKING. THAT IS, IF THESE RATES ARE HIGHER
C     THEY ARE LEFT UNCHANGED; IF THEY ARE LOWER THEY ARE SET TO THE
C     NON-ELEC. OFF-SEASON RATE.  FINALLY, THE ALL ELEC. RATE IS
C     CHECKED AND RESET TO THE NON-ELEC. OFF SEASON ONLY IF IT IS IN-
C     VALID.
      IF (XMPR(5).EQ.-99.) XMPR(5)=XMPR(4)
      IF (XMPR(5).EQ.-99.) XMPR(5)=XMPR(3)
      IF (XMPR(5).EQ.-99.) XMPR(5)=XMPR(2)
      IF (XMPR(5).EQ.-99.) XMPR(5)=XMPR(1)
      IF (XMPR(4).EQ.-99.) XMPR(4)=XMPR(5)
      IF (XMPR(2).LT.XMPR(4)) XMPR(2)=XMPR(4)
      IF (XMPR(3).LT.XMPR(4)) XMPR(3)=XMPR(4)
      IF (XMPR(1).EQ.-99.) XMPR(1)=XMPR(4)
C     WRITE THE XMPR'S
      WRITE(4,600) L,(XMPR(I),I=1,5)
600   FORMAT(I4,1X,5(E13.6,2X))
      WRITE(5,700) L,(N(I),I=1,5)
700   FORMAT(6(1X,I9),20X)
1     CONTINUE
      STOP
      END
```

(3a) Marginal Prices by Primary Sampling Unit

PSU	AMPE78	WMPE78	SMPE78	OSMPE78	AVMPE78
1	0.615394E-01	0.615394E-01	0.623080E-01	0.615394E-01	0.678879E-01
2	0.695582E-01	0.695582E-01	0.695582E-01	0.695582E-01	0.648289E-01
3	0.918956E-01	0.918956E-01	0.101133E+00	0.918956E-01	0.867271E-01
4	0.361693E-01	0.563768E-01	0.563768E-01	0.563768E-01	0.559760E-01
5	0.381916E-01	0.381916E-01	0.381916E-01	0.381916E-01	0.444611E-01
6	0.442855E-01	0.442855E-01	0.535490E-01	0.442855E-01	0.509069E-01
7	0.929877E-01	0.929877E-01	0.100335E+00	0.929877E-01	0.675159E-01
8	0.432652E-01	0.432652E-01	0.432652E-01	0.432652E-01	0.564484E-01
9	0.451717E-01	0.451717E-01	0.494833E-01	0.451717E-01	0.575215E-01
10	0.465870E-01	0.465870E-01	0.535044E-01	0.465870E-01	0.519443E-01
11	0.974315E-01	0.974315E-01	0.977520E-01	0.974315E-01	0.540452E-01
12	0.604029E-01	0.604029E-01	0.680853E-01	0.604029E-01	0.604029E-01
13	0.969005E-01	0.969005E-01	0.990688E-01	0.969005E-01	0.656430E-01
14	0.527022E-01	0.527022E-01	0.527022E-01	0.527022E-01	0.527022E-01
15	0.669697E-01	0.707266E-01	0.770710E-01	0.669697E-01	0.563852E-01
16	0.376967E-01	0.443928E-01	0.443928E-01	0.443928E-01	0.477499E-01
17	0.419157E-01	0.443867E-01	0.419157E-01	0.419157E-01	0.516058E-01
18	0.329057E-01	0.368616E-01	0.368688E-01	0.368616E-01	0.396048E-01
19	0.463114E-01	0.463114E-01	0.464788E-01	0.463114E-01	0.463732E-01
20	0.739906E-01	0.573830E-01	0.638053E-01	0.573830E-01	0.474420E-01
21	0.346334E-01	0.347692E-01	0.346392E-01	0.346334E-01	0.424144E-01
22	0.325585E-01	0.368810E-01	0.368810E-01	0.368810E-01	0.435294E-01
23	0.364602E-01	0.364602E-01	0.441954E-01	0.364602E-01	0.456277E-01
24	0.374158E-01	0.402360E-01	0.402360E-01	0.402360E-01	0.440812E-01
25	0.392836E-01	0.407130E-01	0.534687E-01	0.407130E-01	0.477073E-01
26	0.534913E-01	0.534913E-01	0.534913E-01	0.534913E-01	0.454764E-01
27	0.491435E-01	0.496057E-01	0.491435E-01	0.491435E-01	0.471774E-01
28	0.185823E-01	0.422624E-01	0.440227E-01	0.422624E-01	0.445839E-01
29	0.361029E-01	0.361029E-01	0.453730E-01	0.361029E-01	0.423141E-01
30	0.434595E-01	0.434595E-01	0.452091E-01	0.434595E-01	0.451063E-01
31	0.520863E-01	0.520863E-01	0.520863E-01	0.520863E-01	0.450764E-01
32	0.386342E-01	0.386342E-01	0.386342E-01	0.386342E-01	0.440375E-01
33	0.482186E-01	0.482186E-01	0.482186E-01	0.482186E-01	0.454210E-01
34	0.412887E-01	0.412887E-01	0.462227E-01	0.412887E-01	0.442980E-01
35	0.469680E-01	0.469680E-01	0.469680E-01	0.469680E-01	0.463062E-01
36	0.503137E-01	0.503137E-01	0.503137E-01	0.503137E-01	0.460707E-01
37	0.324032E-01	0.363827E-01	0.363827E-01	0.363827E-01	0.434091E-01
38	0.290702E-01	0.456641E-01	0.551423E-01	0.456641E-01	0.467872E-01
39	0.366758E-01	0.389163E-01	0.437497E-01	0.389163E-01	0.448331E-01
40	0.390315E-01	0.390315E-01	0.477616E-01	0.390315E-01	0.440697E-01
41	0.459190E-01	0.459190E-01	0.480469E-01	0.459190E-01	0.469977E-01
42	0.463545E-01	0.525858E-01	0.525858E-01	0.525858E-01	0.500885E-01
43	0.432786E-01	0.434006E-01	0.433234E-01	0.406584E-01	0.449710E-01
44	0.308716E-01	0.340208E-01	0.308716E-01	0.308716E-01	0.437011E-01
45	0.301146E-01	0.423565E-01	0.430748E-01	0.423565E-01	0.439700E-01
46	0.401177E-01	0.401177E-01	0.401177E-01	0.401177E-01	0.439087E-01
47	0.283400E-01	0.355772E-01	0.355772E-01	0.355772E-01	0.444195E-01
48	0.356759E-01	0.377090E-01	0.377090E-01	0.377090E-01	0.437510E-01
49	0.450659E-01	0.450659E-01	0.450659E-01	0.450659E-01	0.450659E-01
50	0.411910E-01	0.411910E-01	0.411910E-01	0.411910E-01	0.463053E-01
51	0.450371E-01	0.450371E-01	0.450371E-01	0.450371E-01	0.469661E-01
52	0.395219E-01	0.439201E-01	0.439201E-01	0.439201E-01	0.471055E-01
53	0.330019E-01	0.418738E-01	0.431414E-01	0.330019E-01	0.470479E-01
54	0.426420E-01	0.426420E-01	0.426420E-01	0.426420E-01	0.465483E-01
55	0.363876E-01	0.359965E-01	0.359965E-01	0.359965E-01	0.430574E-01

56	0.323260E-01	0.413572E-01	0.454232E-01	0.413572E-01	0.459339E-01
57	0.412232E-01	0.412232E-01	0.513541E-01	0.412232E-01	0.475107E-01
58	0.360636E-01	0.410820E-01	0.431842E-01	0.410820E-01	0.465003E-01
59	0.315132E-01	0.310826E-01	0.345924E-01	0.308467E-01	0.458806E-01
60	0.273507E-01	0.288042E-01	0.288042E-01	0.288042E-01	0.373267E-01
61	0.329603E-01	0.357055E-01	0.351842E-01	0.351842E-01	0.398837E-01
62	0.224490E-01	0.257686E-01	0.339717E-01	0.257686E-01	0.380255E-01
63	0.277734E-01	0.356741E-01	0.361906E-01	0.356741E-01	0.409918E-01
64	0.289309E-01	0.282682E-01	0.282682E-01	0.282682E-01	0.354884E-01
65	0.314629E-01	0.367810E-01	0.367810E-01	0.367810E-01	0.367810E-01
66	0.312201E-01	0.355750E-01	0.393413E-01	0.355750E-01	0.383042E-01
67	0.297718E-01	0.316213E-01	0.339345E-01	0.316213E-01	0.381194E-01
68	0.315099E-01	0.363177E-01	0.363177E-01	0.363177E-01	0.382188E-01
69	0.302503E-01	0.371938E-01	0.371938E-01	0.371938E-01	0.390650E-01
70	0.313468E-01	0.357129E-01	0.358395E-01	0.357129E-01	0.389631E-01
71	0.311292E-01	0.356104E-01	0.356104E-01	0.356104E-01	0.377790E-01
72	0.410165E-01	0.410165E-01	0.420515E-01	0.410165E-01	0.393671E-01
73	0.340119E-01	0.244420E-01	0.259682E-01	0.244420E-01	0.360207E-01
74	0.359174E-01	0.430672E-01	0.441249E-01	0.425731E-01	0.414345E-01
75	0.341577E-01	0.365681E-01	0.511523E-01	0.365681E-01	0.414201E-01
76	0.304010E-01	0.286004E-01	0.284403E-01	0.284403E-01	0.361560E-01
77	0.303919E-01	0.354836E-01	0.398212E-01	0.354836E-01	0.369851E-01
78	0.375651E-01	0.350035E-01	0.354099E-01	0.350035E-01	0.372527E-01
79	0.267736E-01	0.312942E-01	0.294368E-01	0.252713E-01	0.358435E-01
80	0.452574E-01	0.409030E-01	0.409030E-01	0.409030E-01	0.409030E-01
81	0.249419E-01	0.316531E-01	0.355365E-01	0.311559E-01	0.388244E-01
82	0.383938E-01	0.415701E-01	0.415701E-01	0.415701E-01	0.429540E-01
83	0.424909E-01	0.458561E-01	0.460792E-01	0.458561E-01	0.460569E-01
84	0.290708E-01	0.318975E-01	0.318975E-01	0.318975E-01	0.373043E-01
85	0.370014E-01	0.370014E-01	0.397317E-01	0.370014E-01	0.365076E-01
86	0.390098E-01	0.390098E-01	0.390098E-01	0.390098E-01	0.420379E-01
87	0.237153E-01	0.237153E-01	0.237153E-01	0.237153E-01	0.237153E-01
88	0.357874E-01	0.357874E-01	0.357874E-01	0.357874E-01	0.357874E-01
89	0.472943E-01	0.472943E-01	0.472943E-01	0.472943E-01	0.402931E-01
90	0.512574E-01	0.512574E-01	0.547953E-01	0.512574E-01	0.388125E-01
91	0.317062E-01	0.317062E-01	0.317062E-01	0.317062E-01	0.354506E-01
92	0.423982E-01	0.499166E-01	0.516751E-01	0.499166E-01	0.368473E-01
93	0.492906E-01	0.464387E-01	0.464387E-01	0.464387E-01	0.409899E-01
94	0.194822E-01	0.194822E-01	0.211409E-01	0.194822E-01	0.388453E-01
95	0.343593E-01	0.343593E-01	0.343593E-01	0.343593E-01	0.343593E-01
96	0.252773E-01	0.250081E-01	0.243841E-01	0.242158E-01	0.311832E-01
97	0.122910E-01	0.106477E-01	0.102338E-01	0.100343E-01	0.249291E-01
98	0.283751E-01	0.369910E-01	0.369910E-01	0.369910E-01	0.326029E-01
99	0.131349E-01	0.124846E-01	0.127262E-01	0.117981E-01	0.246782E-01
100	0.430073E-01	0.378893E-01	0.378893E-01	0.378893E-01	0.292362E-01
101	0.430062E-01	0.430062E-01	0.430062E-01	0.430062E-01	0.301974E-01
102	0.119395E-01	0.106878E-01	0.106878E-01	0.106878E-01	0.275485E-01
103	0.104731E-01	0.181161E-01	0.176220E-01	0.151837E-01	0.272085E-01

Observations Processed

(3b) Record of Observations Processed

PSU

1	0	52	52	104	106
2	3	54	33	40	36
3	0	44	24	57	71
4	14	96	48	122	111
5	6	76	40	102	102
6	0	130	71	169	173
7	0	60	38	82	102
8	1	30	14	57	48
9	7	43	27	47	48
10	0	71	41	102	96
11	0	27	20	42	50
12	0	22	11	97	61
13	0	25	12	30	20
14	7	81	20	86	134
15	0	38	26	36	35
16	16	108	36	128	138
17	0	65	29	45	61
18	198	118	56	148	183
19	0	91	38	112	126
20	12	43	40	82	82
21	0	75	34	101	75
22	55	117	60	156	207
23	0	90	31	98	152
24	27	51	36	102	85
25	34	56	46	85	120
26	8	147	65	151	175
27	0	58	32	64	39
28	12	76	48	130	126
29	0	88	46	117	117
30	0	78	44	95	110
31	0	77	38	77	98
32	0	33	33	35	22
33	9	74	35	95	54
34	0	113	94	256	201
35	0	74	27	133	61
36	0	84	31	90	110
37	19	90	50	109	134
38	41	96	54	135	138
39	47	91	42	103	112
40	0	130	64	171	185
41	16	79	37	118	101
42	16	134	74	162	182
43	34	127	68	149	150
44	15	82	58	137	164
45	62	152	83	202	211
46	8	187	98	224	241
47	23	68	34	103	109
48	175	127	72	185	192
49	6	71	40	132	104
50	0	83	46	132	145
51	0	86	63	182	161
52	53	56	44	115	108
53	0	70	42	119	112
54	68	27	23	55	53
55	30	127	57	160	182

56	49	94	45	93	114
57	0	150	70	192	189
58	102	40	22	61	64
59	123	41	15	32	57
60	84	170	83	172	193
61	143	85	61	150	140
62	49	70	48	119	109
63	15	100	50	125	124
64	264	18	16	34	40
65	277	3	2	5	5
66	138	119	63	181	182
67	71	72	55	146	134
68	156	62	45	104	113
69	226	106	54	127	145
70	60	66	31	83	83
71	207	82	54	149	139
72	0	81	51	127	136
73	64	78	76	194	149
74	117	124	48	154	154
75	68	67	31	78	80
76	227	39	28	60	63
77	14	131	66	175	172
78	15	55	29	62	77
79	45	50	30	74	64
80	215	8	4	8	10
81	35	119	55	163	186
82	116	119	64	186	164
83	230	136	61	163	148
84	237	53	28	72	67
85	0	126	58	167	156
86	0	6	0	12	8
87	0	7	4	7	6
88	0	94	45	116	115
89	0	10	0	17	0
90	8	37	16	13	8
91	0	73	34	92	92
92	32	148	74	185	184
93	20	83	38	93	85
94	4	18	13	40	27
95	33	146	68	187	192
96	208	74	43	107	108
97	182	32	17	27	30
98	91	50	28	59	68
99	157	57	28	72	78
100	16	131	69	170	170
101	8	158	88	195	232
102	69	0	0	39	22
103	120	28	18	44	41

4. Determination of the Marginal Price of Natural Gas

```
      DIMENSION A(4,20),QUAN(2000),EXPEN(2000)
      DO 1 L=1,103
      READ(1,20) NOBPSU
20    FORMAT(14X,I3,63X)
      N=0
C     THIS LINE CORRECTS FOR THE FACT THAT THE FIRST 5
C     LINES OF THE GAS BILLING DATA ARE MISSING (LRECL 3552)
      IF (L.EQ.1) NOBPSU=NOBPSU-5
      DO 200 KKK=1,NOBPSU
      READ(3,40) NBILLS,((A(M1,M2),M1=1,4),M2=1,20)
40    FORMAT(10X,I2,20(4X,F7.1,3X,F5.2,4X,F2.0,4X,F2.0,145X),20X)
      IF (NBILLS.EQ.99) GO TO 200
      DO 300 J=1,NBILLS
      IF (A(1,J).EQ.0.) GO TO 300
      IF (A(2,J).EQ.0.) GO TO 300
      N=N+1
      QUAN(N)=A(1,J)
      EXPEN(N)=A(2,J)
300   CONTINUE
200   CONTINUE
C     RUN REGRESSIONS AND STORE XMPR
      IF (N.LT.10) GO TO 375
      SUMX=0.
      SUMY=0.
      SXY=0.
      SXX=0.
      DO 500 J=1,N
      SUMX=SUMX+QUAN(J)
      SUMY=SUMY+EXPEN(J)
500   CONTINUE
      XBAR=SUMX/FLOAT(N)
      YBAR=SUMY/FLOAT(N)
      DO 510 L2=1,N
      SXY=SXY+((QUAN(L2)-XBAR)*(EXPEN(L2)-YBAR))
      SXX=SXX+((QUAN(L2)-XBAR)*(QUAN(L2)-XBAR))
510   CONTINUE
      IF (SXX.EQ.0.) GO TO 57
      XMPR=SXY/SXX
      GO TO 58
57    XMPR=0.0
58    IF (SUMX.EQ.0.0)  GO TO 59
      XAPR=SUMY/SUMX
      GO TO 60
59    XAPR=0.0
60    CONTINUE
      IF (XMPR.LE.0) XMPR=XAPR
      IF (XMPR.LE.0.10) GO TO 375
      GO TO 400
375   CONTINUE
      XMPR=0.0
400   CONTINUE
C     WRITE THE XMPR'S
      WRITE(4,600) L,XMPR
600   FORMAT(I4,1X,E13.6,2X,60X)
      WRITE(5,700) L,N
```

```
700     FORMAT(2(1X,I9),60X)
1       CONTINUE
        STOP
        END
```

MPG78

(4a) Marginal Price by Primary Sampling Unit

PSU	
1	0.341614E+00
2	0.275102E+00
3	0.0
4	0.313200E+00
5	0.320076E+00
6	0.319346E+00
7	0.419931E+00
8	0.276353E+00
9	0.327669E+00
10	0.315126E+00
11	0.0
12	0.466833E+00
13	0.431005E+00
14	0.243124E+00
15	0.339862E+00
16	0.246171E+00
17	0.226454E+00
18	0.336241E+00
19	0.289909E+00
20	0.346979E+00
21	0.270784E+00
22	0.314686E+00
23	0.284562E+00
24	0.252753E+00
25	0.292946E+00
26	0.0
27	0.289750E+00
28	0.272271E+00
29	0.264909E+00
30	0.216798E+00
31	0.223869E+00
32	0.257374E+00
33	0.235958E+00
34	0.255409E+00
35	0.231851E+00
36	0.212041E+00
37	0.219284E+00
38	0.207306E+00
39	0.266101E+00
40	0.158331E+00
41	0.169197E+00
42	0.169310E+00
43	0.155860E+00
44	0.236991E+00
45	0.249219E+00
46	0.215887E+00
47	0.242971E+00
48	0.262884E+00
49	0.231630E+00
50	0.209319E+00
51	0.208083E+00
52	0.240105E+00
53	0.248587E+00
54	0.361061E+00
55	0.266388E+00

56	0.335851E+00
57	0.318555E+00
58	0.220303E+00
59	0.286067E+00
60	0.238971E+00
61	0.0
62	0.242265E+00
63	0.157452E+00
64	0.210873E+00
65	0.0
66	0.274921E+00
67	0.198469E+00
68	0.283167E+00
69	0.270030E+00
70	0.273660E+00
71	0.344307E+00
72	0.240427E+00
73	0.154741E+00
74	0.244279E+00
75	0.308869E+00
76	0.218792E+00
77	0.223122E+00
78	0.207089E+00
79	0.235208E+00
80	0.0
81	0.295990E+00
82	0.239429E+00
83	0.315688E+00
84	0.216590E+00
85	0.228616E+00
86	0.183363E+00
87	0.0
88	0.214218E+00
89	0.200969E+00
90	0.193940E+00
91	0.201406E+00
92	0.204096E+00
93	0.179709E+00
94	0.219525E+00
95	0.222310E+00
96	0.343881E+00
97	0.314587E+00
98	0.0
99	0.352934E+00
100	0.170713E+00
101	0.162105E+00
102	0.0
103	0.322192E+00

Observations Processed

(4b) Record of Observations Processed

PSU	
1	267
2	111
3	135
4	286
5	172
6	203
7	193
8	318
9	112
10	134
11	0
12	78
13	72
14	253
15	115
16	306
17	284
18	229
19	185
20	178
21	267
22	342
23	160
24	202
25	175
26	601
27	301
28	335
29	328
30	277
31	336
32	126
33	201
34	598
35	204
36	282
37	413
38	414
39	321
40	485
41	420
42	545
43	462
44	92
45	412
46	603
47	215
48	216
49	286
50	374
51	311
52	248
53	347
54	111
55	295

56	42
57	431
58	121
59	176
60	417
61	0
62	120
63	267
64	44
65	5
66	96
67	350
68	111
69	15
70	75
71	129
72	386
73	325
74	335
75	16
76	35
77	508
78	251
79	176
80	0
81	528
82	44
83	79
84	23
85	212
86	233
87	0
88	370
89	350
90	426
91	284
92	590
93	543
94	112
95	543
96	190
97	37
98	0
99	87
100	526
101	442
102	0
103	34

5. Thermal Load Model for Processed Billing Data

```
      DO 10 I=1,1144
      READ(1,100) XXLAM,W1A,W1,W2,W3A,W3,Q0,Q1,Q2,Q3,Q4,Q5,
     & Q6,Q7
100   FORMAT(14E15.8)
      DO 20 J=1,15
      READ(2,200) HHIDNO,FLAG,START,END,QUAN,EXPEN,HDD65,CDD65
     & ,HDD75,CDD75
200   FORMAT(F6.0,1X,F3.0,1X,2(F6.0,1X),2(F10.2,1X),4(F6.0,1X),5X)
      IF (FLAG.EQ.0.0) GO TO 30
      IF (START.EQ.END) GO TO 30
      XDAYS=END-START
      HDD65=HDD65/XDAYS
      HDD75=HDD75/XDAYS
      CDD65=CDD65/XDAYS
      CDD75=CDD75/XDAYS
      IF ((HDD75.EQ.0.0).OR.(CDD65.EQ.0.0)) GO TO 250
      CALL COEF(HDD75,CDD65,APAR,BPAR)
C     CHECK ESTIMATED TEMPERATURE DISTRIBUTION THROUGH BPAR
      IF(BPAR.EQ.1.0) GO TO 250
      TTT=75.
      CALL ACC(APAR,BPAR,Q0,Q1,Q2,Q3,Q4,Q5,Q6,Q7,T1,T2,T3,TTT,ACUEC)
      TTT=74.
      CALL ACC(APAR,BPAR,Q0,Q1,Q2,Q3,Q4,Q5,Q6,Q7,T1,T2,T3,TTT,DACUEC)
      DACUEC=DACUEC-ACUEC
      T6=70.0+XXLAM
      XLAM=APAR+BPAR*T6
      P6=1.0/(1.0+EXP(-XLAM))
      TEMP=-APAR/BPAR
      IF ((P6.LE.0.0001).OR.((1.0-P6).LE.0.0001)) GO TO 255
      CALL HEAT(BPAR,XLAM,W3A,W1A,W2,SHUEC)
      XLAM=XLAM+BPAR
      CALL HEAT(BPAR,XLAM,W3A,W1A,W2,DSHUEC)
      DSHUEC=DSHUEC-SHUEC
      GO TO 260
250   CONTINUE
      IF (HDD75-CDD65) 251,252,252
251   TEMP=(65.0+CDD65)
      GO TO 253
252   TEMP=(75.0-HDD75)
253   CONTINUE
      TTT=75.
      CALL ACC1(Q0,Q1,Q2,Q3,Q4,Q5,Q6,Q7,T1,T2,T3,TTT,ACUEC,TEMP)
      TTT=74.
      CALL ACC1(Q0,Q1,Q2,Q3,Q4,Q5,Q6,Q7,T1,T2,T3,TTT,DACUEC,TEMP)
      DACUEC=DACUEC-ACUEC
255   CONTINUE
      T6=XXLAM+70.
      TAU=70.
      CALL HEAT1(T6,TAU,TEMP,W3,W1,W2,SHUEC)
      T6=XXLAM+71.
      TAU=71.
      CALL HEAT1(T6,TAU,TEMP,W3,W1,W2,DSHUEC)
      DSHUEC=DSHUEC-SHUEC
260   CONTINUE
      XD=XDAYS*24.0/1000.0
```

```
      SHUEC=SHUEC*XD
      DSHUEC=DSHUEC*XD
      ACUEC=ACUEC*XD
      DACUEC=DACUEC*XD
      WRITE(3,300) HHIDNO,FLAG,START,END,QUAN,EXPEN,HDD65,CDD65,
     & HDD75,CDD75,SHUEC,DSHUEC,ACUEC,DACUEC,T1,T2,T3,T6,TEMP
300   FORMAT(19E15.8)
      GO TO 20
30    Z=0.0
      WRITE(3,300) HHIDNO,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z,Z
20    CONTINUE
10    CONTINUE
      STOP
      END
      SUBROUTINE GAMMA(RRR,GGG)
      TEMP1=AMAX1(0.,RRR)
      TEMP2=EXP((-1.0)*TEMP1)
      GGG=TEMP2*.00643169*(EXP(5.*RRR)-1.)
      GGG=TEMP2*(-.03401569*(EXP(4.*RRR)-1.)+GGG)
      GGG=TEMP2*(.09649159*(EXP(3.*RRR)-1.)+GGG)
      GGG=TEMP2*(-.24595448*(EXP(2.*RRR)-1.)+GGG)
      GGG=TEMP2*(.99949556*(EXP(RRR)-1.)+GGG)
      GGG=(TEMP1*TEMP1/2.)+.82246703+GGG
      RETURN
      END
      SUBROUTINE HEAT(BPAR,RRR1,CO,C1,C2,HHH)
      HHH=CO/(1.0+EXP(-RRR1))
      HHH=HHH+C1*(ALOG(1.0+EXP(RRR1)))/BPAR
      CALL GAMMA(RRR1,GG)
      HHH=HHH+2.0*C2*GG/(BPAR*BPAR)
      RETURN
      END
      SUBROUTINE HEAT1(T6,TAU,TEMP,CO,C1,C2,HHH)
      IF(TEMP.GT.T6) GO TO 10
      HHH=CO+(TAU-TEMP)*C1+C2*(TAU-TEMP)*(TAU-TEMP)
      GO TO 20
10    HHH=0.0
20    RETURN
      END
      SUBROUTINE COEF(H75,C65,A,B)
      BTOP=1.
      BBOT=0.
      B=1.
      DO 10 I=1,30
      G=(1.-EXP(-B*H75))*(1.-EXP(-B*C65))*EXP(B*(H75+C65-10.))-1.
      IF(G) 11,100,12
   11 BBOT=B
      GO TO 13
   12 BTOP=B
   13 IF((BTOP-BBOT).LT..0001) GO TO 100
   10 B=(BTOP+BBOT)/2.
  100 B=(BTOP+BBOT)/2.
      A=B*(H75-75.)+ALOG(1.-EXP(-B*H75))
      RETURN
      END
```

```
      SUBROUTINE ZETA(APAR,BPAR,TA,TB,PA,PB,Z1,Z2)
      XLAMA=AMAX1(-12.0,(APAR+BPAR*TA))
      XLAMA=AMIN1(15.0,XLAMA)
      XLAMB=AMAX1(-12.0,(APAR+BPAR*TB))
      XLAMB=AMIN1(15.0,XLAMB)
      PA=1.0/(1.0+EXP(-XLAMA))
      PB=1.0/(1.0+EXP(-XLAMB))
      IF (((1.0-PA).LE.0.0001).OR.((1.0-PB).LE.0.0001)) GO TO 10
      Z1=(ALOG((1.0-PB)/(1.0-PA)))/BPAR-(TA-TB)*PB
      CALL GAMMA(XLAMA,GA)
      CALL GAMMA(XLAMB,GB)
      Z2=-1.0*(TB-TA)*ALOG((1.0-PA)*(1.0-PB))/BPAR
     & +2.0*(GA-GB)/(BPAR*BPAR)
      RETURN
10    Z1=(-1.0/BPAR)*ALOG(PA)
      CALL GAMMA(XLAMA,GA)
      GB=1.6449341+0.5*(XLAMB*XLAMB)
      Z2=-1.0*(TB-TA)*(ALOG(PA)-XLAMA-XLAMB)/BPAR
     & +2.0*(GA-GB)/(BPAR*BPAR)
      RETURN
      END
      SUBROUTINE ACC(APAR,BPAR,Q0,Q1,Q2,Q3,Q4,Q5,Q6,Q7,
     &  T1,T2,T3,TT,AAU)
      T1=TT-(Q0+Q1+Q3)/Q2
      T2=TT-Q1/Q2
      T3=TT-(Q1-Q3)/Q2
      T4=TT-(Q1+(Q0+Q3)/1.4142136)/Q2
      T5=TT-(Q1-Q3/1.4142136)/Q2
      Q8=(Q6-Q4*(T4-T1)/(T2-T1))/((T4-T1)*(T2-T4))
      Q9=(Q7-Q4-(Q5-Q4)*(T5-T2)/(T3-T2))/((T5-T2)*(T3-T5))
      CALL ZETA(APAR,BPAR,T1,T2,P1,P2,ZZ1,ZZ2)
      CALL ZETA(APAR,BPAR,T2,T3,P2,P3,ZZ3,ZZ4)
      AAU=Q4*ZZ1/(T2-T1)+Q8*ZZ2+Q4*(P3-P2)+(Q5-Q4)*ZZ3/(T3-T2)+Q9*ZZ4+
     & Q5*(1.0-P3)-1.3*Q2*(ALOG(P3))/BPAR
      RETURN
      END
      SUBROUTINE ACC1(Q0,Q1,Q2,Q3,Q4,Q5,Q6,Q7,
     &  T1,T2,T3,TT,AAU,TEMP)
      T1=TT-(Q0+Q1+Q3)/Q2
      T2=TT-Q1/Q2
      T3=TT-(Q1-Q3)/Q2
      T4=TT-(Q1+(Q0+Q3)/1.4142136)/Q2
      T5=TT-(Q1-Q3/1.4142136)/Q2
      Q8=(Q6-Q4*(T4-T1)/(T2-T1))/((T4-T1)*(T2-T4))
      Q9=(Q7-Q4-(Q5-Q4)*(T5-T2)/(T3-T2))/((T5-T2)*(T3-T5))
      IF(TEMP-T1) 1,2,2
1     AAU=0.0
      GO TO 7
2     IF(TEMP-T2) 3,4,4
3     AAU=Q4*(TEMP-T1)/(T2-T1)+Q8*(TEMP-T1)*(T2-TEMP)
      GO TO 7
4     IF(TEMP-T3) 5,6,6
5     AAU=Q4+(TEMP-T2)*(Q5-Q4)/(T3-T2)+Q9*(TEMP-T2)*(T3-TEMP)
      GO TO 7
6     AAU=1.3*(Q0/3.141592654+Q1+Q2*(TEMP-TT))
7     RETURN
      END
```

Appendix II. Conditional Moments in the Generalized Extreme Value Family

In this appendix we establish basic results on the conditional moments of generalized extreme value random variables.[1] We proceed as follows. The generalized extreme value distribution is introduced. We then discuss implications for the marginal extreme value distributions. The first, second and cross moments for G.E.V. variates are derived conditional on the event that a specific alternative is chosen.

Finally we specify a random variable through its linear conditional expectation in the space of G.E.V. random variables and derive its properties. These results in particular provide the distributional framework for the two step consistent estimation techniques to be considered below.

The following theorem due to McFadden (1977) introduces a general family of generalized extreme value choice models.

<u>Theorem 1 (McFadden)</u>

Suppose $G(y_1, y_2, \ldots, y_J)$ is a nonnegative, homogeneous of degree one function of $(y_1, y_2, \ldots, y_J) \geqq 0$. Suppose $\lim_{y_i \longrightarrow +\infty} G(y_1, y_2, \ldots, y_J) = +\infty$ for $i = 1, 2, \ldots, J$. Suppose for any distinct $(i_1, i_2, \ldots, i_k)$ from $\{1, 2, \ldots, J\}$, $\partial^k G/\partial y_{i_1}, \ldots, \partial y_{i_k}$ is nonnegative if k is odd and nonpositive if k is even. Then,

(1) $P_i = e^{V_i} G_i(e^{V_1}, \ldots, e^{V_J})/G(e^{V_1}, \ldots, e^{V_J})$ defines a choice model which is consistent with utility maximization.

<u>Proof Theorem 1</u>.

Theorem 1 is proved in two steps. The first step demonstrates that the function:

$$(2)\qquad F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J) = \exp\left[- G(e^{-\varepsilon_1}, e^{-\varepsilon_2}, \ldots, e^{-\varepsilon_J})\right]$$

is a multivariate extreme value distribution. The details may be found in McFadden (1977).

The second step assumes a population of individuals with utilities $u_i = V_i + \varepsilon_i$, where $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J)$ is distributed F. Let $\underset{\sim}{\varepsilon}$ denote the vector $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J)$ then

$$(3)\qquad P_i = \text{Prob}[u_i \geqq u_j, \forall i \neq j] = \text{Prob}[V_i + \varepsilon_i \geqq V_j + \varepsilon_j, \forall i \neq j]$$

may be written

$$(4)\qquad \int_{\varepsilon_i=-\infty}^{+\infty} F_i(<V_i + \varepsilon_i - V_j>)d\varepsilon_i$$

where F_i denotes the derivative of F with respect to its ith argument, and $<a_j>$ denotes a vector with jth component a_j. From (4) and the definition of the generalized extreme value distribution we have:

$$(5)\qquad P_i = \int_{-\infty}^{+\infty} \exp[-G[<e^{-(V_i+\varepsilon_i-V_j)}>]G_i[<e^{-(V_i+\varepsilon_i-V_j)}>]e^{-\varepsilon_i}\, d\varepsilon_i$$

$$= \int_{-\infty}^{+\infty} e^{-\varepsilon_i}\, G_i[<e^{V_j}>] \exp\left[-G[<e^{V_j}>] \,.\, e^{-V_i}\, e^{-\varepsilon_i}\right] d\varepsilon_i$$

$$= \frac{G_i[<e^{V_j}>]}{G[<e^{V_j}>]}\, e^{V_i} \qquad\qquad \text{Q.E.D.}$$

The second equality in equation (5) uses the fact that G is homogeneous of degree one and the implication that G_i is homogeneous of degree zero. The third equality makes use of the result:

$$(6) \qquad \int_{-\infty}^{+\infty} e^{-\varepsilon}\exp[-e^{-\varepsilon} . \phi]d\varepsilon = \phi^{-1}$$

which follows from the subsitution $u ==> -\phi e^{-\varepsilon}$.

Corollary 1. Multinomial Logit Model

Let $G[y] = [\sum_{j=1}^{J} y_j^{1/\phi}]^{\phi}$. Then

$$P_i = e^{V_i/\phi} \Big/ \sum_{j=1}^{J} e^{V_j/\phi}$$

Proof Corollary 1:

This result is found in McFadden (1976). One need simply verify the linear homogeneity of G and apply (1). Q.E.D.

McFadden shows that when $\varepsilon_j \xrightarrow{\text{Lim}} +\infty$ for $j \neq i$, then from (2), $F[\varepsilon_i] = \exp[-a_i e^{-\varepsilon_i}]$, where $a_i = G[0, \ldots, 0, 1, 0, \ldots, 0]$; one in the ith coordinate. Under the assumptions of Corollary 1, the marginal distribution, $F[\varepsilon_i]$, is $\exp[-e^{-\varepsilon_i}]$ (since $a_i = 1$) which is the cumulative distribution for an extreme value distributed random variate with variance $\pi^2/6$. We note that McFadden's definition of the generalized extreme value distribution is easily modified to encompass marginal distributions with non-normalized variances by choosing:

$$(7) \qquad F[\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J] = \exp[- G[<e^{-\varepsilon_j/\phi}>]].$$

McFadden's proof of Theorem 1 may be modified to demonstrate that (7) is a multivariate extreme value distribution. Alternatively, since (2) is a multivariate extreme value distribution, we see by inspection that (7) is as well. Application of (4) implies the probability choice system

$$(8) \qquad P_i = e^{V_i/\phi} \cdot G_i[<e^{V_j/\phi}>] \; / \; G[<e^{V_j/\phi}>].$$

When $G[<y_j>] = \sum_{j=1}^{J} y_j$, equation (8) implies choice probabilities of the multinomial logit form in Corollary 1.

The marginal distribution for ε_i from (7) is $F(\varepsilon_i) = \exp[-a_i e^{-\varepsilon_i/\phi}]$ which is the cumulative distribution for an extreme value distributed random variate with variance $\frac{\pi^2}{6} \cdot \phi^2$. We have applied the following result:

Theorem 2

A random variate ε with the extreme value distribution $F_\varepsilon[t] = \text{Prob}[\varepsilon \leq t] = \exp[-e^{-(t-\alpha)/\phi}]$ has the properties:

T2a) $E[\varepsilon] = \alpha + \gamma\phi$ where $\gamma = .5772156649 \ldots$ is Euler's constant and

T2b) $\text{Var}[\varepsilon] = \frac{1}{6}\pi^2\phi^2$.

Proof Theorem 2

See McFadden (1973). Q.E.D.

When $G[<y_j>] = \sum_{i=1}^{J} y_i$, (7) implies that ε_i has mean $\gamma\phi$ (since $\alpha = 0$) and variance $\frac{1}{6}\pi^2\phi^2$. Application of Theorem 2 demonstrates that ε_i has a marginal distribution with zero mean when $G[y_1, y_2, \ldots, y_J] = e^{-\gamma} \cdot \sum_{j=1}^{J} y_j$. More generally, ε_i will have mean u and variance $\frac{1}{6}\pi^2\phi^2$ when

$$G[y_1, y_2, \ldots, y_J] = \left(\frac{\exp(u/\phi)}{\exp(\gamma)}\right) \cdot \left(\sum_{j=1}^{J} y_j\right).$$

Let $\delta_j(\underset{\sim}{\varepsilon})$ be an indicator random variable which is one when j is the chosen alternative, i.e., when $V_j + \varepsilon_j \geq V_i + \varepsilon_i$, $\forall i \neq j$, and zero otherwise. We have written δ_j as a function of $\underset{\sim}{\varepsilon}$ to emphasize that δ_j is a random variable whose outcome conditioned on the V_j's depends directly on the realization of $\underset{\sim}{\varepsilon}$. We now derive the conditional moments. Note that without loss of generality it suffices to consider expressions $E[\varepsilon_1 | \delta_1 = 1]$ and $E[\varepsilon_2 | \delta_1 = 1]$ rather than the more general expression $E[\varepsilon_i | \delta_j = 1]$ for $i = j$ and for $i \neq j$.

Lemma 1

Let $\underset{\sim}{\varepsilon}$ be generalized extreme value distributed with cumulative distribution function $F(\underset{\sim}{\varepsilon})$ given in (7). Let g(.) be an arbitrary real-valued function. Then:

L1a) $E[g(\varepsilon_1) | \delta_1(\underset{\sim}{\varepsilon}) = 1]$

$$= E[g(\varepsilon) | \varepsilon \sim EV(\phi(\ln G_1 - \ln P_1), \phi)$$

where EV[a,b] denotes an extreme-valued distributed random variate with location parameter a and scale parameter b.

L1b) Let G be additively separable as $G(y) = G^A(y^A) + y_2$ with $y = (y^A, y)$ and with $G^A(.)$ homogeneous of degree one. Let ε have the corresponding partition, i.e., $\underset{\sim}{\varepsilon} = (\varepsilon^A, \varepsilon_2)$. Then $E[g(\varepsilon_2) | \delta_1(\underset{\sim}{\varepsilon}) = 1] =$

$$\frac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]} \left[E\Big(g(\varepsilon_2) | \varepsilon_2 \sim EV[0, \phi]\Big) - P_2 E\Big(g(\varepsilon_2) | \varepsilon_2 \sim EV[-\phi(\ln P_2), \phi]\Big) \right]$$

Proof Lemma 1

L1a) We make use of the properties of conditional densities. Recall:

(9) $$\int_{-\infty}^{y} \int_{x\varepsilon A} f(x,y)dxdy = PR[x\varepsilon A,\ Y \leq y] = PR[Y \leq y | x\varepsilon A]\ PR[x\varepsilon A]$$

Thus:

(10) $$\frac{1}{PR[x\varepsilon A]} \int_{x\varepsilon A} f(x,y)dx = f(y|x\varepsilon A).$$

Equation (10) implies that:

(11) $$E[Y|x\varepsilon A] = \int_{y} yf(y|x\varepsilon A)dy = \frac{1}{PR[x\varepsilon A]} \int_{y} \int_{x\varepsilon A} yf(x,y)dxdy$$

As an application of (11) we find:

(12) $$E[g(\varepsilon_1)|\delta_1(\underset{\sim}{\varepsilon}) = 1]$$

$$= \frac{1}{P_1} \int_{\varepsilon_1=-\infty}^{\infty} \int_{\varepsilon_2=-\infty}^{V_1-V_2+\varepsilon_1} \cdots \int_{\varepsilon_J=-\infty}^{V_1-V_J+\varepsilon_1} g(\varepsilon_1)dF(\underset{\sim}{\varepsilon})$$

$$= \frac{1}{P_1} \int_{\varepsilon=-\infty}^{\infty} g(\varepsilon)F_1[<\varepsilon + V_1 - V_j>]d\varepsilon$$

$$= \frac{1}{P_1} \int_{\varepsilon=-\infty}^{\infty} g(\varepsilon)e^{-\varepsilon/\phi}\, G_1[<e^{-(\varepsilon+V_1-V_j)/\phi}>]\exp[-G[<e^{-(\varepsilon+V_1-V_j)/\phi}>]]\, \frac{d\varepsilon}{\phi}$$

$$= \frac{1}{P_1} \int_{\varepsilon=-\infty}^{\infty} g(\varepsilon)e^{-\varepsilon/\phi}\, G_1[<e^{V_j/\phi}>]\exp\left(-G[<e^{V_j/\phi}>]e^{-\varepsilon/\phi}\, e^{-V_1/\phi}\right)\frac{d\varepsilon}{\phi}$$

Let $\phi_1 = G[<e^{V_j/\phi}>]e^{-V_1/\phi}$ and $\phi_2 = G_1[<e^{V_j/\phi}>]$

$$(12) = \frac{\phi_2}{P_1} \bullet \int_{-\infty}^{\infty} g(\varepsilon)e^{-\varepsilon/\phi} \exp[-\phi_1 e^{-\varepsilon/\phi}] \frac{d\varepsilon}{\phi}$$

$$= \frac{\phi_2}{P_1\phi_1} \bullet \int_{-\infty}^{\infty} g(\varepsilon)e^{-(\varepsilon-\phi k_1)/\phi} \exp[-e^{-(\varepsilon-\phi k_1)/\phi}] \frac{d\varepsilon}{\phi}$$

where $k_1 = \ln \phi_1$

$$= E[g(\varepsilon)|\varepsilon \sim EV(\phi \ln \phi_1, \phi)]$$

where EV[a, b] denotes an extreme-value distributed random variate with location parameter a and scale parameter b, i.e., $F_\varepsilon[t] = \exp[-e^{-(t-a)/b}]$.

From equation (8), $\phi_2/\phi_1 = G_1/\phi_1 = P_1$. Hence $\ln \phi_1 = (\ln G_1 - \ln P_1)$, so that we can make the substitution in the final equality of (12) to prove the claim.

L1b)

(13) $E(g(\varepsilon_2)|\delta_1(\underset{\sim}{\varepsilon}) = 1)$

$$= \frac{1}{P_1} \int_{\varepsilon_1=-\infty}^{+\infty} \int_{\varepsilon_2=-\infty}^{V_1-V_2+\varepsilon_1} \cdots \int_{\varepsilon_J=-\infty}^{V_1-V_J+\varepsilon_1} g(\varepsilon_2)dF(\underset{\sim}{\varepsilon})$$

$$= \frac{1}{P_1} \int_{\varepsilon_1=-\infty}^{+\infty} \int_{\varepsilon_2=-\infty}^{V_1-V_2+\varepsilon_1} g(\varepsilon_2)F_{12}[\varepsilon_1,\varepsilon_2,V_1-V_3+\varepsilon_1,\ldots,V_1-V_J+\varepsilon_1]d\varepsilon_2 d\varepsilon_1$$

$$= \frac{1}{P_1} \int_{\varepsilon_2=-\infty}^{+\infty} \int_{\varepsilon_2+V_2-V_1}^{+\infty} g(\varepsilon_2)F_{12}[\varepsilon_1,\varepsilon_2,V_1-V_3+\varepsilon_1,\ldots,V_1-V_J+\varepsilon_1]d\varepsilon_1 d\varepsilon_2$$

From equation (7),

(14) $$F(\underset{\sim}{\varepsilon}) = \exp[-\, G[<e^{-\varepsilon_j/\phi}>]]$$

$$= \exp\left[-\, G^A[<e^{-\varepsilon_j^A/\phi}>]\right] \,.\, \exp[-e^{-\varepsilon_2/\phi}],$$ so that:

(15) $$F_{12}(\underset{\sim}{\varepsilon}) = \exp\left[-\, G^A[<e^{-\varepsilon_j^A/\phi}>]\right] G_1^A[<e^{-\varepsilon_j^A/\phi}>]\; e^{-\varepsilon_1/\phi}\; \frac{1}{\phi}$$

$$.\; \exp[-e^{-\varepsilon_2/\phi}] e^{-\varepsilon_2/\phi}\; \frac{1}{\phi}$$

Hence:

(16) $$F_{12}(\varepsilon_1,\ \varepsilon_2,\ V_1-V_3+\varepsilon_1,\ \ldots,\ V_1-V_J+\varepsilon_1)$$

$$= \exp[-\, G^A[e^{-\varepsilon_1/\phi},\ <e^{\frac{-V_1+V_j-\varepsilon_1}{\phi}}>]]$$

$$.\; G_1^A\left[e^{-\varepsilon_1/\phi},\ <e^{\frac{-V_1+V_j-\varepsilon_1}{\phi}}>\right] e^{-\varepsilon_1/\phi}\; \frac{1}{\phi}\; \exp[-e^{-\varepsilon_2/\phi}] e^{-\varepsilon_2/\phi}\; \frac{1}{\phi}$$

$$= \exp\left[-e^{-\varepsilon_1/\phi}\; e^{-V_1/\phi}\;.\; G^A[<e^{V_j/\phi}>]\right] G_1^A[<e^{V_j/\phi}>]\; e^{-\varepsilon_1/\phi}\; \frac{1}{\phi}$$

$$.\; \exp[-e^{-\varepsilon_2/\phi}] e^{-\varepsilon_2/\phi}\; \frac{1}{\phi}$$

(17) $E[g(\varepsilon_2) | \delta_1(\underset{\sim}{\varepsilon}) = 1] =$

$$\frac{G_1^A[<e^{V_j/\phi}>]}{P_1} \int_{\varepsilon_2=-\infty}^{\infty} g(\varepsilon_2) e^{-\varepsilon_2/\phi} \exp[-e^{-\varepsilon_2/\phi}]$$

$$\bullet \int_{\varepsilon_2+V_2-V_1}^{\infty} \exp\left(-e^{-\varepsilon_1/\phi} e^{-V_1/\phi} G^A[<e^{V_j/\phi}>]\right) e^{-\varepsilon_1/\phi} \frac{d\varepsilon_1}{\phi} \frac{d\varepsilon_2}{\phi}$$

$$= \frac{G_1^A[<e^{V_j/\phi}>]}{P_1 \cdot \phi_1^A} \int_{\varepsilon_2=-\infty}^{\infty} g(\varepsilon_2) e^{-\varepsilon_2/\phi} \exp[-e^{-\varepsilon_2/\phi}] \left[1 - \exp[-\phi_1^A e^{\frac{-\varepsilon_2 - V_2 + V_1}{\phi}}]\right] \frac{d\varepsilon_2}{\phi}$$

where $\phi_1^A = e^{-V_1/\phi} \; G^A[<e^{V_j/\phi}>]$ Thus:

$$(17) = \frac{G_1^A[<e^{V_j/\phi}>]}{P_1 \cdot \phi_1^A} \quad E[g(\varepsilon_2) | \varepsilon_2 \sim EV(0,\phi)]$$

$$- \frac{G_1^A[<e^{V_j/\phi}>]}{P_1 \cdot \phi_1^A} \int_{\varepsilon_2=-\infty}^{\infty} g(\varepsilon_2) e^{-\varepsilon_2/\phi} \exp[-e^{-\varepsilon_2/\phi}] \exp[-\phi_1^A e^{\frac{-\varepsilon_2 - V_2 + V_1}{\phi}}] \frac{d\varepsilon_2}{\phi}$$

Now $$\int_{\varepsilon_2=-\infty}^{\infty} g(\varepsilon_2) e^{-\varepsilon_2/\phi} \exp[-e^{-\varepsilon_2/\phi} \cdot \phi_2] \frac{d\varepsilon_2}{\phi} = \frac{1}{\phi_2} E\left[g(\varepsilon_2) | \varepsilon_2 \sim EV(\phi \ln \phi_2, \phi)\right]$$

where we have defined $\phi_2 = (1 + e^{(V_1 - V_2)/\phi} \bullet \phi_1^A)$. Hence:

(18) $E[g(\varepsilon_2) | \delta_1(\underset{\sim}{\varepsilon}) = 1] =$

$$\frac{G_1^A[<e^{V_j/\phi}>]}{P_1 \cdot \phi_1^A} \left[E\left(g(\varepsilon_2) | \varepsilon_2 \sim EV(0,\phi)\right) - \frac{1}{\phi_2} E\left(g(\varepsilon_2) | \varepsilon_2 \sim EV(\phi \ln\phi_2, \ \phi) \right) \right]$$

Note that $G_1^A \ [<e^{V_j/\phi}>] = G_1[<e^{V_j/\phi}>]$ implies:

(19) $$\frac{G_1^A[<e^{V_j/\phi}>]}{P_1 \cdot \phi_1^A} = \frac{G_1[<e^{V_j/\phi}>] \ e^{V_1/\phi}}{G^A[<e^{V_j/\phi}>]P_1} = \frac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]}$$

Also:

(20) $$\phi_2 = (1 + e^{(V_1-V_2)/\phi} \ \phi_1^A) = (1 + e^{-V_2/\phi} \ G^A[<e^{V_j/\phi}>])$$

$$= e^{-V_2/\phi}\left(e^{V_2/\phi} + G^A[<e^{V_j/\phi}>]\right) = e^{-V_2/\phi} \ G[<e^{V_j/\phi}>]$$

Note that $G_2/\phi_2 = P_2$ and $G_2 \equiv 1$ imply:

(21) $$\frac{e^{V_2/\phi}}{G[<e^{V_j/\phi}>]} = \frac{1}{\phi_2} = P_2$$

Combining equations (19) and (21) with equation (18) we have:

(22) $E[g(\varepsilon_2) | \delta_1(\underset{\sim}{\varepsilon}) = 1] =$

$$\frac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]} \left[E\left(g(\varepsilon_2) | \varepsilon_2 \sim EV[0, \ \phi]\right) - P_2 \ E\left(g(\varepsilon_2) | \varepsilon_2 \sim EV[\phi \ln\phi_2, \ \phi]\right) \right]$$

From equation (21) we have:

(23) $\ln\phi_2 = \ln G_2 - \ln P_2 = -\ln P_2.$

Combining (22) and (23) with (21) proves the claim. Q.E.D.

As an application of Lemma 1 we have:

Theorem 3.

Let $\underset{\sim}{\varepsilon}$ be generalized extreme value distributed with cumulative distribution function $F(\underset{\sim}{\varepsilon})$ given in (7). Then:

T3a) $E[\varepsilon_1 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = \phi[\gamma + \ln G_1 - \ln P_1]$

T3b) $E[\varepsilon_1^2 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{\pi^2}{6}\phi^2 + \phi^2[\gamma + \ln G_1 - \ln P_1]^2$

Let G be additively separable as $G(y) = G^A(y^A) + y_2$ with $y = (y^A, y_2)$ and with $G^A(.)$ homogenous of degree one. Let $\underset{\sim}{\varepsilon}$ have the corresponding partition, i.e., $\underset{\sim}{\varepsilon} = (\varepsilon^A, \varepsilon_2)$. Then:

T3c) $E[\varepsilon_2 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = \dfrac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]} \cdot \phi \cdot \left((1-P_2)\gamma + P_2 \ln P_2\right)$

T3d) $E[\varepsilon_2^2 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = \dfrac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]} \cdot \phi^2 \cdot \left(\gamma^2 - P_2(\gamma - \ln P_2)^2 + (1-P_2)\frac{\pi^2}{6}\right)$

Proof Theorem 3:

T3a) Using Lemma 1a with $g(\varepsilon) = \varepsilon$, we have:

(24) $$E[\varepsilon_1 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = E[\varepsilon | \varepsilon \sim EV(\phi(\ln G_1 - \ln P_1), \phi)]$$
$$= \phi[\gamma + \ln G_1 - \ln P_1]$$

where the second equality uses Theorem 2a.

T3b) We take $g(\varepsilon) = \varepsilon^2$ so that:

(25) $$E[\epsilon_1^2|\delta_1(\underset{\sim}{\epsilon}) = 1] = E\left(\epsilon^2|\epsilon \sim EV[\phi(\ln G_1 - \ln P_1), \phi]\right)$$

$$= \left(E[\epsilon|\epsilon \sim EV[\phi(\ln G_1 - \ln P_1), \phi]\right)^2$$

$$+ \text{var}\left(\epsilon|\epsilon \sim EV[\phi(\ln G_1 - \ln P_1), \phi]\right)$$

$$= \phi^2[\gamma + \ln G_1 - \ln P_1]^2 + \frac{\pi^2}{6}\phi^2$$

where the third equality uses Theorem 2b.

T3c) Using Lemma 1b with $g(\epsilon) = \epsilon$ we have:

(26) $$E[\epsilon_2|\delta_1(\underset{\sim}{\epsilon}) = 1]$$

$$= \left(\frac{G}{G^A}\right) \cdot \left[E\left(\epsilon_2|\epsilon_2 \sim EV[0, \phi]\right) - P_2 E\left(\epsilon_2|\epsilon_2 \sim EV[-\phi\ln P_2, \phi]\right)\right]$$

$$= \left(\frac{G}{G^A}\right) \cdot \left(\gamma\phi - P_2(\gamma\phi - \phi\ln P_2)\right)$$

$$= \left(\frac{G}{G^A} \cdot \phi\right) \cdot \left((1-P_2)\gamma + P_2 \ln P_2\right)$$

T3d) Using Lemma 1b with $g(\epsilon) = \epsilon^2$ we have:

(27) $$E[\epsilon_2^2|\delta_1(\underset{\sim}{\epsilon}) = 1]$$

$$= \left(\frac{G}{G^A}\right)\left[E\left(\epsilon_2^2|\epsilon_2 \sim EV[0, \phi]\right) - P_2 E\left(\epsilon_2^2|\epsilon_2 \sim EV[-\phi\ln P_2, \phi]\right)\right]$$

$$= \left(\frac{G}{G^A}\right)\left[\left((\gamma\phi)^2 + \frac{\pi^2}{6}\phi^2\right) - P_2\left(\phi^2(\gamma - \ln P_2)^2 + \frac{\pi^2}{6}\phi^2\right)\right]$$

$$= \left(\frac{G}{G^A}\right)\left[(\gamma\phi)^2 - P_2\phi^2(\gamma - \ln P_2)^2 + (1-P_2)\frac{\pi^2}{6}\phi^2\right]$$

$$= \left(\frac{G}{G^A} \cdot \phi^2\right)\left[\gamma^2 - P_2(\gamma - \ln P_2)^2 + (1-P_2)\frac{\pi^2}{6}\right] \qquad \text{Q.E.D.}$$

Comments: Theorem 3 imposes strong separability in the functional form for G to obtain a closed form conditional expectation. If in fact G has the additive form $G[y] = G^A[y_2^A] + y_2$ then ε_2 is independent from ε^A. If we do not impose strong separability then $F_{12}(\underset{\sim}{\varepsilon})$ in equation (13) becomes:

$$(28) \qquad F_{12}(\underset{\sim}{\varepsilon}) = \exp\left(-G[<e^{-\varepsilon_j/\phi}>]\right) e^{-\varepsilon_1/\phi}\, e^{-\varepsilon_2/\phi}\, \frac{1}{\phi^2} \cdot \left(G_1[<e^{-\varepsilon_j/\phi}>]\, G_2[<e^{-\varepsilon_j/\phi}>] - G_{12}[<e^{-\varepsilon_j/\phi}>]\right)$$

Following the proof of Lemma 1b we see that the analogue of (16) corresponding to equation (28) does not permit an easy integration in (17).

However, it is possible to extend the results of Theorems 3c and 3d by assuming $G[y] = G^A[y^A] + \alpha y_2$.

We present the results in Theorem 4.

Theorem 4.

Let ε be generalized extreme value distributed with cumulative distribution function $F(\underset{\sim}{\varepsilon})$ given in (7).

Let G be additively separable as $G(y) = G^A(y^A) + \alpha y_2$ where $y = (y^A, y_2)$ and with $G^A(.)$ homogeneous of degree one. Let $\alpha^* = \phi \ln \alpha$. Then:

T4a) $$E[\varepsilon_2|\delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]}\left[(\gamma\phi + \alpha^*)(1 - P_2) + \phi P_2 \ln P_2)\right]$$

T4b) $$E[\varepsilon_2^2|\delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]}\left[\frac{\pi^2}{6}\phi^2(1 - P_2) + (\gamma\phi + \alpha^*)^2\ (1 - P_2)\right.$$
$$\left. + 2\phi(\gamma\phi + \alpha^*)\ .\ P_2 \ln P_2 - \phi^2 P_2(\ln P_2)^2\right]$$

Proof Theorem 4:

The proof of Theorem 4 requires minor modifications in the arguments which demonstrate Lemma 1b, Theorem 3c, and Theorem 3d. It is therefore omitted. Q.E.D.

As a corollary to Theorems 3 and 4 we derive the conditional moments for the multinomial logit and nested logit models.

Corollary 2. Conditional Moments in the Multinomial Logit Model

Let $G[y] = \alpha\left[\sum_{j=1}^{J} y_j\right]$. Then:

C2a) $E[\varepsilon_1|\delta_1(\underset{\sim}{\varepsilon}) = 1] = (\alpha^* + \gamma\phi) - \phi \ln P_1$ where $\alpha^* = \phi \ln \alpha$.

C2b) $$E[\varepsilon_1^2|\delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{1}{6}\pi^2\phi^2 + (\alpha^* + \gamma\phi)^2 + \phi^2(\ln P_1)^2$$
$$- 2(\alpha^* + \gamma\phi).\phi(\ln P_1)$$

C2c) $E[\varepsilon_2|\delta_1(\underset{\sim}{\varepsilon}) = 1] = (\alpha^* + \gamma\phi) + \phi P_2 \ln P_2/(1-P_2)$

C2d) $E[\varepsilon_2^2 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{1}{6}\pi^2\phi^2 + (\alpha^* + \gamma\phi)^2 - P_2\phi^2(\ln P_2)^2/(1-P_2)$

$$+ 2(\alpha^* + \gamma\phi)(\phi \ln P_2)P_2/(1-P_2)$$

Proof Corollary 2:

C2a) $G_1 = \alpha$ and $\phi \ln G_1 = \phi \ln \alpha = \alpha^*$. Apply Theorem 3a.

C2b) Use Theorem 3b and $G_1 = \alpha$ to find:

$$E[\varepsilon_1^2 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{\pi^2}{6}\phi^2 + \phi^2[\gamma + \ln\alpha - \ln P_1]^2$$

$$= \frac{\pi^2}{6}\phi^2 + \phi^2(\gamma + \ln\alpha)^2 - 2\phi^2(\gamma + \ln\alpha)(\ln P_1) + \phi^2(\ln P_1)^2$$

$$= \frac{\pi^2}{6}\phi^2 + (\gamma\phi + \alpha^*)^2 - 2(\gamma\phi + \alpha^*)\phi(\ln P_1) + \phi^2(\ln P_1)^2 .$$

C2c) Apply Theorem 4a with $G^A[y^A] = \alpha\left[\sum_{j \neq 2} y_j\right]$ so that:

$$E[\varepsilon_2 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]}(\gamma\phi + \alpha^*)(1-P_2) + \phi P_2 \ln P_2 .$$

(29) $$\frac{G[<e^{V_j/\phi}>]}{G^A[<e^{V_j/\phi}>]} = \alpha \sum_{j=1}^{J} e^{V_j/\phi} \Big/ \alpha\left[\sum_{j \neq 2}^{J} e^{V_j/\phi}\right] = 1/(1-P_2) \text{ from equation (8).}$$

Thus $E[\varepsilon_2 | \delta_1(\underset{\sim}{\varepsilon}) = 1] = (\gamma\phi + \alpha^*) + \phi P_2 \ln P_2/(1-P_2)$.

C2d) Apply Theorem 4b with $G^A[y^A] = \alpha\left[\sum_{j\neq 2}^{J} y_j\right]$ and (29):

$$E[\epsilon_2^2 | \delta_1(\underset{\sim}{\epsilon}) = 1] = \frac{\pi^2}{6}\phi^2 + (\gamma\phi + \alpha^*)^2 + 2\phi(\gamma\phi + \alpha^*)P_2 \ln P_2/(1-P_2)$$

$$- \phi^2 P_2(\ln P_2)^2/(1-P_2) \;. \qquad \text{Q.E.D.}$$

As a second illustration of Theorems 3 and 4 we consider a two-level nested logit model with three alternatives:

(30) $$G[y_1, y_2, y_3] = \left[y_1^{1/(1-\sigma)} + y_3^{1/(1-\sigma)}\right]^{(1-\sigma)} + y_2$$

Following McFadden (1977) one may verify that (30) satisfies the conditions of Theorem 1. From (8),

(31) $$P[2|1,2,3] = \frac{e^{V_2/\phi}}{\left[e^{V_1/\phi(1-\sigma)} + e^{V_3/\phi(1-\sigma)}\right]^{(1-\sigma)} + e^{V_2/\phi}}$$

(32) $$P[1|1,2,3] = \frac{\left[e^{V_1/\phi(1-\sigma)} + e^{V_3/\phi(1-\sigma)}\right]^{(1-\sigma)}}{\left[e^{V_1/\phi(1-\sigma)} + e^{V_3/\phi(1-\sigma)}\right]^{(1-\sigma)} + e^{V_2/\phi}} \;.$$

$$\frac{e^{V_1/\phi(1-\sigma)}}{\left[e^{V_1/\phi(1-\sigma)} + e^{V_3/\phi(1-\sigma)}\right]}$$

$$= P[(1,3)|(1,2,3)] \;.\; P[1|(1,3)]$$

where $P(i|A)$ denotes the probability that i is chosen from the set A.

From equation (30) we calculate:

$$(33) \qquad G_1 = \left[e^{V_1/\phi(1-\sigma)} + e^{V_3/\phi(1-\sigma)}\right]^{-\sigma} \cdot e^{V_1\sigma/\phi(1-\sigma)} = P[1|1,3]^{\sigma}$$

Further we define $G^A[y_1, y_2, y_3] = \left[y_1^{1/(1-\sigma)} + y_3^{1/(1-\sigma)}\right]^{(1-\sigma)}$ so that:

$$(34) \qquad (\frac{G}{G^A})[<e^{V_j/\phi}>] = \frac{\left[e^{V_1/\phi(1-\sigma)} + e^{V_3/\phi(1-\sigma)}\right]^{(1-\sigma)} + e^{V_2/\phi}}{\left[e^{V_1/\phi(1-\sigma)} + e^{V_3/\phi(1-\sigma)}\right]^{(1-\sigma)}}$$

$$= 1 + \frac{P[2|(1,2,3)]}{P[(1,3)|(1,2,3)]}$$

$$= (1 - P[2|(1,2,3)])^{-1}$$

Application of Theorem 3a and Theorem 3b for G given by (30) implies:

$$(35) \qquad E[\varepsilon_1|\delta_1(\underset{\sim}{\varepsilon}) = 1] = \phi\left(\gamma + \sigma . \ln P(1|1,3) - \ln P(1|1,2,3)\right)$$

$$(36) \qquad E[\varepsilon_1^2|\delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{\pi^2}{6}\phi^2 + \phi^2\left(\gamma + \sigma . \ln P(1|1,3) - \ln P(1|1,2,3)\right)^2$$

Application of Theorem 3c and Theorem 3d using (34) imply:

$$(37) \qquad E[\varepsilon_2|\delta_1(\underset{\sim}{\varepsilon}) = 1] = \phi[\gamma + P_2\ln P_2/(1-P_2)] \quad \text{and}$$

$$(38) \qquad E[\varepsilon_2^2|\delta_1(\underset{\sim}{\varepsilon}) = 1] = \phi^2\left[\frac{\pi^2}{6} + \left(\gamma^2 - P_2(\gamma - \ln P_2)^2\right)/(1-P_2)\right]$$

In equations (35) and (36), one observes that the nested logit model implies a closed-form expression in the conditional probabilities of reaching alternative one from different nodes of the tree.

The conditional expectations in (35) and (36) differ from their counterparts derived in Corollary 2a and Corollary 2b for the multinomial logit model by the term $\sigma \ln P(1|(1,3))$. As σ tends to zero in the limit, the nested logit model converges to the multinomial logit model and the term $\sigma \ln P(1|(1,3))$ vanishes.

Comparison of (37) and (38) with the corresponding expressions in Corollary 2 reveals equal conditional expectations for both models. In other words, the variate ε_2 behaves as if it were given from a multinomial logit specification rather than equation (30). This is of course the essence of the separability assumption.

The calculations involved in (35) - (38) are easily modified to trees of any depth. As an illustration consider the nested logit model:

$$(39) \qquad G(y) = \sum_{m=1}^{M} a_m \left[\sum_{i \varepsilon B_m} y_i^{1/(1-\sigma_m)} \right]^{1-\sigma_m}$$

where $B_m \subseteq \{1,2 \ldots, J\}$, $\bigcup_{m=1}^{M} B_m = \{1,2, \ldots, J\}$, $a_m > 0$, and $0 \leq \sigma_m < 1$. McFadden (1976) derives the choice probabilities for equation (39) and shows that they satisfy:

$$(40) \qquad P_i = \sum_{m}^{M} \sum_{i \varepsilon B_m} e^{V_i/(1-\sigma_m)} a_m \left[\sum_{j \varepsilon B_m} e^{V_j/(1-\sigma_m)} \right]^{-\sigma_m} \Bigg/ \sum_{n=1}^{M} a_n \left[\sum_{k \varepsilon B_n} e^{V_k/(1-\sigma_n)} \right]^{(1-\sigma_n)}$$

$$= \sum_{m}^{M} \sum_{i \varepsilon B_m} P[i|B_m]P[B_m] \qquad \text{where:}$$

$$(41) \qquad P[i|B_m] = \begin{cases} e^{V_i/(1-\sigma_m)} \Big/ \sum_{j \varepsilon B_m} e^{V_j/(1-\sigma_m)} & \text{if } i \varepsilon B_m \\ 0 & \text{otherwise} \end{cases}$$

and where:

$$(42) \qquad P[B_m] = a_m \left[\sum_{j \varepsilon B_m} e^{V_j/(1-\sigma_m)} \right]^{(1-\sigma_m)} \Bigg/ \sum_{n=1}^{M} a_n \left[\sum_{k \varepsilon B_n} e^{V_k/(1-\sigma_n)} \right]^{(1-\sigma_n)}$$

From (39) we have:

$$(43) \qquad G_i(y) = \sum_{m\, i \in B_m}^{M} a_m \Big[\sum_{j \in B_m} y_j^{1/(1-\sigma_m)}\Big]^{-\sigma_m} \bullet y_i^{\sigma_m/(1-\sigma_m)} \quad \text{so that:}$$

$$(44) \qquad G_i(<e^{V_j}>) = \sum_{m=1}^{M} a_m P[i|B_m]^{\sigma_m}$$

The form of the derivative in (44) generalizes to higher order tress. As an example consider a three-level tree structure implied by

$$(45) \qquad G = \sum_a \left[\sum_d \left[\sum_m \quad y_{mda}^{1/(1-\sigma)} \right]^{(1-\sigma)/(1-\delta)} \right]^{(1-\delta)}$$

In this case one may show

$$(46) \qquad G_{mda}[<e^{V_j}>] = \sum_a \sum_d P[d|a]^{\delta} \,.\, P[m|da]^{\sigma}$$

where G_{mda} denotes the derivative of G in (45) with respect to y_{mda}. Furthermore, equation (34) will generalize to cover all cases in which G exhibits strong separability. Suppose for example $G = G^A + a_{M+1}y_{M+1}$, then $P_{M+1} = a_{M+1}e^{V_{M+1}/\phi}/G$ and $((G-G^A)/G)(<e^{V_j/\phi}>) = P_{M+1}$. Thus $(G/G^A)(<e^{V_j/\phi}>) = (1 - P_{M+1})^{-1}$ as in (34).

We now consider the conditional moment of the product of two generalized extreme value random variables. Rather than calculate $E[\varepsilon_1\varepsilon_2|\delta_1(\underset{\sim}{\varepsilon}) = 1]$ we will alternatively find $E[(\varepsilon_2 - \varepsilon_1)^2|\delta_1(\underset{\sim}{\varepsilon}) = 1]$ and use the relation $(\varepsilon_2 - \varepsilon_1)^2 = \varepsilon_2^2 - 2\varepsilon_1\varepsilon_2 + \varepsilon_1^2$ along with Theorems 3 and 4. The difference $(\varepsilon_2 - \varepsilon_1)$ has the well known logistic distribution when ε_1 and ε_2 are independent identically extreme value distributed. Our next result finds the joint distribution function for $(Y_2, Y_3, \ldots, Y_J) = (\varepsilon_2 - \varepsilon_1, \varepsilon_3 - \varepsilon_1, \ldots, \varepsilon_j - \varepsilon_1)$ when $\underset{\sim}{\varepsilon}$ has the generalized extreme value distribution.

Theorem 5. Generalized Logistic Distribution

Let $Y_j = \varepsilon_j - \varepsilon_1$ for j = 2, 3, ..., J where $\underset{\sim}{\varepsilon}$ has the generalized extreme vaue distribution given by G(y) and equation (7). Then:

$$H[w_2, w_3, \ldots, w_J] = \text{Prob}[Y_2 \leq w_2, Y_3 \leq w_3, \ldots, Y_J \leq w_J]$$

$$= G_1[<e^{-w_j/\phi}>]/G[<e^{-w_j/\phi}>] \qquad \text{where } w_1 \equiv 0.$$

Proof Theorem 5

$$H = \text{Prob}[Y_2 \leq w_2, \ldots, Y_J \leq w_J]$$

$$= \int_{\varepsilon_1=-\infty}^{\infty} \int_{\varepsilon_2=-\infty}^{\varepsilon_1+w_2} \cdots \int_{\varepsilon_J=-\infty}^{\varepsilon_1+w_J} dF(\underset{\sim}{\varepsilon})$$

$$= \int_{\varepsilon_1=-\infty}^{\infty} F_1[\varepsilon, \varepsilon+w_2, \ldots, \varepsilon+w_J]d\varepsilon$$

$$= \int_{\varepsilon=-\infty}^{\infty} \exp\left[-G[<e^{-(\varepsilon+w_j)/\phi}>]\right] G_1[<e^{(-\varepsilon-w_j)/\phi}>]e^{-\varepsilon/\phi} \frac{d\varepsilon}{\phi}$$

$$= \int_{\varepsilon=-\infty}^{\infty} \exp\left[-e^{-\varepsilon/\phi} G[<e^{-w_j/\phi}>]\right] G_1[<e^{-w_j/\phi}>] e^{-\varepsilon/\phi} \frac{d\varepsilon}{\phi}$$

$$= \frac{G_1[<e^{-w_j/\phi}>]}{G[<e^{-w_j/\phi}>]} \qquad \text{Q.E.D.}$$

Two familiar results follow immediately from Theorem 5.

Corollary 3

C3a) $H\,[V_1-V_2, V_1-V_3, \ldots, V_1-V_J] = P_1$

C3b) $(Y_2, Y_3, \ldots, Y_J)$ has the logistic distribution when

$G[y] = \sum_{j=1}^{J} y_j$.

Proof Corollary 3:

C3a) $$H[V_1-V_2, \ldots, V_1-V_J] = \frac{G_1[<e^{-(V_1-V_j)/\phi}>]}{G[<e^{-(V_1-V_j)/\phi}>]}$$

$$= e^{V_1}\, G_1[<e^{V_j/\phi}>]/G[e^{V_j/\phi}>]$$

$$= P_1$$

where the first equality applies the result of Theorem 5, the second equality applies the homogeneity properties of G, and the third equality applies (8).

C3b) Since $G[y] = \sum_{j=1}^{J} y_j$, $G_1[y] = 1$. Theorem 5 implies

$H[w_2, \ldots, w_J] = 1/\sum_{j=1}^{J} e^{-w_j/\phi}$ which is the multivariate logistic distribution. Q.E.D.

Theorem 6

Let $\underset{\sim}{\varepsilon}$ be generalized extreme value distributed with $G[y] = \alpha y_1 + \alpha y_2 + \alpha G^A[<y_j^A>]$ where G^A is homogeneous of degree one and where y =

(y_1, y_2, y^A). Then:

$$E\left[(\varepsilon_2 - \varepsilon_1)^2 \middle| \delta_1(\underset{\sim}{\varepsilon}) = 1\right] =$$

$$= \phi^2[\ln((1-P_2)/P_1)]^2 - 2\phi^2 \ln((1-P_2)/P_1)\left(P_2 \ln P_2/(1-P_2) + \ln(1-P_2)\right)$$

$$+ \phi^2/(1-P_2) \int_{-\infty}^{\ln((1-P_2)/(P_2))} h(z)dz \text{ where } h(z) = \frac{z^2 e^{-z}}{[1+e^{-z}]^2}.$$

Comment: We have assumed that ε_1 and ε_2 are independent from each other and from ε^A by necessity. A closed form solution for the conditional cross moment will not exist under weaker assumptions.

<u>Proof Theorem 6:</u>

$$E[(\varepsilon_2 - \varepsilon_1)^2 \mid \delta_1(\underset{\sim}{\varepsilon}) = 1]$$

$$= \frac{1}{P_1} \int_{\varepsilon_1=-\infty}^{\infty} \int_{\varepsilon_2=-\infty}^{V_1-V_2+\varepsilon_1} (\varepsilon_2 - \varepsilon_1)^2 F_{12}[\varepsilon_1, \varepsilon_2, V_1-V_3+\varepsilon_1, \dots,$$

$$V_1-V_J+\varepsilon_1]d\varepsilon_2 d\varepsilon_1$$

We now make a logistic transformation: $z_1 \longleftarrow \varepsilon_1$, $z_2 \longleftarrow \varepsilon_2 - \varepsilon_1$.

It is easily verified that this transformation has unit Jacobian. Thus:

$$E[(\varepsilon_2 - \varepsilon_1)^2 \mid \delta_1(\underset{\sim}{\varepsilon}) = 1]$$

$$= \frac{1}{P_1} \int_{z_1=-\infty}^{\infty} \int_{z_2=-\infty}^{V_1-V_2} z_2^2 F_{12}[z_1, z_1+z_2, V_1-V_3+z_1, \dots,$$

$$V_1-V_J+z_1]dz_2 dz_1$$

$$= \frac{1}{P_1} \int_{z_2=-\infty}^{V_1-V_2} z_2^2 \int_{z_1=-\infty}^{\infty} F_{12}[z_1,\ z_1+ z_2\ V_1-V_3+z_1,\ \ldots,\ V_1- V_J+z_1]dz_1dz_2$$

Let $H[w_2, \ldots, w_J] = \int_{\varepsilon=-\infty}^{\infty} F_1[\varepsilon, \varepsilon+w_2, \ldots, \varepsilon+w_J]d\varepsilon$. Then:

$$E[(\varepsilon_2 - \varepsilon_1)^2 \mid \delta_1(\underset{\sim}{\varepsilon}) = 1] = \frac{1}{P_1} \int_{z_2=-\infty}^{V_1-V_2} z_2^2 \cdot H_2[z_2, V_1-V_3, \ldots, V_1- V_J]dz_2$$

Since $G[y_1, y_2, \ldots, y_J] = \alpha y_1 + \alpha y_2 + \alpha G^A[<y_j^A>]$, $G_1 = \alpha$ and by Theorem 5:

$$H[w_2, \ldots, w_J] = \alpha \left[\alpha + \alpha e^{-w_2/\phi} + \alpha G^A[<e^{-w_j/\phi}>]\right]^{-1}$$

Thus $$H_2[w_2, \ldots, w_J] = e^{-w_2/\phi}\left[1 + G^A[<e^{-w_j/\phi}>] + e^{-w_2/\phi}\right]^{-2}$$

and $$E[(\varepsilon_2 - \varepsilon_1)^2|\delta_1 = 1] =$$

$$\frac{\phi^2}{P_1} \int_{y=-\infty}^{(V_1-V_2)/\phi} y^2e^{-y}/[A + e^{-y}]^2\, dy = \frac{\phi^2}{P_1A^2} \int_{y=-\infty}^{(V_1-V_2)/\phi} y^2e^{-y}/[1+e^{-\ln A-y}]^2dy$$

where $A = 1 + G^A[<e^{-(V_1-V_j)/\phi}>]$ and we have made the transformation $z_2/\phi \longrightarrow y$. Note that:

$$(1-P_2)/(P_1) = \frac{G - \alpha e^{V_2/\phi}}{\alpha e^{V_1/\phi}} = \frac{\alpha e^{V_1/\phi} + \alpha G^A}{\alpha e^{V_1/\phi}} = 1 + e^{-V_1/\phi} \bullet G^A[<e^{V_j/\phi}>] = A$$

Let $z = y + \ln A$. Then:

$$E(Y_2^2|\delta_1 = 1) = \frac{\phi^2}{P_1 A^2} \cdot \int_{-\infty}^{((V_1-V_2)/\phi)+\ln A} \frac{(z - \ln A)^2 A e^{-z}}{[1 + e^{-z}]^2} dz$$

$$E[Y_2^2|\delta_1 = 1] = \frac{\phi^2}{P_1 A} \cdot \int_{-\infty}^{((V_1-V_2)/\phi)+\ln A} \frac{[z^2 - 2z \ln A + (\ln A)^2]e^{-z}}{[1 + e^{-z}]^2} dz$$

Since $(V_1-V_2)/\phi = \ln(P_1/P_2)$ and $A = (1-P_2)/P_1$ it follows that

$$(V_1-V_2)/\phi + \ln A = \ln (P_1/P_2) + \ln ((1-P_2)/P_1) = \ln ((1-P_2)/P_2)$$

Let $x = \ln((1-P_2)/P_2)$. It follows that $E[Y_2^2|\delta_1 = 1]$

$$= \frac{\phi^2}{P_1 A} \int_{-\infty}^{x} \frac{z^2 e^{-z}}{[1 + e^{-z}]^2} dz + \frac{\phi^2}{P_1 A} \int_{-\infty}^{x} \frac{e^{-z}}{[1 + e^{-z}]^2} dz \cdot (\ln A)^2$$

$$\frac{-2(\ln A)\phi^2}{P_1 A} \int_{-\infty}^{x} \frac{e^{-z}}{[1 + e^{-z}]^2} dz$$

$$= \frac{\phi^2}{(1-P_2)} \int_{-\infty}^{x} \frac{z^2 e^{-z}}{[1 + e^{-z}]^2} dz + \phi^2(\ln((1-P_2)/P_1))^2$$

$$\frac{-2(\ln A)\phi^2}{(1-P_2)} \int_{-\infty}^{x} \frac{z e^{-z}}{[1 + e^{-z}]^2} dz$$

We use the result that: (Integration by Parts)

$$\int_{t=-\infty}^{x} \frac{te^{-t}dt}{(1+e^{-t})^2} = \frac{x}{1+e^{-x}} - \ln[1+e^{x}]$$

$$\int_{t=-\infty}^{\ln((1-P_2)/P_2)} \frac{te^{-t}dt}{(1+e^{-t})^2} = \frac{\ln((1-P_2)/P_2)}{(1-P_2)^{-1}} + \ln P_2 = \left[P_2 \ln P_2 + (1-P_2)\ln(1-P_2)\right]$$

Hence: $E(Y_2^2 \mid \delta_1(\underset{\sim}{\varepsilon}) = 1) = \phi^2(\ln((1-P_2)/P_1))^2$

$$- \frac{2\phi^2}{(1-P_2)} \ln((1-P_2)/P_1)[P_2 \ln P_2 + (1-P_2)\ln(1-P_2)]$$

$$+ \frac{\phi^2}{(1-P_2)} \bullet \int_{-\infty}^{\ln((1-P_2)/P_1)} h(z)dz$$

$$E(Y_2^2 \mid \delta_1(\underset{\sim}{\varepsilon}) = 1) = \phi^2(\ln((1-P_2)/P_1))^2 - 2\phi^2\ln((1-P_2)/P_1)\left[P_2\ln P_2/(1-P_2) + \ln(1-P_2)\right]$$

$$+ \frac{\phi^2}{(1-P_2)} \bullet \int_{-\infty}^{\ln((1-P_2)/P_2)} h(z)dz \qquad \text{Q.E.D.}$$

<u>Theorem 7</u>.

Let $G[y] = \alpha \sum_{j=1}^{J} y_j$ and let $\alpha^* = \phi \ln\alpha$. Then:

$E[(\varepsilon_1\varepsilon_2)^2 \mid \delta_1(\underset{\sim}{\varepsilon}) = 1]$

$$= \Big[\frac{1}{6}\pi^2\phi^2 + (\alpha^* + \gamma\phi)^2 - \left[P_2\phi^2/(1-P_2)\right](\ln P_2)^2$$

$$+ 2(\alpha^* + \gamma\phi)(\phi\ln P_2)\ P_2/(1-P_2)$$

$$+ \frac{1}{6}\pi^2\phi^2 + (\alpha^* + \gamma\phi)^2 + \phi^2(\ln P_1)^2 - 2(\alpha^* + \gamma\phi)(\phi \ln P_1)$$

$$- \phi^2(\ln((1-P_2)/P_1))^2 + 2\phi^2 \ln((1-P_2)/P_1)\, P_2 \ln P_2/(1-P_2) + \ln(1-P_2)$$

$$\left. - \frac{\phi^2}{(1-P_2)} \int_{-\infty}^{\ln((1-P_2)/P_2)} \frac{z^2 e^{-z}}{[1 + e^{-z}]^2}\, dz \right] \left(\frac{1}{2}\right)$$

Proof Theorem 7: Note that $\varepsilon_1\varepsilon_2 = (\varepsilon_1^2 + \varepsilon_2^2 - y_2^2)\frac{1}{2}$, and use the results of Corollary 2 and Theorem 6 after applying conditional expectations. Q.E.D.

The integral $\int_{-\infty}^{x} h(z)dz$ where $h(z) = \dfrac{z^2 e^{-z}}{(1 + e^{-z})^2}$ is in fact related to $E[y^2 | y < x]$ where y has a univariate logistic distribution. A closed form expression for this distribution does not exist. It is however related to a series expansion involving terms in the incomplete gamma distribution. See Hay (1980) and Lee (1981). Using an alternate series expansion we provide a more useful form of the integral.

Theorem 8

$$\int_0^{\ln \lambda^{-1}} \frac{u^2 e^{-u}}{(1 + e^{-u})^2}\, du = \frac{\pi^2}{6} - \frac{\lambda(\ln \lambda)^2}{(1+\lambda)} - 2(\ln\lambda)(\ln(1+\lambda))$$

$$+ 2 \sum_{i=0}^{\infty} (-1)^i \frac{\lambda^{i+1}}{(i+1)^2}$$

For $\ln \lambda^{-1} > 0$ or $\lambda^{-1} > 1$ or $0 < \lambda < 1$.

Proof 8:

From the formula for the sum of a geometric series we have $(1+x)^{-1} = \sum_{i=0}^{\infty} (-1)^i x^i$ for $|x| < 1$. Differentiating and integrating term by term provides two useful relations:

$$\frac{1}{(1+x)^2} = \sum_{i=1}^{\infty} (-1)^{i+1}\, i\, x^{i-1} = \sum_{i=0}^{\infty} (-1)^i (i+1) x^i \text{ for } |x| < 1$$

and

$$\ln(1+x) = \sum_{i=0}^{\infty} \frac{(-1)^i}{(i+1)} x^{i+1}$$

Take $x = e^{-u}$ for $u > 0$, then

$$\frac{1}{(1 + e^{-u})^2} = \sum_{i=0}^{\infty} (-1)^{i+1} (i+1) e^{-ui} \text{ . Thus:}$$

$$\int_0^{\ln \lambda^{-1}} \frac{u^2 e^{-u}}{(1+e^{-u})^2}\, du = \int_0^{\ln \lambda^{-1}} u^2 \sum_{i=0}^{\infty} (-1)^i (i+1) e^{-u(i+1)} du$$

$$= \sum_{i=0}^{\infty} (-1)^i (i+1) \int_0^{\ln \lambda^{-1}} u^2 e^{-u(i+1)} du$$

Next use the fact that $\int y^2 e^{-iy} dy = \frac{-1}{i}\left[y^2 + \frac{2}{i} y + \frac{2}{i^2}\right] e^{-iy}$ so that:

$$\int_0^{\ln \lambda^{-1}} \frac{u^2 e^{-u}}{(1+e^{-u})^2}\, du$$

$$= \sum_{i=0}^{\infty} (-1)^i (i+1) \left[\frac{-1}{(i+1)} y^2 + \frac{2}{(i+1)} y + \frac{2}{(i+1)^2} \right] e^{-(i+1)y} \Bigg|_0^{\ln \lambda^{-1}}$$

$$= \sum_{i=0}^{\infty} (-1)^{i+1}\left[\left[(\ln \lambda^{-1})^2 + \frac{2}{(i+1)} \ln \lambda^{-1} + \frac{2}{(i+1)^2}\right]\lambda^{i+1} - \frac{2}{(i+1)^2}\right]$$

$$= -2\left[\sum_{i=0}^{\infty} (-1)^{i+1}/(i+1)^2\right] + (\ln \lambda^{-1})^2 \bullet \sum_{i=0}^{\infty} \left[(-1)^{i+1} \lambda^{i+1}\right]$$

$$+ 2\left[\ln \lambda^{-1}\right]\sum_{i=0}^{\infty} (-1)^{i+1}/(i+1)\ \lambda^{i+1} + 2 \sum_{i=0}^{\infty} (-1)^{i+1} \lambda^{i+1}/(i+1)^2$$

$$= \frac{\pi^2}{6} - \left[\frac{\lambda\ (\ln \lambda)^2}{(1+\lambda)} - 2(\ln\lambda)(\ln(1+\lambda)) + 2 \sum_{i=0}^{\infty} (-1)^i \lambda^{i+1}/(i+1)^2\right]$$

where we have used the fact that:

$$\sum_{i=0}^{\infty} (-1)^i/(i+1)^2 = \pi^2/12. \qquad \text{Q.E.D.}$$

For reference below we let:

$$G(\lambda) = \left[\frac{\lambda(\ln \lambda)^2}{(1+\lambda)} - 2(\ln\lambda)\ln(1+\lambda) + 2 \sum_{i=0}^{\infty} (-1)^i \lambda^{i+1}/(i+1)^2\right]$$

Application of Theorem 6 for the case of binary alternatives gives:

<u>Theorem 9</u>

Consider the case in which m = 2. Then

$$E(y_2^2 \mid \delta_1=1) = \begin{cases} \phi^2/P_1\ [\pi^2/3 - G(P_2/P_1)] & \text{for } P_1 > P_2 \\ \phi^2/P_1\ [G(P_1/P_2)] & \text{for } P_1 < P_2 \\ \phi^2/P_1\ [\pi^2/6] & \text{for } P_1 = P_2 \end{cases}$$

Proof Theorem 9

Using Theorem 6, $E(y_2^2 | \delta_1 = 1) = \frac{\phi^2}{P_1} \int_{-\infty}^{\ln(P_1/P_2)} h(z)dz$ where we have imposed the restriction $P_1 + P_2 = 1$ implied by this case of binary alternatives. For $P_1 > P_2$:

$$E(y_2^2 | \delta_1 = 1) = \frac{\phi^2}{P_1} \int_{-\infty}^{0} h(z)dz + \frac{\phi^2}{P_1} \int_{0}^{\ln(P_1/P_2)} h(z)dz$$

We let $\lambda^{-1} = P_1/P_2$ so that $\lambda = P_2/P_1$. Application of Theorem 8 implies $E(y_2^2 | \delta_1 = 1) = \phi^2/P_1[\pi^2/6] + \phi^2/P_1\left[\pi^2/6 - G(P_2/P_1)\right]$.

For $P_1 < P_2$:

$$E(y_2^2 | \delta_1 = 1) = \frac{\phi^2}{P_1} \int_{-\infty}^{0} h(z)dz - \frac{\phi^2}{P_1} \int_{\ln(P_1/P_2)}^{0} h(z)dz$$

$$= \frac{\phi^2}{P_1}\frac{\pi^2}{6} - \frac{\phi^2}{P_1}[\pi^2/6 - G(P_1/P_2)] = \frac{\phi^2}{P_1} G(P_1/P_2)$$

Finally, at $P_1 = P_2$, note that $G(1) = \pi^2/6$ implies continuity for $E(y_2^2 | \delta = 1)$. Q.E.D.

We now introduce a random variable η and suppose that conditional on ε, η has mean $\frac{\sqrt{6}\,\sigma}{\pi\phi} \sum_{i=1}^{m} R_i \varepsilon_i$ and variance $\sigma^2(1 - \sum_{i=1}^{m} R_i^2)$ with $\sum_{i=1}^{m} R_i = 0$ and $\sum_{i=1}^{m} R_i^2 < 1$.

It will be convenient to assume that $<\varepsilon_i>$ are independently, identically extreme value distributed and that $E(\varepsilon_i) = 0$.
From Theorem 2, this is accomplished by assuming that the location parameter $\alpha = -\gamma\phi$. Note that $\frac{\sqrt{6}}{\pi}\frac{\sigma}{\phi} = \frac{\sigma}{\sigma_\varepsilon}$ where σ_ε is the square root of the variance of ε_i. Unconditional moments are presented in Theorem 10.

Theorem 10 (Dubin and McFadden)

T10a) $E(\eta) = 0$

T10b) $E(\eta)^2 = \sigma^2$

T10c) $\text{Correl}\ (\eta, \varepsilon_i) = R_i$

Proof Theorem 10:

T10a) $$E(\eta) = \underset{\underset{\sim}{\varepsilon}}{E}\ [\underset{\eta}{E}(\eta|\underset{\sim}{\varepsilon})] = \underset{\underset{\sim}{\varepsilon}}{E}\left(\frac{\sigma}{\sigma_\varepsilon} \sum_{i=1}^{m} R_i\varepsilon_i\right) = 0$$

T10b) $$E(\eta^2|\underset{\sim}{\varepsilon}) = \text{var}\ (\eta|\underset{\sim}{\varepsilon}) + (E(\eta|\underset{\sim}{\varepsilon}))^2$$

$$E(\eta^2) = \underset{\underset{\sim}{\varepsilon}}{E}\left[\sigma^2(1 - \sum_{i=1}^{m} R_i^2) + \left(\frac{\sigma}{\sigma_\varepsilon} \sum_{i=1}^{m} R_i\varepsilon_i\right)^2\right]$$

$$= \sigma^2(1 - \sum_{i=1}^{m} R_i^2) + \frac{\sigma^2}{\sigma_\varepsilon^2} \sum_{i=1}^{m} R_i^2\sigma_\varepsilon^2 = \sigma^2$$

T10c) $$E(\eta\varepsilon_i) = \underset{\underset{\sim}{\varepsilon}}{E}\left[\underset{\eta}{E}(\eta\varepsilon_i|\underset{\sim}{\varepsilon})\right] = \underset{\underset{\sim}{\varepsilon}}{E}\left[\varepsilon_i E(\eta|\underset{\sim}{\varepsilon})\right]$$

$$= \underset{\underset{\sim}{\varepsilon}}{E}\left[\varepsilon_i\frac{\sigma}{i\sigma_\varepsilon} \sum_{i=1}^{m} R_i\varepsilon_i\right] = \frac{\sigma}{\sigma_\varepsilon} R_i\sigma_\varepsilon^2 = \sigma\ R_i\sigma_\varepsilon$$

$\text{Correl}\ (\eta, \varepsilon_i) = E(\eta\varepsilon_i)/\sigma\sigma_\varepsilon = R_i$ Q.E.D.

We now derive the first moment of η conditional on the event that a particular alternative is chosen.

Theorem 11 (Dubin and McFadden)

$$E(\eta|\delta_i(\underset{\sim}{\varepsilon}) = 1) = \frac{\sqrt{6}\,\sigma}{\pi}\left[\sum_{j=1}^{m} \frac{R_j P_j}{(1-P_j)} \ln P_j \; - R_i \frac{\ln P_i}{(1-P_i)}\right]$$

Proof Theorem 11

Let $A_i \equiv \{\underset{\sim}{\varepsilon} \,|\, \delta_i(\underset{\sim}{\varepsilon}) = 1\}$ Then:

$$E(\eta|\delta_i = 1) = \frac{1}{P_i} \int_{A_i} E(\eta|\underset{\sim}{\varepsilon}) \prod_{j=1}^{m} f(\varepsilon_j)\, d\underset{\sim}{\varepsilon}$$

$$E(\eta|\delta_i = 1) = \frac{1}{P_i} \int_{A_i} \left(\frac{\sigma}{\sigma_\varepsilon} \sum_{j=1}^{m} R_j \varepsilon_j\right) \prod_{j=1}^{m} f(\varepsilon_j) d\underset{\sim}{\varepsilon}$$

$$= \frac{\sigma}{\sigma_\varepsilon} \sum_{j=1}^{m} \frac{R_j}{P_i} \int_{A_i} \varepsilon_j \prod_{j=1}^{m} f(\varepsilon_j) d\underset{\sim}{\varepsilon}$$

$$= \frac{\sigma}{\sigma_\varepsilon} \sum_{j=1}^{m} E[\varepsilon_j|\delta_i(\underset{\sim}{\varepsilon}) = 1] \,.\, R_j$$

$$= \frac{\sigma}{\sigma_\varepsilon} \sum_{j\neq i}^{m} E[\varepsilon_j|\delta_i(\underset{\sim}{\varepsilon})=1]R_j + \frac{\sigma}{\sigma_\varepsilon} E[\varepsilon_i|\delta_i(\underset{\sim}{\varepsilon})=1]R_i$$

Using the results of Corollary 2:

$$E(\eta|\delta_i = 1) = \frac{\sigma}{\sigma_\varepsilon} \sum_{j\neq i}^{m} \frac{\phi R_j P_j . \ln P_j}{(1-P_j)} - \frac{\sigma}{\sigma_\varepsilon} R_i \phi \ln P_i$$

where we have imposed $\alpha = -\gamma\phi$. Noting that $\sigma_\varepsilon = \frac{\pi\phi}{\sqrt{6}}$, we have:

$$E(\eta|\delta_i(\underset{\sim}{\varepsilon}) = 1) = \frac{\sqrt{6}\,\sigma}{\pi}\left[\left(\sum_{j\neq i}^{m} \frac{R_j P_j . \ln P_j}{(1-P_j)}\right) - R_i \ln P_i\right]$$

$$= \frac{\sqrt{6}\,\sigma}{\pi}\left[\left(\sum_{j=1}^{m} \frac{R_j P_j . \ln P_j}{(1-P_j)}\right) - \frac{R_i \ln P_i}{(1-P_i)}\right] \qquad \text{Q.E.D.}$$

Let δ_{ij} be the Kronecker delta. Then we may rewrite the result of Theorem 11 as:

$$E(\eta|\delta_i(\underset{\sim}{\varepsilon}) = 1) = \frac{\sqrt{6}\,\sigma}{\pi}\left[\left(\sum_{j\neq i}^{m} \frac{R_j P_j . \ln P_j}{(1-P_j)}\right) + \frac{R_i . \ln P_i (P_i - 1)}{(1-P_i)}\right]$$

$$= \frac{\sqrt{6}\,\sigma}{\pi}\left[\sum_{j=1}^{m} \frac{R_j . \ln P_j}{(1-P_j)} (P_j - \delta_{ij})\right].$$

We now consider the conditional second moments of η. Recall that

$E(\eta^2|\delta_i = 1) = \frac{1}{P_i} \cdot \int_{A_i} E(\eta^2|\underset{\sim}{\varepsilon}) f(\underset{\sim}{\varepsilon}) d\underset{\sim}{\varepsilon}$ where $f(\underset{\sim}{\varepsilon}) = \prod_{i=1}^{m} f(\varepsilon_i)$. We use the relation $E(\eta^2|\underset{\sim}{\varepsilon}) = Var(\eta|\underset{\sim}{\varepsilon}) + (E(\eta|\underset{\sim}{\varepsilon}))^2 = \sigma^2(1 - \sum_{i=1}^{2} R_i^2) + \frac{\sigma^2}{\sigma_\varepsilon^2}(\sum_{i=1}^{2} R_i\varepsilon_i)^2$

to obtain:

$$E(\eta^2|\delta_i = 1) = \sigma^2(1 - \sum_{t=1}^{m} R_t^2) + \frac{\sigma^2}{\sigma_\varepsilon^2}\sum_{t=1}^{m} R_t^2\, E(\varepsilon_t^2 \,|\, \delta_i = 1)$$

$$+ \frac{2\sigma^2}{\sigma_\varepsilon^2} \sum_{\substack{t=1 \\ s>t}}^{m} R_t R_s E(\varepsilon_t \varepsilon_s \,|\, \delta_i = 1)$$

We continue with the case in which $m = 2$:

Theorem 12 (Dubin and McFadden)

$$E(\eta^2 | \delta_1(\underset{\sim}{\varepsilon})) = \sigma^2 + 2\sigma^2 R_2^2 H(P_1,\delta_1) \qquad \text{where } H(P_1,\delta_1) =$$

$$\begin{cases} 1/P_1 - 1 - 3/\pi^2 (1/P_1) \bullet G[\frac{(1-P_1)}{P_1}] & \text{if } \delta_1 = 1 \text{ and } P_1 > 1/2 \\ -1 + 3/\pi^2 (1/P_1) \bullet G[\frac{P_1}{(1-P_1)}] & \text{if } \delta_1 = 1 \text{ and } P_1 \leq 1/2 \\ -1 + 3/\pi^2 (1/(1-P_1))G[\frac{(1-P_1)}{P_1}] & \text{if } \delta_1 = 0 \text{ and } P_1 > 1/2 \\ P_1/(1-P_1) - 3/\pi^2 (1/(1-P_1))G[\frac{P_1}{(1-P_1)}] & \text{if } \delta_1 = 0 \text{ and } P_2 \leqq 1/2 \end{cases}$$

Proof Theorem 12:

$$E[\eta^2 | \delta_1] = \sigma^2(1 - (R_1^2 + R_2^2)) + \frac{\sigma^2}{\sigma_\varepsilon^2}\Big[R_1^2 E(\varepsilon_1^2 \mid \delta_1=1) +$$

$$R_2^2 E(\varepsilon_2^2 \mid \delta_1=1) + 2R_1R_2(\varepsilon_1\varepsilon_2 | \delta_1=1)\Big]$$

In the binary case $P_1 + P_2 = 1$ and $R_1 + R_2 = 0$. Application of Corollary 2, and Theorems 7 and 9 implies:

$$E[\eta^2 | \delta_1 = 1) = \sigma^2(1 - 2R_2^2) + (\sigma_2^2/\sigma_\varepsilon^2)\cdot R_2^2 \begin{cases} \frac{\phi^2}{P_1}[\pi^2/3 - G((1-P_1)/P_1)] & \text{for } P_1 > 1/2 \\ \frac{\phi^2}{P_1} G(P_1/(1-P_1)) & \text{for } P_1 \leqq 1/2 \end{cases}$$

Using $\sigma_\varepsilon^2 = \pi^2\phi^2/6$ and rewriting, yields the first two parts of the claim. It is then easy to derive the expression for $E(\eta^2|\delta_1 = 0)$ using

$$\left[E(\eta^2 | \delta_1=1)P_1 + E(\eta^2 | \delta_1=0)(1-P_1)\right] = E(\eta^2) = \sigma^2. \quad \text{Q.E.D.}$$

We now relax the assumption that $<\epsilon_i>$ are independently, identically extreme value distributed and assume that $<\epsilon_i>$ have the sequential form of the generalized extreme value family. It has been demonstrated that conditional moments for the generalized extreme value family require quite strong assumptions to insure tractability. Indeed, the strong separability used for the function G in Theorems 4 and 6 if applied symmetrically to all components of G would imply the simple multinomial logit specification. The joint assumption that η have a linear conditional expectation in the space of $<\epsilon_i>$ and that $<\epsilon_i>$ are not independently, identically extreme value distributed goes beyond computational feasibility.

A simple alternative for the sequential form of the generalized extreme value family assumes that η has a linear conditional expectation in the space of the "induced" independent extreme value random variables which generate the conditional probabilities. This assumption is motivated by two considerations: (i) multinomial logit models tend to "robustly" fit data generated within a non-independent error system and (ii) that the simple multinomial logit probability form is implied by but does not imply an independent extreme value error structure. The first observation comes from a growing body of econometric and Monte-Carlo evidence while the second observation is usefully illustrated by the bivariate extreme value distribution:

(47) $$G(y) = \left[y_1^{1/(1-\sigma)} + y_2^{1/(1-\sigma)} \right]^{(1-\sigma)}$$

The probability choice system for (47) implies:

(48) $$P_1 = e^{V_1/\phi(1-\sigma)} \Big/ (e^{V_1/\phi(1-\sigma)} + e^{V_2/\phi(1-\sigma)})$$

which is observationally equivalent to the multinomial logit probability choice system:

(49) $$P_1 = e^{V_1/\phi} / (e^{V_1/\phi} + e^{V_2/\phi})$$

since the scale parameters $\phi(1-\sigma)$ and ϕ are <u>not</u> identified in (48) and (49) respectively. Equation (49) is generated by the independent form of the generalized extreme value family by:

(50) $$G[y] = y_1 + y_2 \; .$$

Equation (47) implies that the stochastic components of utility are correlated while (50) implies independence; yet the binary probabilistic choice systems are observationally equivalent.

To illustrate the methodology consider the nested logit model (39). The second level conditional probabilities in (41) may be thought of being generated by the independent extreme value random variables $<\varepsilon_j^{B_m}>$ with variance $(\pi^2/6)(1 - \sigma_m)^2$. Specifically,

(51) $$P[i \mid B_m] = \text{Prob } [V_i + \varepsilon_i^{B_m} \geqq V_j + \varepsilon_j^{B_m} \text{ for } i,j \in B_m \text{ and } j \neq i].$$

Finally, suppose that $\eta = \sum_{i=1}^{M} \lambda_i \eta_i$ where:

(52) $$E[\eta_i \mid <\varepsilon_j^{B_i}>] = \left(\sum_{j \in B_m} \varepsilon_j^{B_m} \cdot R_j^{B_m} \right) \cdot \frac{\sigma^2}{(1-\sigma_m)^2}$$

Equation (52) implies an error structure which may be analyzed through Theorems 10, 11, and 12.

Footnotes

1. In the course of the exposition several theorems related to the independent form of the generalized extreme value family, i.e., the multinomial logit model, are derived. Specifically, Corollary 2 and Theorems 8, 10, 11, and 12, which involve conditional moments in the multinomial logit model, have been derived jointly with Daniel McFadden and are presented in Dubin and McFadden (1979). It should further be noted that Theorems 8, 10, and 11 have been independently demonstrated by Hay (1980).

Appendix III. Two-Stage Single Equation Estimation Methods: An Efficiency Comparison

In this appendix we consider various two-stage consistent estimation techniques applied to a single equation. We begin with a linear in parameters form:

(1) $$y_t = f[z_t, \delta_t]\,\beta + \eta_t , \qquad t = 1, 2, \ldots, T$$

where:

β = column vector of K_1 parameters,

z_t = row vector of K_0 explanatory variables,

y_t = scalar dependent variable,

η_t = scalar equation error, and

δ_t = scalar dummy variable.

The function f allows non-linear interaction between the elements of z_t and δ_t and maps into a row vector of structural explanatory variables. We assume that the dummy variable δ_t is determined by a random event and takes the value one to indicate that the latent variable y_t^* is less than zero. Equation (1) and the stochastic specification for y_t^* form a dummy endogenous simultaneous equation system. We now consider several two-step procedures which provide consistent estimates of the parameters β under the assumption that the dummy indicator variable is endogenous.

We define the following matrices:

(2) $$W_\delta = \langle f(z_t, \delta_t)\rangle$$

(3) $$W_p = \langle f(z_t, P_t)\rangle$$

(4) $$W_{\hat{p}} = \langle f(z_t, \hat{P}_t)\rangle$$

The order of the matrices W_δ, W_p, and $W_{\hat{p}}$ is $T \times K_1$. The matrix W_p is constructed by replacing the indicator δ_t in W_δ by its expected value denoted p_t. The matrix $W_{\hat{p}}$ is constructed by replacing the indicator δ_t in W_δ by an estimate of the true probability denoted $\hat{P}_t$.

Define two least squares projections:

$$(5) \qquad W = W_p(W_p'W_p)^{-1} W_p'W_\delta \qquad \text{and}$$

$$(6) \qquad \hat{W} = W_{\hat{p}}(W_{\hat{p}}'W_{\hat{p}})^{-1} W_{\hat{p}}'W_\delta$$

Let $y = \langle y_t \rangle$ and $\eta = \langle \eta_t \rangle$.

We express equation (1) alternately as:

$$(7.0) \qquad y = W_\delta\beta + \nu^0 \qquad \text{where } \nu^0 = \eta$$

$$(7.1) \qquad y = W\beta + \nu^1 \qquad \text{where } \nu^1 = \eta + (W_\delta - W)\beta$$

$$(7.2) \qquad y = \hat{W}\beta + \nu^2 \qquad \text{where } \nu^2 = \eta + (W_\delta - \hat{W})\beta$$

$$(7.3) \qquad y = W_p\beta + \nu^3 \qquad \text{where } \nu^3 = \eta + (W_\delta - W_p)\beta$$

$$(7.4) \qquad y = W_{\hat{p}}\beta + \nu^4 \qquad \text{where } \nu^4 = \eta + (W_\delta - W_p)\beta - (W_{\hat{p}} - W_p)\beta$$

In the presence of correlation between δ_t and η_t, ordinary least squares applied to (7.0) will yield inconsistent estimates of β. We consider in turn the ordinary least squares estimators of equations (7.1) to (7.4). It should be noted that the estimators for (7.2) and (7.4) are viable estimators of equation (1). One would not be able to use the least squares estimates of (7.1) and (7.3) as P_t is unobservable.

Ordinary least squares applied to (7.1) through (7.4) produces:

(8.1) $\hat{\beta}^1 = (W'W)^{-1}(W'y)$

(8.2) $\hat{\beta}^2 = (\hat{W}'\hat{W})^{-1}(\hat{W}'y)$

(8.3) $\hat{\beta}^3 = (W_p'W_p)^{-1}(W_p'y)$

(8.4) $\hat{\beta}^4 = (W_{\hat{p}}'W_{\hat{p}})^{-1}(W_{\hat{p}}'y)$

We observe that (8.1) and (8.2) are instrumental variable estimators with instrument matrices W and $\hat{W}$ respectively. In the first stage of (8.1), the endogenous right-hand side variables in (1) are projected onto the exogenous set of instruments W_p. The resultant instrument matrix is given by W in (5). In the second stage, the instrument matrix is used with (7.0) to obtain:

(9) $\hat{\beta}_{IV} = (W'W_\delta)^{-1}\ (W'y)$

Equation (9) is identical to (8.1) since $(W'W_\delta) = (W'W)$. With this observation, (8.1) and (7.0) imply:

(10) $(\hat{\beta}^1 - \beta) = (W'W_\delta)^{-1}(W'\eta)$

Alternatively, equations (8.1) and (7.1) imply:

$$(\hat{\beta}^1 - \beta) = (W'W)^{-1}(W'\nu^1)$$

However, $W'\nu^1 = W'(\eta + (W_\delta - W)\beta) = W'\eta$ since the residual portion of ν^1, $(W_\delta - W)\beta$, is orthogonal to W.

These comments apply directly for (8.2) and produce the instrumental variable estimator:

(11) $(\hat{\beta}^2 - \beta) = (\hat{W}'W_\delta)^{-1}\ (\hat{W}'\eta)$

The estimators $\hat{\beta}^3$ and $\hat{\beta}^4$ are defined by (8.3) and (8.4). Combining these expressions with (7.3) and (7.4) we obtain:

(12) $(\hat{\beta}^3 - \beta) = (W_p'W_p)^{-1}\, W_p'\nu^3$

(13) $(\hat{\beta}^4 - \beta) = (W_{\hat{p}}'W_{\hat{p}})^{-1}\, W_{\hat{p}}'\nu^4$

Alternatively, equation (12) (or (13)) may be derived by combining (8.3) and (7.0):

$$\hat{\beta}^3 = (W_p'W_p)^{-1}(W_p'y) = (W_p'W_p)^{-1}\,[W_p'W_\delta\beta + W_p'\eta] \quad \text{so that:}$$

$$(\hat{\beta}^3 - \beta) = (W_p'W_p)^{-1}[W_p'W_\delta\beta + W_p'\eta - (W_p'W_p)\beta]$$

$$= (W_p'W_p)^{-1}[W_p'[\eta + (W_\delta - W_p)\beta]]$$

$$= (W_p'W_p)^{-1}\,[W_p'\nu^3]$$

It is important to note that the residual portions of the errors ν^3 and ν^4 are not orthogonal to W_p and $W_{\hat{p}}$ except asymptotically. Since $W_p'\nu^3 \neq W_p'\eta$ it is not possible to interpret (7.3) and and (7.4) as instrumental variable estimators. We refer to (7.3) and (7.4) as reduced form estimators.

We introduce two additional instrumental variable estimators:

(14) $\hat{\beta}^5 = [\omega'Z'W_\delta]^{-1}[\omega'Z'y] = \beta + [\omega'Z'W_\delta]^{-1}[\omega'Z'\eta]$

(15) $\hat{\beta}^6 = [W_{\hat{p}}'X'XW_{\hat{p}}]^{-1}\,[W_{\hat{p}}'X'Xy] = \beta + [W_{\hat{p}}'X'XW_{\hat{p}}]^{-1}[W_{\hat{p}}'X'X\nu^4]$

The instrument matrix for the estimator in (14) is $Z\omega$ where $Z = \langle z_t \rangle$ is order $T \times K_0$ and ω is order $K_0 \times K_1$. The instrument matrix for estimator (15) is $X'XW_{\hat{p}}$ where X is an exogenous matrix of order $L \times T$. The matrices ω and X are specified below. Note that (15) is an instrumental variables estimator of equation (7.4).

We now consider the conditional expectation correction estimator of Amemiya (1979) and Heckman (1973). Assume that η_t in (1) has conditional expectation:

(16) $$E[\eta_t \mid \delta_t] = g(z_t,\delta_t,P_t)\gamma$$

where g is a differentiable function of δ_t and the reduced form variables z_t, P_t and γ is a column vector of K_2 parameters.

We rewrite equation (1) as:

(17) $$y_t = f(z_t,\delta_t)\beta + g(z_t,\delta_t,P_t)\gamma + \nu_t^7$$

where $\nu_t^7 = \eta_t - g(z_t,\delta_t,P_t)\gamma$

When P_t is replaced by its estimate $\hat{P}_t$ we have:

(18) $$y_t = f(z_t,\delta_t)\beta + g(z_t,\delta_t,\hat{P}_t)\gamma + \nu_t^8$$

where $\nu_t^8 = \eta_t - g(z_t,\delta_t,P_t)\gamma + [g(z_t,\delta_t,P_t) - g(z_t,\delta_t,\hat{P}_t)]\gamma$

Notationally, let $W_g = \langle g(z_t,\delta_t,P_t)\rangle$ and $W_{\hat{g}} = \langle g(z_t,\delta_t,\hat{P}_t)\rangle$. W_g and $W_{\hat{g}}$ are of order T x K . Also denote $\tilde{\eta}_t = \eta_t - g(z_t,\delta_t,P_t)\gamma$. Note that $E[\tilde{\eta}_t \mid \delta_t] = 0$. Equations (17) and (18) may be rewritten in matrix form as:

(19) $$y = [W_\delta \vdots W_g][\cdot\tfrac{\beta}{\gamma}\cdot] + \nu^7 \quad \text{and}$$

(20) $$y = [W_\delta \vdots W_{\hat{g}}][\cdot\tfrac{\beta}{\gamma}\cdot] + \nu^8 .$$

We present in Table 1 the various two-stage estimators. We use the notation:

$$W^* = [W_\delta \vdots W_g] , \qquad \hat{W}^* = [W_\delta \vdots W_{\hat{g}}],$$

and $\beta^* = [\dot{\dot{\beta}}_{\gamma}]$.

To derive the asymptotic distribution of each estimator we need the following assumptions:

(A1) f is differentiable;

(A2) β is interior to a compact parameter space;

(A3) z_t is uniformly bounded with a convergent empirical distribution function; and

(A4) $\underset{T \to \infty}{\text{PLIM}} \left(\frac{W_\delta' W_p}{T}\right) = A_1$,

$\underset{T \to \infty}{\text{PLIM}} \left(\frac{W_p' W_p}{T}\right) = A_2$, with A_1 and A_2 positive definite.

From equation (5) we find:

(21) $\text{PLIM}\left(\frac{W'W}{T}\right) = A_1 A_2^{-1} A_1'$.

To demonstrate equation (21) observe that:

(22) $\frac{W'W}{T} = \frac{W'W_\delta}{T} = \left[\frac{W_\delta' W_p}{T}\right]\left[\frac{W_p' W_p}{T}\right]^{-1}\left[\frac{W_p' W_\delta}{T}\right]$

and use the fact that the probability limit of a product is the product of the limits when all limits are finite.

From equation (6) we find:

(23) $\text{PLIM}\left(\frac{\hat{W}'\hat{W}}{T}\right) = \text{LIM}\left(\frac{W'W}{T}\right) = A_1 A_2^{-1} A_1'$

Equation (23) follows from Lemma 4 of Amemiya (1973) and uses the fact that PLIM $\hat{P}_t = P_t$.

When $f(z_t, \delta_t)$ is linear in δ_t, A_1 equals A_2. This follows as:

TABLE 1

Two Stage Estimators For: $y_t = f[z_t, \delta_t]\beta + \eta_t$

(i)	$\hat{\beta}^1 - \beta = (W'W_\delta)^{-1}\,(W'\nu^1)$	I.V.
(ii)	$\hat{\beta}^2 - \beta = (\hat{W}'W_\delta)^{-1}\,(\hat{W}'\nu^2)$	I.V.E.
(iii)	$\hat{\beta}^3 - \beta = (W_p{}'W_p)^{-1}(W_p{}'\nu^3)$	R.F.
(iv)	$\hat{\beta}^4 - \beta = (W_{\hat{p}}{}'W_{\hat{p}})^{-1}(W_{\hat{p}}{}'\nu^4)$	R.F.E.
(v)	$\hat{\beta}^5 - \beta = (\omega'Z'W_\delta)^{-1}(\omega'Z'\nu^5)$	I.V.
(vi)	$\hat{\beta}^6 - \beta = (W_{\hat{p}}{}'X'XW_{\hat{p}})^{-1}(W_{\hat{p}}{}'X'X\nu^6)$	I.V.E.+ R.F.E.
(vii)	$\hat{\beta}^7 - \beta^* = (W^{*\prime}W^*)^{-1}\,(W^{*\prime}\nu^7)$	A.H.
(viii)	$\hat{\beta}^8 - \beta^* = (\hat{W}^{*\prime}\hat{W}^*)^{-1}(\hat{W}^{*\prime}\nu^8)$	A.H.E.

NOTES:

(1) $\nu^1 = \nu^2 = \nu^5 = \eta$

(2) $\nu^3 = \eta + (W_\delta - W_p)\beta$

(3) $\nu^4 = \nu^6 = \eta + (W_\delta - W_p)\beta - (W_{\hat{p}} - W_p)\beta$

(4) $\nu^7 = \tilde{\eta}$

(5) $\nu^8 = \tilde{\eta} + (W_g - W_{\hat{g}})\gamma$

IV: Instrumental Variables
IVE: Instrumental Variables Estimated
RF: Reduced Form
RFE: Reduced Form Estimated
AH: Amemiya-Heckman
AHE: Amemiya-Heckman Estimated

(24) $$f(z_t,\delta_t) = [f_0(z_t)\delta_t,\ f_1(z_t)] \quad \text{implies:}$$

(25) $$E[f(z_t,\delta_t)] = [f_0(z_t)P_t,\ f_1(z_t)] = f(z_t,P_t) \quad \text{so that:}$$

(26) $$A_1 = \text{PLIM}(\frac{W_\delta{}'W_p}{T}) = \text{PLIM}(\frac{W_p{}'W_p}{T}) = A_2 .$$

Furthermore, (23) and (24) imply:

(27) $$\text{PLIM}(\frac{W'W}{T}) = A_2 .$$

For the asymptotic distributions of the Amemiya-Heckman estimators we assume:

(A5) g is differentiable;

(A6) γ is interior to a compact parameter space;

(A7) $$\text{PLIM}(\frac{W^{*}{}'W^{*}}{T}) =$$

$$\begin{bmatrix} \text{PLIM}(\frac{W_\delta{}'W_\delta}{T}) & \vdots & \text{PLIM}(\frac{W_\delta{}'W_g}{T}) \\ \cdots & \vdots & \cdots \\ \text{PLIM}(\frac{W_g{}'W_\delta}{T}) & \vdots & \text{PLIM}(\frac{W_g{}'W_g}{T}) \end{bmatrix}$$

$$= \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_4 \end{bmatrix}$$ with A_3, A_4 and A_5 positive definite.

From (A7) we have:

(28) $$\text{PLIM}(\frac{\hat{W}^{*}{}'\hat{W}^{*}}{T}) = \text{PLIM}(\frac{W^{*}{}'W^{*}}{T}) = \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}$$

Equation (28) follows from the definition of $\hat{W}^{*}$ and uses the consistency of $\hat{P}_t$ together with Lemma 4 of Amemiya (1973).

The two-stage estimators presented in Table 1 have the form:

$$(29) \qquad \hat{\beta}^k - \beta = (W_1^k)^{-1} (W_2^{k\prime} \nu^k)$$

for appropriate choices of the matrices W_1^k and W_2^k. We rewrite equation (29) as:

$$(30) \qquad \sqrt{T}(\hat{\beta}^k - \beta) = \left(\frac{W_1^k}{T}\right)^{-1} \left(\frac{W_2^{k\prime} \nu^k}{T}\right)$$

Under the conditions of the Lindberg-Feller central limit theorem it is possible to show that:

$$(31) \qquad \frac{W_2^{k\prime} \nu^k}{\sqrt{T}} \xrightarrow{L} N\left[0, \text{Lim } E[\frac{1}{T} W_2^{k\prime} \nu^k \nu^{k\prime} W_2^k]\right]$$

We now postulate an error structure for the probability model:

(A8) $P_t = \text{Prob}[y_t^* < 0] = V[z_t, \alpha]$ where V is a given function of the exogenous variables z_t and a column vector of L parameters α.

We suppose that the probability model (A8) is estimated by maximum likelihood. For the maximum likelihood estimator of the parameters α, the following useful approximation results:

Lemma 1

Let $\hat{P}_t$ be the estimated value of P_t i.e. the value of P_t which results when $V[z_t, \alpha]$ is replaced by $V[z_t, \hat{\alpha}]$ where $\hat{\alpha}$ is the maximum likelihood estimate of α. Then:

$$(32) \qquad (\hat{P} - P) = Y' V Y D_0^{-1} (\delta - P) \qquad \text{where:}$$

$$\hat{P} = \langle \hat{P}_t \rangle, \quad P = \langle P_t \rangle, \quad D_0 = \text{diag } P_t(1 - P_t) \quad , \; \delta = \langle \delta_t \rangle,$$

$V = E[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)']$ and:

(33) $$\underset{L \times T}{Y} = \left[\left(\frac{\partial P_1}{\partial \alpha}\right)' \;\vdots\; \left(\frac{\partial P_2}{\partial \alpha}\right)' \;\vdots\; \cdots \;\vdots\; \left(\frac{\partial P_T}{\partial \alpha}\right)' \right]$$

N.B. $V \neq V[z_t, \alpha]$.

Proof Lemma 1

The log likelihood function, L, is given by

(34) $$L = \frac{1}{T} \sum_t \delta_t \ln P_t + (1-\delta_t) \ln(1-P_t)$$

From equation (34):

(35) $$L\alpha = \frac{1}{T} \sum_t \left[\left(\frac{\delta_t}{P_t}\right)\left(\frac{\partial P_t}{\partial \alpha}\right) - \frac{(1-\delta_t)}{(1-P_t)} \left(\frac{\partial P_t}{\partial \alpha}\right) \right]$$

(36) $$L_\alpha' = \frac{1}{T} \sum_t \frac{(\delta_t - P_t)}{P_t(1-P_t)} \left(\frac{\partial P_t}{\partial \alpha}\right)' = \frac{1}{T} Y D_0^{-1} (\delta - P)$$

To complete the derivation we use a first-order Taylor expansion for $\hat{P}_t$ around P_t:

(37) $$\hat{P}_t - P_t \overset{D}{=} \left(\frac{\partial P_t}{\partial \alpha}\right) (\hat{\alpha} - \alpha)$$

and apply the usual asymptotic argument to establish:

(38) $$\hat{\alpha} - \alpha \overset{D}{=} - L_{\alpha\alpha}^{-1} L_\alpha' .$$ Combining (37) and (38), we find:

(39) $$\hat{P}_t - P_t \overset{D}{=} -\left(\frac{\partial P_t}{\partial \alpha}\right) L_{\alpha\alpha}^{-1} L_\alpha'$$

Finally, we substitute the expression for L_α' given in (36) into (39) and use $\dot{V} = -\frac{1}{T} L_{\alpha\alpha}^{-1}$ to obtain the matrix form (32). Q.E.D.

We now consider the binary logit model as an illustration of (A8). To generate a logit probability model, we assume that: $y_t^* = (V_{2t} - V_{1t}) + (\varepsilon_{2t} - \varepsilon_{1t})$ where $V_{jt} = V_j[z_t, \alpha]$ is a given function of z_t and α and where ε_{jt} are random variables independent and identically extreme value distributed with variance $(\pi^2/6)\phi^2$. Note that $\delta_t = 1$ if only if $y_t^* < 0$ so that $V_{1t} + \varepsilon_{1t} > V_{2t} + \varepsilon_{2t}$. It then follows that:

(40) $$P_t = \text{Prob}[\delta_t = 1] = \text{Prob}[V_{1t} + \varepsilon_{1t} > V_{2t} + \varepsilon_{2t}]$$

$$= e^{V_{1t}/\phi}/[e^{V_{1t}/\phi} + e^{V_{2t}/\phi}]$$

$$= 1/[1 + e^{-(V_{1t}-V_{2t})/\phi}]$$

$$= 1/[1 + e^{-V_t}] \qquad \text{where } V_t \equiv (V_{1t} - V_{2t})/\phi.$$

Furthermore, equation (40) implies:

(41) $$\frac{\partial P_t}{\partial \alpha} = P_t(1 - P_t)\left[\frac{\partial V_t}{\partial \alpha}\right]$$

We have demonstrated the following result:

Lemma 2

Let $X = \left[\left(\frac{\partial V_1}{\partial \alpha}\right)' \vdots \left(\frac{\partial V_2}{\partial \alpha}\right)' \vdots \cdot \cdot \cdot \cdot \cdot \vdots \left(\frac{\partial V_T}{\partial \alpha}\right)'\right]$. Then :

(42) $$(\hat{P} - P) \overset{D}{=} D_0 X'VX(\delta - P)$$

when P_t is given by the binary logit model (40).

Proof Lemma 2

For the binary logit model (41) implies that $Y = XD_0$. Substituting into (32) proves the Lemma. Q.E.D.

Note that when $Y = XD_0$, $V^{-1} = (XD_0X')$ since $V^{-1} = E[L_\alpha' L_\alpha \cdot T^2] = YD_0^{-1}Y'$ from (36).

Consider the important special case in which $V_j[z_t, \alpha]$ is linear so that $V_j[z_t, \alpha]/\phi = w_{jt}\alpha$ where w_{jt} is an L component row vector of explanatory variables which vary by alternative and observation. Then $V_t = V_1[z_t, \alpha]/\phi - V_2[z_t, \alpha]/\phi = (w_{1t} - w_{2t})\alpha$. Suppose $z_t = [w_{1t}, w_{2t}, w_t^*]$.

$$\text{Then } \partial V_t/\partial\alpha = z_t \begin{bmatrix} I \\ -I \\ 0 \end{bmatrix} \text{ and } X' = \begin{bmatrix} 2V_1/2\alpha \\ \cdots \\ 2V_2/2\alpha \\ \vdots \\ 2V_T/2\alpha \end{bmatrix} = Z\rho \text{ where } \rho = \begin{matrix} \\ \\ K_0 \end{matrix}\begin{bmatrix} I \\ -I \\ 0 \end{bmatrix}^{L} .$$

Throughout the remainder of this section we use the binary logit probability model. To return to the general framework one need simply substitute $X = YD_0^{-1}$.

In Lemma 3 we evaluate the expressions of $E[\nu^k \nu^{k\prime}]$ for the limiting distribution in equation (31).

Lemma 3

Let $E[\eta\eta'] = A$ with A diagonal. Then:

L3a) $E[\nu^3\nu^{3\prime}] = A + 2D_1D_3 + D_1^2D_0$

L3b) $E[\nu^4\nu^{4\prime}] = A + D_1^2 D_0 + 2D_1 D_3$

$$- [D_1 D_0 X'VXD_3 + D_3 X'VXD_0 D_1 + D_1 D_0 X'VXD_0 D_1]$$

L3c) $E[\nu^7\nu^{7\prime}] = A + D_4$

L3d) $E[\nu^8\nu^{8\prime}] = A + D_4 + D_2 D_0 X'VXD_0 D_2$

where $D_1 = \text{diag}\left\{f'(z_t, P_t)\beta\right\}$ $\quad D_3 = \text{diag}\left\{E[\eta_t(\delta_t - P_t)]\right\}$

$D_2 = \text{diag}\left\{g'(z_t, \delta_t, P_t)\gamma\right\}$ $\quad D_4 = \text{diag}\left\{(E[\eta_t|\delta_t])^2\right\}$

<u>Proof Lemma 3</u>

L3a) $\nu_t^3 = \eta_t + [f(z_t, \delta_t) - f(z_t, P_t)]\beta$

We make a first-order Taylor approximation to $f(z_t, .)$ to obtain:

$$f(z_t, \delta_t) - f(z_t, P_t) \overset{D}{=} f'(z_t, P_t)(\delta_t - P_t)$$

where $f'(z_t, P_t) = \partial f(z_t, s)/\partial s \Big|_{s = P_t}$.

Thus $\nu_t^3 \overset{D}{=} \eta_t + (f'(z_t, P_t)\beta)(\delta_t - P_t)$

Let $D_1 = \text{diag}\left\{f'(z_t, P_t)\beta\right\}$ so that $\nu^3 = \eta + D_1(\delta - P)$. Then:

$$E[\nu^3\nu^{3\prime}] = E[(\eta + D_1(\delta - P))(\eta' + (\delta - P)'D_1)]$$

$$= E[\eta\eta'] + D_1 E[(\delta - P)\eta] + E[\eta(\delta - P)']D_1 +$$

$$D_1 E[(\delta - P)(\delta - P)']D_1$$

Now let $D_3 = E[\eta(\delta - P)']$ and note that $E[(\delta - P)(\delta - P)'] = D_0$ since $E[(\delta_t - P_t)^2] = P_t(1 - P_t)$ and $E[(\delta_t - P_t)(\delta_s - P_s)] = 0$ for $t \neq s$.

Thus: $E[\nu^3 \nu^{3\prime}] = A + D_1 D_3 + D_3 D_1 + D_1 D_0 D_1 = A + 2D_1 D_3 + D_1^2 D_0$.

L3b) Recall $\nu_t^4 = \eta_t + (f(z_t, \delta_t) - f(z_t, P_t))\beta$

$$- (f(z_t, \hat{P}_t) - f(z_t, P_t))\beta$$

We use the approximation of Lemma 3a to obtain:

$\nu^4 \overset{D}{=} \eta + D_1(\delta - P) - D_1(\hat{P} - P)$. From Lemma 2:

$$\nu^4 \overset{D}{=} \eta + D_1(\delta - P) - D_1 D_0 X'VX(\delta - P) = \eta + D_1[I - D_0 X'VX](\delta - P)$$

Thus:

$$E(\nu^4 \nu'^4) = E\left([\eta + D_1(I - D_0 X'VX)(\delta - P)] \cdot [\eta' + (\delta - P)'(I - X'VXD_0)D_1] \right)$$

$$= A + D_1[I - D_0 X'VX]D_0[I - X'VXD_0]D_1$$

$$+ D_3[I - D_0 X'VXD_0]D_1 + D_1[I - D_0 X'VX]D_3$$

But $D_1[I - D_0 X'VX]D_0[I - X'VXD_0]D_1 = D_1^2 D_0 - D_1 D_0 X'VXD_0 D_1$ since $V = (XD_0 X')^{-1}$. Finally:

$$E(\nu^4 \nu^{4\prime}) = A + D_1^2 D_0 + 2D_1 D_3 - [D_1 D_0 X'VXD_3 + D_3 X'VXD_0 D_1 + D_1 D_0 X'VXD_0 D_1]$$

L3c) $\nu_t^7 = \tilde{\eta}_t = \eta_t - E[\eta_t | \delta_t]$

since $E[\eta_t] = 0$, $E[\nu^7\nu^{7\prime}] = A + D_4$ where $D_4 = \text{diag}\ \{(E[\eta_t|\delta_t])^2\}$.

L3d) $\nu_t^8 = \tilde{\eta}_t - (g(z_t\ \delta_t, \hat{P}_t) - g(z_t, \delta_t, P_t))\gamma$

We make a first-order Taylor approximation to $g(z_t, \delta_t, .)$ to obtain

$g(z_t,\delta_t,\hat{P}_t) - g(z_t,\delta_t,P_t) = g'(z_t,\delta_t,P_t)(\hat{P}_t - P_t)$ where:

$$g'(z_t,\delta_t,P_t) \overset{D}{=} \partial g(z_t,\delta_t,s)/\partial s \Big|_{s=P_t} .$$

Hence $\nu_t^8 \overset{D}{=} \tilde{\eta}_t - (g'(z_t,\delta_t,P_t)\gamma)(\hat{P}_t - P_t)$.

Let $D_2 = \text{diag}\left\{g'(z_t,\delta_t,P_t)\gamma\right\}$ so that:

$$\nu^8 \overset{D}{=} \tilde{\eta} - D_2(\hat{P} - P) \overset{D}{=} \tilde{\eta} - D_2D_0X'VX(\delta - P)$$

As $E[\tilde{\eta}(\delta - P)'] = \text{diag}\left\{E[\tilde{\eta}_t(\delta_t - P_t)]\right\} = 0$,

$$E[\nu^8\nu^{8\prime}] = A + D_4 + D_2D_0X'VXD_0D_2 \qquad \text{Q.E.D.}$$

From Lemma 3 and equation (31) we are able to find the asymptotic distributions for the two-stage estimators listed in Table 1.

<u>Theorem 1</u>

Let $B_1 = A_1'^{-1}A_2A_1^{-1}$, $A = \sigma^2 I$,

$$B_2 = \text{PLIM}\left[\frac{1}{T} W_p'(2D_1D_3 + D_1^2D_0)W_p\right],$$

$$B_3 = \text{PLIM}\left[\frac{1}{T} W_p'[D_1D_0X'VXD_3 + D_3X'VXD_0D_1 + D_1D_0X'VXD_0D_1]W_p\right]$$

T1a) $\sqrt{T}\ (\hat{\beta}^1 - \beta) \overset{L}{\dashrightarrow} N[0, \sigma^2 B_1]$

T1b) $\sqrt{T}\,(\hat{\beta}^2 - \beta) \overset{L}{\longrightarrow} N[0,\ \sigma^2 B_1]$

T1c) $\sqrt{T}\,(\hat{\beta}^3 - \beta) \overset{L}{\longrightarrow} N[0,\ \sigma^2 A_2^{-1} + A_2^{-1} B_2 A_2^{-1}]$

T1d) $\sqrt{T}\,(\hat{\beta}^4 - \beta) \overset{L}{\longrightarrow} N[0,\ \sigma^2 A_2^{-1} + A_2^{-1} B_2 A_2^{-1} - A_2^{-1} B_3 A_2^{-1}]$

T1e) $\sqrt{T}\,(\hat{\beta}^5 - \beta) \overset{L}{\longrightarrow} N[0,\ \sigma^2 (C_1' C_2^{-1} C_1)^{-1}]$

T1f) $\sqrt{T}\,(\hat{\beta}^6 - \beta) \overset{L}{\longrightarrow} N[0,\ \sigma^2 (C_3' C_3)^{-1} (C_3' C_4 C_3)(C_3' C_3)^{-1}]$.

where $C_1 = \mathrm{PLIM}(\frac{Z'W_\delta}{T})$ $\quad C_2 = \mathrm{PLIM}(\frac{Z'Z}{T})$,

and $C_3 = \mathrm{PLIM}(\frac{Z'W_\delta}{T})$ $\quad C_4 = \mathrm{PLIM}(\frac{XX'}{T})$ and $\omega = (Z'Z)^{-1} Z'W_\delta$.

Let $D_5 = D_2 D_0 X'VXD_0 D_2$

$$\begin{bmatrix} \mathrm{PLIM}(\frac{1}{T} W_\delta' D_4 W_\delta) & \vdots & \mathrm{PLIM}(\frac{1}{T} W_\delta' D_4 W_g) \\ \cdots & \vdots & \cdots \\ \mathrm{PLIM}(\frac{1}{T} W_g' D_4 W_\delta) & \vdots & \mathrm{PLIM}(\frac{1}{T} W_g' D_4 W_g) \end{bmatrix} = \begin{bmatrix} B_4 & \vdots & B_5 \\ \cdots & \vdots & \cdots \\ B_5' & \vdots & B_6 \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{PLIM}(\frac{1}{T} W_\delta' D_5 W_\delta) & \vdots & \mathrm{PLIM}(\frac{1}{T} W_\delta' D_5 W_g) \\ \cdots & \vdots & \cdots \\ \mathrm{PLIM}(\frac{1}{T} W_g' D_5 W_\delta) & \vdots & \mathrm{PLIM}(\frac{1}{T} W_g' D_5 W_g) \end{bmatrix} = \begin{bmatrix} B_7 & \vdots & B_8 \\ \cdots & \vdots & \cdots \\ B_8' & \vdots & B_9 \end{bmatrix}$$

T1g) $$\sqrt{T}\,(\hat{\beta}^7 - \beta^*) \xrightarrow{L} N\left[0,\ \sigma^2 \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1} + \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1} \bullet \begin{bmatrix} B_4 & \vdots & B_5 \\ \cdots & \vdots & \cdots \\ B_4' & \vdots & B_5 \end{bmatrix} \bullet \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1} \right]$$

T1h) $$\sqrt{T}\,(\hat{\beta}^8 - \beta^*) \xrightarrow{L} N\left[0,\ \sigma^2 \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1} + \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1} \begin{bmatrix} B_4+B_7 & \vdots & B_5+B_8 \\ \cdots & \vdots & \cdots \\ B_5'+B_8' & \vdots & B_6+B_9 \end{bmatrix} \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1} \right]$$

Proof Theorem 1

T1a) $$\text{PLIM}\left(\frac{1}{T} W'W_\delta\right)^{-1} = (A_1 A_2^{-1} A_1')^{-1} = A_1'^{-1} A_2 A_1^{-1} = B_1$$

where we have used (21), (22), and the continuity property of matrix inversion. Also,

$$\text{Lim } E\left[\frac{1}{T} W'\nu^1\nu^{1\prime}W\right] = \text{Lim }\left[\frac{1}{T} W'E(\nu^1\nu^{1\prime})W\right] = \sigma^2 B_1^{-1}$$

Finally, write $\sqrt{T}\,(\hat{\beta}^1 - \beta) = \left(\dfrac{W'W_\delta}{T}\right)^{-1}\left(\dfrac{W'\nu^1}{T}\right)$ and apply (31) so that $\sqrt{T}\,(\hat{\beta}^1 - \beta) \longrightarrow N[0,\ \sigma^2 B_1^2]$.

In T1b to T1h we calculate the appropriate probability limits but omit the details relating to the application of (31).

T1b) $\text{PLIM}(\frac{1}{T}\hat{W}'W_{\delta})^{-1} = B_1$ $\qquad$ $\text{Lim } E[\frac{1}{T}\hat{W}'\nu^2\nu^{2\prime}\hat{W}] = \sigma^2 B_1^{-1}.$

T1c) $\text{PLIM}(\frac{1}{T}W_p'W_p)^{-1} = A_2^{-1}$ $\qquad$ $\text{Lim } E[\frac{1}{T}W'\nu^3\nu^{3\prime}W_p] = \sigma^2 A_2 + B_2.$

T1d) $\text{PLIM}(\frac{1}{T}W_{\hat{p}}'W_{\hat{p}})^{-1} = A_2^{-1}$ $\qquad$ $\text{Lim } E[\frac{1}{T}W_{\hat{p}}'\nu^4\nu^{4\prime}W_{\hat{p}}] = \sigma^2 A_2 + B_2 - B_3$

T1e) $$\text{PLIM}\left[\left[\frac{W_{\delta}'Z}{T}\right]\left(\frac{Z'Z}{T}\right)^{-1}\left[\frac{Z'W_p}{T}\right]\right]^{-1} = (C_1'C_2^{-1}C_1)^{-1}$$

$$\text{PLIM}\left[\left[\frac{W_{\delta}'Z}{T}\right]\left(\frac{Z'Z}{T}\right)^{-1}\right] = C_1'C_2^{-1}$$

$$\text{LIM } E\left[\frac{1}{T}Z'\nu^5\nu^{5\prime}Z\right] = \sigma^2 C_2$$

$$\text{PLIM}\left(\left[\frac{W_{\delta}'X'}{T}\right]\left[\frac{XW_p}{T}\right]\right)^{-1} = C_3'C_3.$$

T1f) $$X\nu^6 = X\nu^4 = X\eta + D_1X[I - D_0X'VX](\delta - P)$$

$$= X\eta \text{ since } X[I - D_0X'VX] = 0$$

$$\text{LIM } E\left[\frac{1}{T}X\nu^6\nu^{6\prime}X'\right] = \sigma^2 C_4$$

T1g) $$\text{PLIM}(\frac{1}{T}W^{*\prime}W^*)^{-1} = \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1} \quad \text{from (A7).}$$

$$\text{LIM } E[\frac{1}{T}W^{*\prime}\nu^7\nu^{7\prime}W^*] =$$

$$\text{LIM } E \begin{bmatrix} \frac{1}{T} W_\delta{}'\nu^7\nu^7{}'W_\delta & \vdots & \frac{1}{T} W_\delta{}'\nu^7\nu^7{}'W_g \\ \cdots & \vdots & \cdots \\ \frac{1}{T} W_g{}'\nu^7\nu^7{}'W_\delta & \vdots & \frac{1}{T} W_g{}'\nu^7\nu^7{}'W_g \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix} + \begin{bmatrix} B_4 & \vdots & B_5 \\ \cdots & \vdots & \cdots \\ B_5' & \vdots & B_6 \end{bmatrix} .$$

T1h) $$\text{PLIM}\left(\frac{1}{T} \hat{W}^*{}'\hat{W}^*\right)^{-1} = \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1}$$

$$\text{LIM } E\left[\frac{1}{T} \hat{W}^*{}'\nu^8\nu^8{}'\hat{W}^*\right] =$$

$$= \sigma^2 \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix} + \begin{bmatrix} B_4 & \vdots & B_5 \\ \cdots & \vdots & \cdots \\ B_5' & \vdots & B_6 \end{bmatrix} + \begin{bmatrix} B_7 & \vdots & B_8 \\ \cdots & \vdots & \cdots \\ B_8' & \vdots & B_9 \end{bmatrix}$$

Q.E.D.

Comment: We have taken $\omega = (Z'Z)^{-1}Z'W_\delta$ which is the least squares projection of W_δ onto the linear span of Z. Among instrumental variable estimators of equation (1) which use instruments linear in Z, $\hat{\beta}^5$ in Theorem 1e is optimal having the smallest asymptotic covariance matrix.

It is useful to find the asymptotic distributions of the eight estimators under the null hypothesis in which η and $\underset{\sim}{\varepsilon}$ are uncorrelated. This is accomplished in Corollary 1.

Corollary 1

Let $B_2^N = \text{PLIM}[\frac{1}{T} W_p'(D_1^2 D_0) W_p]$

$B_3^N = \text{PLIM}[\frac{1}{T} W_p'[D_1(D_0 X'VXD_0)D_1]W_p]$

Under the null hypothesis in which η and $\underset{\sim}{\varepsilon}$ are uncorrelated:

C1a) $\sqrt{T}\,(\hat{\beta}^1 - \beta) \xrightarrow{L} N[0,\ \sigma^2 B_1]$

C1b) $\sqrt{T}\,(\hat{\beta}^2 - \beta) \xrightarrow{L} N[0,\ \sigma^2 B_1]$

C1c) $\sqrt{T}\,(\hat{\beta}^3 - \beta) \xrightarrow{L} N[0,\ \sigma^2 A_2^{-1} + A_2^{-1} B_2^N A_2^{-1}]$

C1d) $\sqrt{T}\,(\hat{\beta}^4 - \beta) \xrightarrow{L} N[0,\ \sigma^2 A_2^{-1} + A_2^{-1}(B_2^N - B_3^N)A_2^{-1}]$

C1e) $\sqrt{T}\,(\hat{\beta}^5 - \beta) \xrightarrow{L} N[0,\ \sigma^2 (C_1' C_2^{-1} C_1)^{-1}]$

C1f) $\sqrt{T}\,(\hat{\beta}^6 - \beta) \xrightarrow{L} N\left[0,\ \sigma^2\ (C_3'C_3)^{-1}(C_3'C_4C_3)(C_3'C_3)^{-1}\right]^{-1}$

C1g) $\sqrt{T}\,(\hat{\beta}^7 - \beta^*) \xrightarrow{L} N\left[0,\ \sigma^2 \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1}\right]$

C1h) $\sqrt{T}\,(\hat{\beta}^8 - \beta^*) \xrightarrow{L} N\left[0,\ \sigma^2 \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1}\right]$

Furthermore, B_2^N and $B_2^N - B_3^N$ are positive definite and positive semi-definite matrices respectively.

Proof Corollary 1

Under the null hypothesis, $E[\eta_t | \delta_t] = 0$ so that $D_3 = D_4 = 0$. Since $E[\eta_t | \delta_t] = g(z_t, \delta_t, P_t)\gamma$ it follows that $\gamma = 0$ and hence $D_2 = \text{diag}\left\{g'(z_t, \delta_t, P_t)\gamma\right\} = 0$. Furthermore $D_2 = 0$ implies $D_5 = 0$ so that $B_4 = B_5 = B_6 = B_7 = B_8 = B_9 = 0$. Making the appropriate substitutions in T1a–T1h demonstrate C1a–C1h. Note that the instrumental variable estimators: C1a, C1b, C1e, C1f remain unchanged and that $D_3 = 0$ implies $B_2 = B_2^N$ and $B_3 = B_3^N$. Finally, B_2^N is positive definite since $D_1^2 D_0$ is diagonal with positive terms. To prove that $B_2^N - B_3^N$ is positive semi-definite we write $B_2^N - B_3^N = \text{PLIM}(\frac{1}{T} W_p'[D_1(D_0 - D_0X'VXD_0)D_1)W_p]$ and demonstrate that $D_1(D_0 - D_0X'VXD\ )D_1$ is positive semi-definite. Note that $D_1(D_0 - D_0X'VXD_0)D_1 = D_1D_0^{1/2}[I - D_0^{1/2}X'VXD_0^{1/2}]D_0^{1/2}D_1$, and that the matrix $[I - D_0^{1/2}X'VXD_0^{1/2}]$ is idempotent, and hence positive semi-definite. Q.E.D.

Estimator Efficiency Orderings

1. Comparing T1a with T1b we see that asymptotically estimator one and estimator two have identical distributions. Thus one does no harm asymptotically by using the estimated rather than the actual probabilities. The limiting distributions are identical under the null hypothesis in which η and $\underset{\sim}{\varepsilon}$ are uncorrelated. When $f(z_t, \delta_t)$ is linear as in (24) the limiting distribution for T1a, T1b, C1a, C1b is $N[0, \sigma^2 A_2^{-1}]$.

2. Comparing T1c with T1d we see that asymptotically the distributions of the reduced form estimators are different when estimated rather than actual probabilities are employed. However, it is not possible to determine whether one does better or worse (in the positive definite sense) using the estimated probabilities. The difference of the covariance matrices is indefinite since $V(\hat{\beta}^4) - V(\hat{\beta}^3) = -A_2^{-1}B_3A_2^{-1}$ and B_3 need not be definite. Under the null hypothesis, we see from C1c and C1d that $V(\hat{\beta}^4) - V(\hat{\beta}^3) = -A_2^{-1}B_3^N A_2^{-1}$ which is <u>negative</u> definite when D_1 is scalar.

3. Comparing T1g with T1h we find that the asymptotic covariance matrices differ by a matrix which is positive definite. Hence, the Amemiya-Heckman estimator is more efficient when actual probabilities are used rather than estimated probabilities. To demonstrate this claim we note that D_4 is positive-definite since it is a diagonal matrix with positive terms and that $D_5 = D_2D_0X'VXD_0D_2$ is positive definite since V is the variance-covariance for the estimated logistic parameters $\hat{\alpha}$.

The definitions of $\begin{bmatrix} B_4 & \vdots & B_5 \\ \cdots & \vdots & \cdots \\ B_5' & \vdots & B_6 \end{bmatrix}$ and $\begin{bmatrix} B_7 & \vdots & B_8 \\ \cdots & \vdots & \cdots \\ B_8' & \vdots & B_9 \end{bmatrix}$

imply that each is positive definite as a consequence of the definiteness of D_4 and D_5.

From T1g and T1h we see that $V(\hat{\beta}^8) - V(\hat{\beta}^7) =$

$$= \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1} \begin{bmatrix} B_7 & \vdots & B_8 \\ \cdots & \vdots & \cdots \\ B_8' & \vdots & B_9 \end{bmatrix} \begin{bmatrix} A_3 & \vdots & A_4 \\ \cdots & \vdots & \cdots \\ A_4' & \vdots & A_5 \end{bmatrix}^{-1}$$

which is positive definite. Under the null hypothesis, C1g and C1h indicate that $\hat{\beta}^7$ and $\hat{\beta}^8$ have identical asymptotic distributions.

<u>Efficiency Orderings for the Reduced Form and Instrumental Variable Estimators</u>

1. Comparing T1c with T1a (or T1b), we see that the difference $V(\hat{\beta}^3) - V(\hat{\beta}^1) = A_2^{-1}B_2A_2^{-1} + \sigma^2(A_2^{-1} - B_2)$. When $f(z_t, \delta_t)$ is linear, $A_2^{-1}-B_2 = 0$ so that $V(\hat{\beta}^3) - V(\hat{\beta}^1) = A_2^{-1}B_2A_2^{-1}$. Since $B_2 =$ $\text{PLIM}(\frac{1}{T} W_p'(2D_1D_3 + D_1^2D_0)W_p)$ we cannot determine whether $A_2^{-1}B_2A_2$ is definite. Under the null hypothesis and assuming linearity for $f(z_t, \delta_t)$ we find $V(\hat{\beta}^3) - V(\hat{\beta}^1) = A_2^{-1}B_2A_2^{-1}$ which is positive definite from Corollary 1. Hence the reduced form estimator using known probabilities is less efficient than instrumental variable estimators $\hat{\beta}^1$ (or $\hat{\beta}^2$) under the null hypothesis.

2. Comparing T1d with T1a (or T1b) we see that the difference $V(\hat{\beta}^4) - V(\hat{\beta}^3)$ is indefinite. In this case one cannot determine whether the matrix $A_2^{-1}(B_2 - B_3)A_2^{-1}$ is definite. Under the null hypothesis and assuming linearity for $f[z_t, \delta_t]$ we find that $V(\hat{\beta}^4) - V(\hat{\beta}^1) =$ $A_2^{-1}(B_2^N - B_3^N)A_2^{-1}$ which is positive definite from Corollary 1. Hence the reduced form estimator using estimated probabilities is less efficient than instrumental variable estimators $\hat{\beta}^1$ (or $\hat{\beta}^2$) under the null hypothesis.

3. We now compare the instrumental variable estimators, $\hat{\beta}^2$ and $\hat{\beta}^5$, which differ by choice of instrument matrices. The instrument matrix for $\hat{\beta}^2$ is $W_{\hat{p}}(W_{\hat{p}}'W_{\hat{p}})^{-1}W_{\hat{p}}'W_\delta$ while the instrument matrix for $\hat{\beta}^5$ is $Z(Z'Z)^{-1}Z'W_\delta$.

Intuitively, one would conclude that instruments provided by the "structural" span in W_p would contain more information than those provided by the "reduced form" span in Z since W_δ is expected to be more highly correlated with W_p than with Z. In the case in which $W_\delta = [Z \vdots \delta]$ (so that the dummy indicator variable is isolated in the equation $y = W_\delta \beta + \eta$) and in which P_t is determined by a linear probability model: $P_t = Z_t \Delta$ it can be shown that the instrumental variable estimators $\hat{\beta}^2$ and $\hat{\beta}^5$ have identical limiting distributions. One suspects that a measure of the efficiency differential between the two estimators is provided by the degree of robustness in using a linear probability model to approximate logistic probabilities.

4. Reviewing points one and two above, we have not been able to make a positive statement about the relative efficiency of the reduced form estimators except under the null hypothesis. We can however compare the joint instrumental variable and reduced form estimator $\hat{\beta}^6$ with a pure instrumental variable estimator.

Suppose we use $(X'XW_{\hat{p}})$ with order $T \times K_1$ as instruments for $y = W_\delta \beta + \eta$. The resultant estimator is:

$\hat{\beta} - \beta = [W_{\hat{p}}'X'XW_\delta]^{-1}[W_{\hat{p}}'X'X\eta]$ which has the asymptotic distribution:

$T\ (\hat{\beta} - \beta) \dashrightarrow N[0,\ \sigma^2(C_3'C_3)^{-1}(C_3'C_4C_3)(C_3'C_3)^{-1}]$ which is precisely the asymptotic distribution of $T\ \ (\hat{\beta}^6 - \beta)$. The equivalence of the asymptotic distributions is due to the orthogonality of X and the residual portion of the error term ν^6. Recall:

$$X\nu^6 = X\nu^4 = X[\eta + D_1(I - D_0X'VX)\ .\ (\delta - P)]$$

$$= X\eta + D_1(X - XD_0X'VX)(\delta - P) = X\eta$$

since $(X - XD_0X'VX) = 0$.

We conclude that the reduced form estimator using estimated probabilities differs from a pure instrumental variable estimator by a projection with the matrix X.

Estimator Efficiency Orderings for the Amemiya-Heckman and Instrumental Variable Procedures

The Amemiya-Heckman estimators $\hat{\beta}^7$ and $\hat{\beta}^8$ are least squares estimators of the transformed equations $y = W_\delta\beta + W_g\gamma + \nu^7$ and $y = W_\delta\beta + W_{\hat{g}}\gamma + \nu^8$. We concentrate our attention on the efficiency of estimating β regarding γ as nuisance parameters.

Estimation of β by the Amemiya-Heckman methods implies:

$$(\hat{\beta}^7 - \beta) = (W_\delta'M_gW_\delta)^{-1}(W_\delta'M_g\nu^7) \text{ and}$$

$$(\hat{\beta}^8 - \beta) = (W_\delta'M_{\hat{g}}W_\delta)^{-1}(W_\delta'M_{\hat{g}}\nu^8) \text{ where}$$

$$M_g = [I - W_g(W_g'W_g)^{-1}W_g'] \text{ and } M_{\hat{g}} = [I - W_{\hat{g}}(W_{\hat{g}}'W_{\hat{g}})^{-1}W_{\hat{g}}']$$

Consider the asymptotic distribution of $\hat{\beta}^7$. Since $E[\nu^7\nu^{7\prime}] = \sigma^2 I + D_4$ it follows that:

$$\sqrt{T}\,(\hat{\beta}^7 - \beta) \xrightarrow{L} N[0, \sigma^2E_1 + E_2] \text{ where}$$

$$E_1 = \text{PLIM}(\tfrac{1}{T}\, W_\delta'M_gW_\delta)^{-1} \text{ and}$$

$$E_2 = \text{PLIM}(\tfrac{1}{T}\, W_\delta'M_gD_4M_gW_\delta).$$

Clearly E_2 is positive definite so that $V(\hat{\beta}^7) > \sigma^2 E_1$. Under the null hypothesis, $D_4 = 0$ so that $V(\hat{\beta}^7) = \sigma^2 E_1$ which exceeds the covariance matrix of the ordinary least squares estimator for $y = W_\delta \beta + \eta$.

2. It is not possible to order the Amemiya-Heckman with the instrumental variable estimator. Consider the difference in covariance matrices:

$$V(\hat{\beta}^7) - V(\hat{\beta}^1) = \sigma^2 E_1 + E_2 - \sigma^2 B_1 = \sigma^2 (E_1 - B_1) + E_2$$

When $f(z_t, \delta_t)$ is linear, $B_1 = A_2^{-1}$ and $E_1 - B_1 = \text{PLIM}\,(\frac{1}{T} W_\delta' M_g W_\delta)^{-1} - \text{PLIM}(\frac{1}{T} W_p' W_p)^{-1}$. Now $E_1 - B_1 \geq 0$, if and only if $B_1^{-1} - E_1^{-1} \geq 0$. But $B_1^{-1} - E_1^{-1} = \text{PLIM}(\frac{1}{T} W_p' W_p) - \text{PLIM}(\frac{1}{T} W_\delta' W_\delta) + \text{Plim}(\frac{1}{T} [W_\delta' W_g (W_g' W_g)^{-1} W_g' W_\delta])$ and $\text{PLIM}(\frac{1}{T} W_p' W_p) \leq \text{PLIM}(\frac{1}{T} W_\delta' W_\delta)$ so that B_1^{-1} need not be greater than E_1^{-1} in the positive definite sense.

If as an empirical matter, the difference between $\text{PLIM}(\frac{1}{T} W_p' W_p)$ and $\text{PLIM}(\frac{1}{T} W_\delta' W_\delta)$ is small relative to $\text{PLIM}(\frac{1}{T} W_\delta' W_g (W_g' W_g)^{-1} (W_g' W_\delta))$ then E_1 will exceed B_1 implying that the instrumental variable estimator has better efficiency than the Amemiya-Heckman estimator.

The difference $V(\hat{\beta}^7) - V(\hat{\beta}^1)$ is further influenced by the positive matrix $E_2 = \text{PLIM}(\, W_\delta' M_g D_4 M_g W_\delta)$. The diagonal elements of D_4 are squares of the conditional expectation $E[\eta_t | \delta_t]$. Thus the power of the Amemiya-Heckman estimator as considered relative to the instrumental variable procedure is greatest when the hypothesized correlation in the error structure is largest.

GLOSSARY

CHAPTER 2:

VARIABLE	DESCRIPTION
AKWH75	monthly consumption of electricity in 1975
RATE	measured marginal price in 1975
AVPRICE	measured average price in 1975
WMPE75	winter tail-end block price for electricity in 1975
INCOME	monthly income of household head
RSP	measured rate structure premium
WHE	electric water heat dummy
SHE	electric space heat dummy
ROOMS	number of rooms in household
PERSONS	number of persons in household
CAC	central air-conditioning dummy
CDDCAC	(annual cooling degree days) * (CAC)
RACNUM	number of room air conditioners
CDDRACNUM	(annual cooling degree days) * (RACNUM)
AUTOWSH	automatic washing machine dummy
AUTODSH	automatic dishwasher dummy
FOODFRZ	food freezer dummy
ELECRNGE	electric range dummy
ECLTHDR	electric clothes dryer dummy
BWTV	black and white television dummy
CLRTV	color television dummy

CHAPTER 3:

Room Air-Conditioning Choice Model:

RMOPCST	operating cost for room air-conditioning (1967$)
RMCPCST	capital cost for room air-conditioning (1967$)
RMOPCST1	RMOPCST/(base load usage)
RMCPCST1	RMCPCST/(base load usage)
CDD78	cooling degree days in 1978
RINCOME	income (1967$)/$10^3$
NHSLDMEM	number of household members

Water Heat Choice Model:

WHOPCST	water heat operating costs
WHOPCST1	water heat operating cost divided by usage
WHCPCST	water heat capital cost
WHCPCST1	water heat capital cost divided by usage
SHE	(space heat fuel electricity)*(ALT1)
SHG	(space heat fuel gas)*(ALT2)
SHO	(space heat fuel oil)*(ALT3)

ALTERNATIVE

1	electric water heat
2	natural gas water heat
3	oil water heat

Water Heat Inclusive Values:

WHINCVE	water heat inclusive value given electricity
WHINCVG	water heat inclusive value given natural gas
WHINCVO	water heat inclusive value given oil

Space Heat Choice Mode:

SHOPCST	space heat operating costs
SHCPCST	space heat capital costs
SHOPCST1	SHOPCST/usage
SHCPCST1	SHCPCST/usage
SHOPCST2	SHOPCST/operating cost of HVAC 18
SHCPCST2	SHOPCST/operating cost of HVAC 18

Alternative

1	elec. forced air/no central air	HVAC #13
2	gas forced air/no central air	HVAC #1
3	oil forced air/no central air	HVAC #7
4	elec. baseboard/no central air	HVAC#18
5	gas hot water/no central air	HVAC #3
6	oil hot water/no central air	HVAC #9
7	elec. forced air/central air	HVAC #14
8	gas forced air/central air	HVAC #2
9	oil forced air/central air	HVAC #8
10	electric heat pump	HVAC #15

Space Heat Inclusive Value:

SHINCVNC	space heat inclusive value given no central air-conditioning
SHINCVC	space heat inclusive value given central air-conditioning

Central Air Choice Model:

CACOPC	central air-conditioning operating cost
CACCST	central air-conditioning capital cost

SU18	(HVAC 18 dummy)(UEC18)
SU13	(HVAC 13 dummy)(UEC13)
SU14S	(HVAC 14 dummy)(UEC14S)
SU15S	(HVAC 15 dummy)(UEC15S)
SU14A	(HVAC 14 dummy)(UEC14A)
SU15A	(HVAC 15 dummy)(UEC15A)
SUWHE	(Water heat electric dummy)(UECWH)
SURMAC	(Room air conditioner dummy)(UECRMAC)

SU18P, SU13P, SU14SP, SU15SP, SU14AP, SU15AP, SUWHEP, and SURMACP are variables multiplied by service prices.

SU18Y, SU13Y, SU14SY, SU15SY, SU14AY, SU15AY, SUWHEY, and SURMACY are variables multiplied by income.

MPE	Marginal price of electricity ($/KWH)
EDAYS	Number of days in aggregated period
NHSLDMEM	Number of household members
NETEQUAN	Net electricity usage (KWH)

SUSHE = SU18 + SU13 + SU14S + SU15S
SUSHEP = SU18P + SU13P + SU14SP + SU15SP
SUSHEY = SU18Y + SU13Y + SU14SY + SU15SY

SUCAC = SU14A + SU15A + SU2A + SU8A
SUCACP = SU14AP + SU15AP + SU2AP + SU8AP
SUCACY = SU14AY + SU15AY + SU2AY + SU8AY

SU1	(HVAC 1 dummy)(UEC1)
SU2	(HVAC 2 dummy)(UEC2)
SU3	(HVAC 3 dummy)(UEC3)
SUWHG	(Water heat gas dummy)(UECWH)

SU1P, SU2P, SU3P, and SUWHGP are variables multiplied by service prices.

SU1Y, SU2Y, SU3Y, and SUWHGY are variables multiplied by income.

MPG	Marginal price of natural gas ($/Therms)
GDAYS	Number of days in aggregated period
NETGQUAN	Net natural gas usage (Therms)

SUSHG = SU1 + SU2 + SU3
SUSHGP = SU1P + SU2P + SU3P
SUSHGY = SU1Y + SU2Y + SU3Y

REFERENCES

1. Acton, J.P., B.M. Mitchell, and R.S. Mowill (1976), Residential Demand for Electricity in Los Angeles: An Econometric Study of Disaggregated Data, The Rand Corporation, R-1899-NSF.

2. Acton, J.P., B.M. Mitchell, and R.S. Mowill (1978), Residential Electricity Demand under Declining Block Tariffs: An Econometric Study Using Micro-Data, The Rand Corporation, P-6203.

3. Amemiya, T. (1973), "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica 41, pp. 997-1016.

4. Amemiya, T. (1974a), "Bivariate Probit Analysis: Minimum Chi-Square Methods," Journal of the American Statistical Association 69, pp. 940-944.

5. Amemiya, T. (1974b), Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal," Econometrica 42, pp. 999-1012.

6. Amemiya, T. (1974c), "A Note on the Fair and Jaffee Model," Econometrica 42, pp. 759-762.

7. Amemiya, T. (1978a), "The Estimation of a Simultaneous Equation Generalized Probit Model," Econometrica 46, pp. 1193-1205.

8. Amemiya, T. (1978b), "On a Two-Step Estimation of a Multivariate Logit Model," Journal of Econometrics, Vol. 8, no. 1, p. 13-21.

9. Amemiya, T. (1979), "The Estimation of a Simultaneous Equation Tobit Model," International Economic Review in press.

10. Amemiya, T. and G. Sen (1977), "The Consistency of the Maximum Likelihood Estimator in a Disequilibrium Model," Tech. Report 238, Institute for Math. Studies in Social Sciences, Stanford.

11. Anderson, K.P. (1973), Residential Energy Use: An Econometric Analysis, The Rand Corporation, R-1297-NSF.

12. Annual Climatological Data - National Summary 1978-1979.

13. Association for Space Heat, Refrigeration, Air-Conditioning Manufacturers (1977), Handbook and Product Directory, Fundamentals, New York: ASHRAE.

14. Balestra, P. and Nerlove, M. (1966), "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica 34, pp. 585-612.

15. Barnes, R., R. Gillingham, and R. Hagemann (1981), "The Short-Run Residential Demand for Electricity," Review of Economics and Statistics.

16. Becker, G.S. (1965), "A Theory of the Allocation of Time," Economic Journal, 75, no. 299, pp. 493-517.

17. Berndt, E.R. (1978), "The Demand for Electricity: Comment and Further Results," Resources Paper No. 28, Department of Economics, University of British Columbia.

18. Blackorby, C., G. Lady, D. Nissen, and R.R. Russell (1970), "Homothetic Separability and Consumer Budgeting," Econometrica 38, pp. 468-472.

19. Blumstein, C., C. York, and W. Kemp, "An Assessment of the National Interim Energy Consumption Survey," undated draft, Energy Resources Group, University of California - Berkeley.

20. Brownstone, D. (1980), An Econometric Model of Consumer Durable Choice and Utilization Rate, unpublished Ph.D. thesis, University of California, Berkeley.

21. Burtless and Hausman (1978), "The Effect of Taxation on Labor Supply," J. Polit. Econ., 86, pp. 1103-30.

22. California State Energy Conservation (1979), "California Energy Demand 1978-2000," working paper.

23. Cambridge Systematics/West (1979), "An Analysis of Household Survey Data in Household Time-of-Day and Annual Electricity Consumption," Cambridge Systematics/West, working paper.

24. Cambridge Systematics, Inc. (1981), "Development of a Residential Appliance Survey for New England Electric System."

25. Chern, W.S., and W. Lin (1976), "Energy Demand for Space Heating: An Econometric Analysis," American Statistical Association, Proceedings of the Business and Economic Statistics Section.

26. Chow, G.C. (1957), Demand for Automobiles in the United States, ("Contribution to Economic Analysis," Vol. XIII Amsterdam: North-Holland Publishing Co.).

27. Cowing, T., J. Dubin, and D. McFadden (1981a), "Residential Energy Demand Modeling and the NIECS Data Base: An Evaluation," Massachusetts Institute of Technology Energy Laboratory Report No. MIT-EL-82-009.

28. Cowing, T., J. Dubin, and D. McFadden (1981b), "A Review and Evaluation of "An Assessment of the National Interim Energy Consumption Survey," by Carl Blumstein, Carl York, William Kemp - Energy Resources Group, University of California, Berkeley," Massachusetts Institute of Technology, DMCF-81-3.

29. Cowing, T., J. Dubin, and D. McFadden (1981c), "The NIECS Weather Information," Massachusetts Institute of Technology, DMCF-81-4.

30. Cowing, T., J. Dubin, and D. McFadden (1981d), "Report on Procedures Used to Identify State Locations for NIECS Households," Massachusetts Institute of Technology, DMCF-81-5.

31. Cowing, T., J. Dubin, and D. McFadden (1981e), "Estimating HVAC and Water Heating Equipment and Installation Costs Using NIECS Data," Massachusetts Institute of Technology, DMCF-81-8.

32. Cragg, J.G. (1971), "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," Econometrica 39, pp. 829-844.

33. Cragg, J.G. and R. Uhler (1970), "The Demand for Automobiles," Canadian Journal of Economics 3, pp. 386-406.

34. Diewert, W.D. (1974), "Intertemporal Consumer Theory and the Demand for Durables," Econometrica 42, pp. 497-516

35. Dubin, J. and D. McFadden (1979), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," Massachusetts Institute of Technology, Department of Economics Working Paper, presented at the 4th World Conference of the Econometric Society, Aix-en-Provence, France.

36. Duncan, G. (1980a), "Formulation and Statistical Analysis of the Mixed Continuous/Discrete Variable Model in Classical Production Theory," Econometrica 48, pp. 839-851.

37. Duncan, G. (1980b), "Mixed Continuous Discrete Choice Models in the Presence of Hedonic or Exogenous Price Functions," Washington State University, working paper.

38. Duncan, G. and Duane Leigh (1980), "Wage Determination in the Union and Non-Union Sectors: A Sample Selectivity Approach," Industrial and Labor Relations Review (forthcoming).

39. Fair, R.C. and D.M. Jaffee (1972), "Methods of Estimation for Markets in Disequilibrium," Econometrica 40, pp. 497-514.

40. Fisher, F. and C. Kaysen (1962), A Study in Econometrics: The Demand for Electricity in the U.S., North-Holland Publishing Company.

41. Fuss, M. and D. McFadden (1978), "Flexibility versus Efficiency in Ex-Ante Plant Design, in Fuss and McFadden (eds.), Production Economics: A Dual Approach to Theory and Applications, North-Holland Publishing Company, Amsterdam.

42. Geary, P.T. and M. Morishima (1973), "Demand and Supply under Separability," in Theory of Demand: Real and Monetary, M. Morishima and others (eds.), Oxford: Clarendon.

43. George, S. (1979), Short-Run Residential Electricity Demand: A Policy Oriented Look, Ph.D. Dissertation, University of California, Davis.

44. George, S. (1981), "A Review of the Conditional Demand Approach to Electricity Demand Estimation," (revised version of a paper given at the EPRI Conference on "End-Use Modeling and Conservation Analysis," Atlanta, Georgia).

45. Goett, A. (1979), "A Structured Logit Model of Appliance Investment and Fuel Choice," Working Paper, Cambridge Systematics, Inc. West.

46. Goett, A., D. McFadden, and G.L. Earl (1980), "The Residential End-Use Energy Policy System: Model Description and Analysis of Energy Usage and Efficiency Choice."

47. Goldfeld, S.M. and R.E. Quandt (1972), Nonlinear Methods on Econometrics, Amsterdam: North-Holland Publishing Company.

48. Goldfeld, S.M. and R.E. Quandt (1973), "The Estimation of Structural Shifts by Switching Regressions," Annals of Economic and Social Measurement 2, pp. 475-485.

49. Goldfeld, S.M. and R.E. Quandt (1976), "Techniques for Estimating Switching Regressions" in S.M. Goldfeld and R.E. Quandt (eds.) Studies in Non-Linear Estimation, Cambridge: Ballinger.

50. Gorman, W.M. (1959), "Separable Utility and Aggregation," Econometrica 27, pp. 469-481.

51. Griliches, Z. (1960), "The Demand for a Durable Input: Farm Tractors in the United States, 1921-57," in Demand for Durable Goods, Harberger (eds.).

52. Hartman, Raymond S. (1978), A Critical Review of Single Fuel and Interfuel Substitution Residential Energy Demand Models, Massachusetts Institute of Technology Energy Laboratory Report, MIT-EL-78-003.

53. Hartman, R.S. (1979a), "Discrete Consumer Choice Among Alternative Fuels and Technologies for Residential Energy Using Appliances," Massachusetts Institute of Technology Energy Laboratory Working Paper, MIT-EL-79-049WP.

54. Hartman, Raymond S. (1979b), "Frontiers in Energy Demand Modeling," Annual Review of Energy, Forthcoming.

55. Hartman, R.S., and A. Werth (1979), "Short-Run Residential Demand for Fuels: A Disaggregated Approach," Working Paper, MIT-EL-79-01WP. Also in Land Economics, 1981.

56. Hausman, J. (1978), "Specification Tests in Econometrics," Econometrica 46, pp. 1251-1271.

57. Hausman, J. (1979), "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables," The Bell Journal of Economics, Volume 10, No. 1, pp. 33-54.

58. Hausman, J. (1981), "Exact Consumer's Surplus and Deadweight Loss," American Economic Review 71, pp. 663-676.

59. Hausman, J., M. Kinnucan, and D. McFadden (1979), "A Two-Level Electricity Demand Model," Journal of Econometrics 10, pp. 263-289.

60. Hausman, J. and D. McFadden (1981), "Specification Tests for the Multinomial Logit Model," Department of Economics, Massachusetts Institute of Technology Working Paper No. 292.

61. Hausman, J. and Taylor, W.E. (1981), "Panel Data and Unobservable Effects, Econometrica 49, No. 5.

62. Hausman, J. and Trimble, J. (1981), "Appliance Purchase and Usage Adaptation to a Permanent Time of Day Electricity Rate Schedule, mimeo Massachusetts Institute of Technology.

63. Hay, J. (1979), "An Analysis of Occupational Choice and Income," Ph.D. Dissertation, Yale University.

64. Heckman, J. (1976a), "Simultaneous Equations Model with Continuous and Discrete Endogenous Variables and Structural Shifts" in Goldfeld and Quandt (eds.), Studies in Non-Linear Estimation, Cambridge: Ballinger.

65. Heckman, J. (1976b), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimation for Such Models," Annals of Economic and Social Measurement 5, pp. 475-492.

66. Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," Econometrica 46.

67. Heckman, J. (1979), "Sample Selection Bias as a Specification Error," Econometrica 47.

68. Hirst, E. and J. Carney (1978), "The Oak Ridge National Laboratory Engineering-Economic Model of Residential Energy Use," ORNL/CON-24.

69. Hirst, E., R. Goeltz, and J. Carney, "Residential Energy Use and Conservation Actions: Analysis of Disaggregate Household Data," ORNL/CON-68 Oak Ridge National Laboratory.

70. Houthakker, H.S. (1951a), "Electricity Tariffs in Theory and Practice," Economic Journal 61, pp. 1-25.

71. Houthakker, H.S. (1951b), "Some Calculations of Electricity Consumption in Great Britain," JRSS Series A 114, 351-371.

72. Houthakker, H.S. and L. Taylor (1970), Consumer Demand in the U.S., 2nd edition, Harvard University Press.

73. King, Mervyn (1980), "An Econometric Model of Tenure Choice and Demand for Housing as a Joint Decision," University of Birmingham Working Paper.

74. Lancaster, K. (1966), "A New Approach to Consumer Theory," J. Polit. Econ., 74, pp. 132-57.

75. Lee, L.F. (1978), "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables," International Economic Review 19, pp. 415-433.

76. Lee, L.F. (1981), "Simultaneous Equations Models with Discrete and Censored Variables," in C. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data, Cambridge, The Massachusetts Institute of Technology Press.

77. Lee, L.F. and R.P. Trost (1978), "Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand," Journal of Econometrics 8, pp. 357-382.

78. Lee, L.F., G.S. Madalla, and R. Trost (1980), "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," Econometrica 48, pp. 491-503.

79. Li, M. (1977), "A Logit Model of Home Ownership," Econometrica 45, pp. 1081-1097.

80. Maddala, G. and F. Nelson (1974), "Maximum Likelihood Methods for Markets in Disequilibrium," Econometrica 42, pp. 1013-1030.

81. Maddala, G. and F. Nelson (1975), "Switching Regression Models with Exogenous and Endogenous Switching," Proceedings of the American Statistical Association, Business and Economics Section, pp. 423-426.

82. McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, Frontiers in Econometrics, Academic Press, New York.

83. McFadden, D. (1974), "The Measurement of Urban Travel Demand," Journal of Public Economics 3, pp. 303-328.

84. McFadden, D. (1978), "Modelling the Choice of Residential Location," in A. Karlquist, et al., Spatial Interaction Theory and Residential Location, Amsterdam: North-Holland, pp. 75-96.

85. McFadden. (1979a), "Econometric Net Supply Systems for Firms with Continuous and Discrete Commodities," mimeo, Department of Economics, Massachusetts Institute of Technology.

86. McFadden, D. (1979b), "Quantitative Methods for Analyzing Travel Behavior of Individuals," in D. Hensher and P. Stopher (eds.), Behavioral Travel Modeling, Croom-Helm.

87. McFadden, D. (1981a), "An Evaluation of the ORNL Residential Energy Use Model," Draft Report, Massachusetts Institute of Technology Energy Laboratory, Energy Model Analysis Program.

88. McFadden, D. (1981b), "Econometric Models of Probabilistic Choice" in C. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data, Massachusetts Institute of Technology Press.

89. McFadden, D., D. Kirshner, and C. Puig (1978), "Determinants of the Long-Run Demand for Electricity," Proceedings of the American Statistical Association.

90. McFadden, D. and J. Dubin (1982), "A Thermal Model for Single-Family Owner-Occupied Detached Dwellings in the National Interim Energy Consumption Survey," Working Paper, Department of Economics, Massachusetts Institute of Technology.

91. McFadden, D. and C. Winston (1981), "Joint Estimation of Discrete and Continuous Choices in Freight Transportation," Working Paper, Department of Economics, Massachusetts Institute of Technology.

92. Mount, T.D., L.D. Chapman, and T.J. Tyrell (1973), "Electricity Demand in the United States: An Econometric Analysis," Oak Ridge National Laboratory.

93. Muellbauer, John (1974), "Household Production Theory, Quality, and the 'Hedonic Technique'," American Economic Review 64, pp 977-994.

94. Muth, Richard F. (1966), "Household Production and Consumer Demand Functions," Econometrica 34, pp. 699-708.

95. National Electric Rate Book, August 1978, DOE/EIA-0110 (78), Department of Energy-Energy Data Reports-New Jersey.

96. Nordin, J.A. (1976), "A Proposed Modification of Taylor's Demand Analysis: Comment," Bell Journal of Economics.

97. Parti, M. and C. Parti (1980), "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector," Bell Journal of Economics.

98. Pesaran, M.H. (1975), "On the General Problem of Model Selection," Review of Economic Studies 41, 153-171.

99. Pollak, R.A. and M.L. Wachter (1975), "The Relevance of the Household Production Function and Its Implications for the Allocation of Time," J. Polit. Econ., 83, pp. 255-277.

100. Response Analysis Corporation (1976), Lifestyles and Household Energy Use: Report on Methodology, Report #3819, Princeton, New Jersey.

101. Response Analysis Corporation (1981), Report on Methodology, 6 vols. (Princeton, New Jersey).

102. Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiations in Price Competition," J. Polit. Econ., 82, pp. 34-55.

103. Taylor, L.D. (1975), "The Demand for Electricity: A Survey," Bell Journal of Economics 6, Number 1, 74-110.

104. Taylor, L.D., G.R. Blattenberger, and P.K. Verleger (1977), "The Residential Demand for Energy," Electric Power Research Institute, Palo Alto.

105. U.S. Department of Energy, Energy Information Administration, Office of Energy Markets and End Use, DOE/EIA-0272, Exploring the Variability in Energy Consumption, July 1981.

106. U.S. Department of Energy, Energy Information Administration, Office of Energy Markets and End Use, DOE/EIA-0272/5, Exploring the Variability in Energy Consumption: A Supplement, October 1981.

107. U.S. Department of Energy, Energy Information Administration, Residential Energy Consumption Survey: Characteristics of the Housing Stocks and Households, 1978, Feb. 1980, DOE/EIA-0207/2.

108. U.S. Department of Energy, Energy Information Administration, Resdiential Energy Consumption Survey: Conservation, February 1980, DOE/EIA-0207/3.

109. U.S. Department of Energy, Energy Information Administration, Residential Energy Consumption Survey: Consumption and Expenditures, April 1979 Through March 1979, July 1980, DOE/EIA-1017/5.

110. U.S. Department of Energy, Energy Information Administration, Residential Energy Consumption Survey: 1979-1980 Consumption and Expenditures, Part I: National Data (Including Conservation), April 1981, DOE/EIA-0262/1.

111. U.S. Department of Energy, Energy Information Administration, Residential Energy Consumption Survey: 1979-1980 Consumption and Expenditures, Part II: Regional Data, May 1981, DOE/EIA-0262/2.

112. U.S. Department of Energy, Energy Information Administration, Single-Family Households: Fuel Inventories and Expenditures: National Interim Energy Consumption Survey, Dec. 1979, DOE/EIA-0207/1.

113. U.S. Department of Energy, "STATE Energy Fuel Prices by Major Economic Sectors from 1960-1977," DOE-EIA-0190, July, 1979.

114. U.S. Department of Energy, Office of Energy Markets and End Use Energy Information Administration, "Technical Documentation for the Residential Energy Consumption Survey: National Interim Energy Consumption Survey 1978-1979, Household Monthly Energy Consumption and Expenditures - Public Use Data Tapes - User's Guide," August, 1981.

115. Werth, Alix (1978), "Residential Demand for Electricity and Natural Gas in the Short Run: An Econometric Analysis, Massachusetts Institute of Technology Energy Laboratory Working Paper No. MIT-EL-78-031WP.

116. Wilder, R.P., and J.F. Willenborg (1975), "Residential Demand for Electricity: A Consumer Panel Approach," Southern Economic Journal.

117. Wilson, J. (1971), "Residential Demand for Electricity," Quarterly Review of Economics and Business.

118. Wu, D. (1973), "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," Econometrica 41, 733-750.