Estimates of Future Water Demand for Selected Water-Service Areas in the Upper Duck River Basin, Central Tennessee

By SUSAN S. HUTSON

With a section on Methodology Used to Develop Population Forecasts for Bedford, Marshall, and Maury Counties, Tennessee, From 1993 Through 2050

By GREGORY E. SCHWARZ

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 96-4140

Prepared in cooperation with the
DUCK RIVER DEVELOPMENT AGENCY

Memphis, Tennessee
1996

U.S. DEPARTMENT OF THE INTERIOR
BRUCE BABBITT, Secretary

U.S. GEOLOGICAL SURVEY
Gordon P. Eaton, Director

For additional information write to:

District Chief
U.S. Geological Survey
810 Broadway, Suite 500
Nashville, Tennessee 37203

Copies of this report may be purchased from:

U.S. Geological Survey
Branch of Information Services
Box 25286
Denver, Colorado 80225-0286

CONTENTS

FIGURES

TABLES

CONVERSION FACTORS

Multiply	By	To obtain
inch (in.)	25.4	millimeter
foot (ft)	0.3048	meter
mile (mi)	1.609	kilometer
square miles (mi^2)	640.0	acres
acre feet (acre ft)	0.001233	cubic meters
cubic foot (ft^3)	0.02832	cubic meter
cubic foot per second (ft^3/s)	0.02832	cubic meter per second
cubic foot per second (ft^3/s)	0.64632	million gallons per day
gallon (gal)	0.0037854	cubic meter
gallon per minute (gal/min)	0.06309	liter per second
gallon per day (gal/d)	0.003785	cubic meter per day
million gallons (Mgal)	0.3785	cubic meter
million gallons per day (Mgal/d)	0.04381	cubic meter per second

GLOSSARY

Significant terms, defined according to their meaning in this report, are listed below:

Bill difference is the difference in the customer's actual total bill and what would be charged if all units of water were sold at the marginal price (U.S. Army Corps of Engineers, 1988).

Coefficient is a statistically derived measure of a property or characteristic of water use utilized as a factor in the computation of water demand.

Constant is the y-intercept in the demand models related to gallons of water per housing unit per day (U.S. Army Corps of Engineers, 1988).

Constant dollars. Current dollar figures reflect actual prices or costs prevailing during the specified year(s). Constant dollar figures are estimates representing an effort to remove the effects of price changes from statistical series reported in dollar terms. Constant dollar series are derived by dividing current dollar estimates by the appropriate price index for the appropriate period of time. The result is presumably a series that would exist if prices were the same throughout time (Vickers, 1989).

Gross per capita water use. The quantity of water used within the public-supply distribution system. The value is calculated as the sum of water withdrawn by a public-supply system plus water purchased from other systems minus water sold to other public-supply systems. The water is sold to residential, commercial, and industrial customers or provided free as public-use water, and includes water lost in the distribution system.

Housing density is the number of housing units per acre (U.S. Army Corps of Engineers, 1988).

Marginal price. Annual marginal price is estimated by averaging, over all billing periods, the price in effect for the last units of water used plus any wastewater charges or surcharges. (U.S. Army Corps of Engineers, 1988).

Median income is a type of average which divides the distribution into two equal parts; one-half of the households fall below the median income and one-half of the households exceed the median income (U.S. Department Commerce, 1992b).

Multiple-coefficient demand models include the price of water to the user, as well as related economic factors such as income, among the explanatory variables. Demand models are usually constructed according to econometric methods, where the structure of the model and the list of potential explanatory variables reflect assumptions regarding causality rather than simply arising from observed correlation (U.S. Army Corps of Engineers, 1988).

Municipal water is public-supply water delivered to residential, commercial, and industrial users. The amount of water also includes public/unaccounted water.

Per capita income is the average annual rate of income per person.

Per capita water use is the average daily rate of use of water per person.

Price elasticity is a dimensionless measure of the relation between a percentage of change in water use and a percentage of change in price when other factors affecting water demand remain unchanged. The same concept may be applied to express the responsiveness of water use to changes in other variables (Boland and others, 1984).

Public/unaccounted sector is free-service water and distribution losses which include leakage, pipe flushing, and apparent losses caused by cumulative meter misregistration.

Single-coefficient (unit) or multiple-coefficient requirement models may be expressed as a function of one or more explanatory variables. The models do not include the price of water or other economic factors as explanatory variables. The models imply that water use is an absolute requirement, unaffected by economic choice (U.S. Army Corps of Engineers, 1988).

Standard Industrial Classification (SIC) is the statistical classification standard underlying all establishment-based Federal economic statistics classified by industry. The SIC is used to promote the comparability of establishment data describing various facets of the U.S. economy. The classification covers the entire field of economic activities and defines industries in accordance with the structure and composition of the economy (Office of Management and Budget, 1987).

Water demand is the relation between water use and price when all other factors are held constant. Demand is a negative functional relation; increased price results in decreased water use (Boland and others, 1984).

Water use. In the restrictive sense, the term refers to water that is actually used for a specific purpose, such as for domestic use or industrial processing. More broadly, water use pertains to human's interaction with and influence on the hydrologic cycle, and includes elements such as water withdrawal, delivery, consumptive use, wastewater release, reclaimed wastewater, return flow, and instream use.

Estimates of Future Water Demand for Selected Water-Service Areas in the Upper Duck River Basin, Central Tennessee

By Susan S. Hutson

Abstract

Estimates of future water demand were determined for selected water-service areas in the upper Duck River basin in central Tennessee through the year 2050. The Duck River is the principal source of publicly-supplied water in the study area providing a total of 15.6 million gallons per day (Mgal/d) in 1993 to the cities of Columbia, Lewisburg, Shelbyville, part of southern Williamson County, and several smaller communities. Municipal water use increased 19 percent from 1980 to 1993 (from 14.5 to 17.2 Mgal/d). Based on certain assumptions about socioeconomic conditions and future development in the basin, water demand should continue to increase through 2050.

Projections of municipal water demand for the study area from 1993 to 2015 were made using econometric and single-coefficient (unit-use) requirement models of the per capita type. The models are part of the Institute for Water Resources-Municipal and Industrial Needs System, IWR-MAIN. Socioeconomic data for 1993 were utilized to calibrate the models.

Projections of water demand in the study area from 2015 to 2050 were made using a single-coefficient requirement model. A gross per capita use value (unit-requirement) was estimated for each water-service area based on the results generated by IWR-MAIN for year 2015. The gross per capita estimate for 2015 was applied to population projections for year 2050 to calculate water demand. Population was projected using the log-linear form of the Box-Cox regression model.

Water demand was simulated for two scenarios. The scenarios were suggested by various planning agencies associated with the study area. The first scenario reflects a steady growth pattern based on present demographic and socioeconomic conditions in the Bedford, Marshall, and Maury/southern Williamson water-service areas. The second scenario considers steady growth in the Bedford and Marshall water-service areas and additional industrial and residential development in the Maury/southern Williamson water-service area beginning in 2000.

For the study area, water demand for scenario one shows an increase of 121 percent (from 17.2 to 38 Mgal/d) from 1993 to 2050. In scenario two, simulated water demand increases 150 percent (17.2 to 43 Mgal/d) from 1993 to 2050.

INTRODUCTION

Water use for municipal purposes is increasing in the upper Duck River basin (fig. 1). Water for domestic, industrial, and commercial uses (municipal purposes) from public-supply systems increased 19 percent, from 14.5 million gallons per day (Mgal/d) in 1980 to 17.2 Mgal/d in 1993. Projected residential, industrial, and commercial developments in the basin suggest that water use is likely to continue to increase. Uncertainty exists among officials from local agencies in the basin about the adequacy of existing water supplies to meet future demands. Long-term forecasts can help decision-makers determine if the Duck River, the principal source of publicly-supplied water in the basin, can supply the future demands.

In 1989, the U.S. Geological Survey (USGS) conducted an investigation to document trends in water availability, use, and future demand in the upper

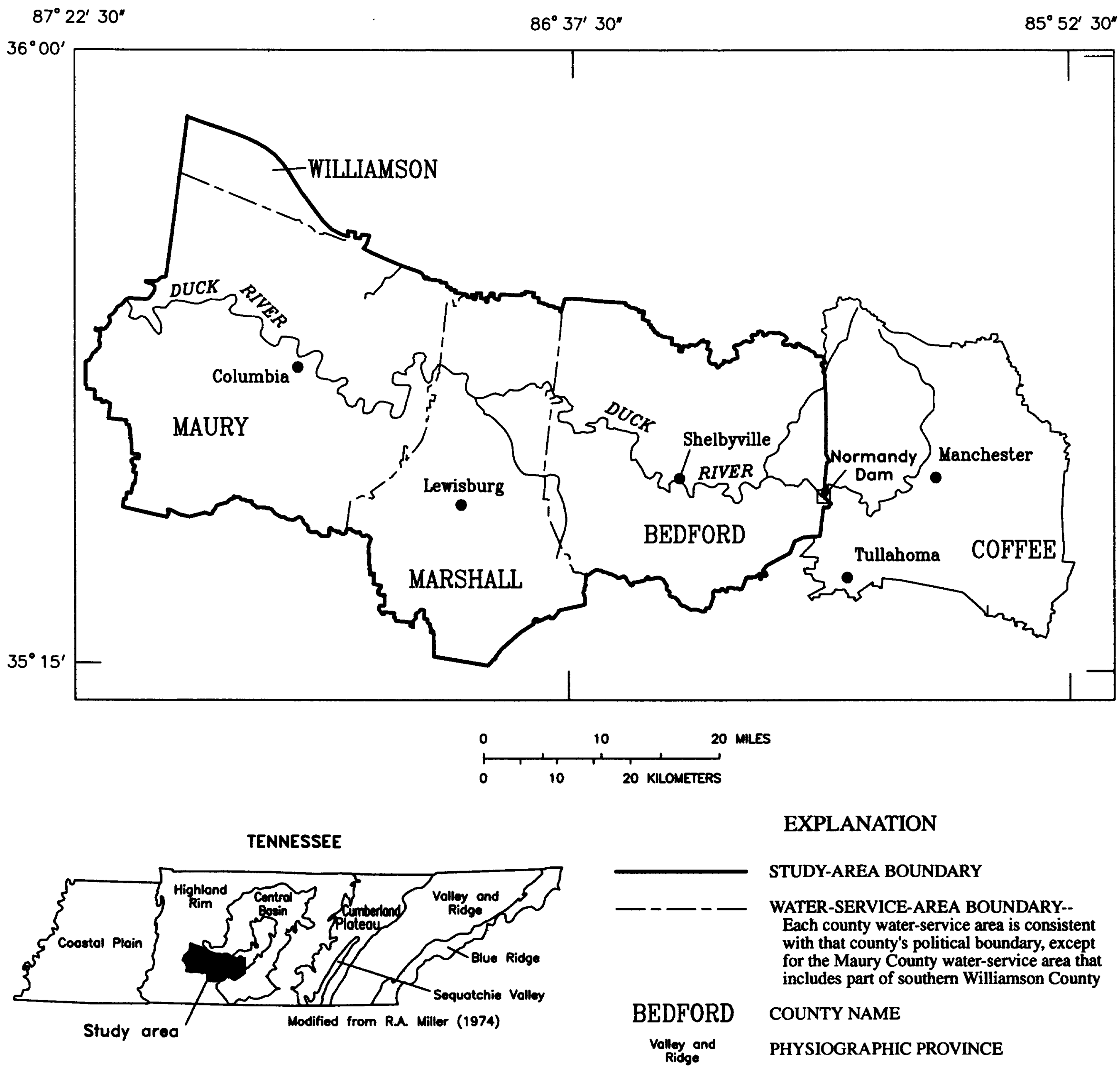


Figure 1. Upper Duck River study area and the Maury/southern Williamson, Marshall, and Bedford water-service areas.

Duck River basin (Hutson, 1993)[1]. The study provided estimates of future water demand through 2015. The study concluded that increases in withdrawals from the Duck River downstream of the city of Shelbyville would reduce minimum flows at the public water-supply intakes for the cities of Lewisburg and Columbia (fig. 1).

The effect of reduced flows on water quality was not addressed in the Hutson (1993) study. However, discharge permits that affect water quality are based on current minimum flows. Reduced minimum flows along the Duck River would reduce the carrying capacity of the river which would necessitate further costly treatment by municipalities and industry (such as tertiary treatment of effluent). Otherwise, deterioration of water quality in the river would ensue. Adverse water-quality conditions would be further exacerbated by any additional loadings of treated effluent that would emanate from economic development associated with increased water use (Larry M. Richardson, Manager, Water Resources Operations, Tennessee Valley Authority, written commun., 1992).

In 1994, the USGS began an investigation to reassess municipal water demand in the basin using more recent (1990 and 1993) socioeconomic and demographic data and to extend the water demand projections to 2050. The investigation was conducted in cooperation with the Duck River Development Agency (DRDA). Part of the mission of the USGS is to assess the water use of the nation. In addition to collecting water-use data, this study evaluated and developed methodologies and techniques for estimating water use and for projecting future water demand that may be applied nationwide.

Purpose and Scope

The purpose of this report is to provide estimates of future municipal water demand for the year 2050. Municipal water demand is publicly-supplied water delivered to residential, industrial, and commercial users and includes conveyance losses in the distribution systems. The study area is comprised of Bedford, Marshall, and Maury Counties, and part of southern Williamson County (fig. 1). The investigation was limited to this area because municipal water demand from the upper Duck River will most probably increase as a result of growth primarily in these counties (Steve Parks, Director, Duck River Development Agency, oral commun., 1994). The investigation includes an inventory of the municipal water use in the study area and excludes any assessment of the effect of withdrawals on streamflow or of the availability of streamflow for waste assimilation.

Estimates of municipal water demand for 2015 were made with the Institute for Water Resources-Municipal and Industrial Needs System (IWR-MAIN or System) water-use models. These models were designed by Planning and Management Consultants, Ltd., under the auspices of the U.S. Army Corps of Engineers, Institute for Water Resources (U.S. Army Corps of Engineers, 1988).

The structure of the IWR-MAIN System limits projections to 25 years, for example, 1990 to 2015. A single-coefficient requirement model, therefore, was devised to extend the forecast to 2050. Estimates of municipal water-use demand for 2050 were calculated as the product of the gross per capita water use derived from the IWR-MAIN forecasts for 2015 and a population estimate for 2050 for each water-service area.

Three types of models for projecting population over a long period of time (58 years, 1993 to 2050) were evaluated. The projections of population produced by the log-linear case of the Box-Cox model were used in the single-requirement models because the log-linear case yields the more reasonable and similar estimates.

The terms **water use**, **water demand**, and **water requirement** are commonly interchanged. Technically the terms are distinct. Water use refers to the water that is actually used for a specific purpose (See Glossary). Water demand is the relation between water use and price when all other factors are held constant. Water requirement is water use as an absolute requirement unaffected by economic choice.

Water-use values are expressed in the text as three-significant figures and in the tables as three-significant figures or as two decimal places. Percentages appear in the text and tables as integers. Simulated water demand for the calibration year, 1993, is expressed as three-significant figures or as two decimal places. Simulated water demand for years 2000 through 2050 is expressed as two-significant figures or as one decimal place to reflect the uncertainty of the accuracies of the projections over the length of time. Values may not add to totals shown because of independent rounding.

---

[1]References begin on page 43.

Approach

The following tasks were designed to accomplish the project objectives:

- Collect municipal water-use data for each public-supply system in the water-service areas for 1980, 1985, 1989, 1990, and 1993.
- Calibrate the IWR-MAIN water-demand models using demographic, economic, and water-use data for 1993. Estimate **municipal water** demand for 2015 using the calibrated models. Calculate **gross per capita use** for each water-service area.
- Project population to 2050 using the log-linear form of the Box-Cox statistical model.
- Input gross per capita use and population projections to a **single-coefficient requirement model** to forecast water demand for 2050.

HYDROLOGIC SETTING

The upper Duck River basin drains parts of the Highland Rim and Central Basin physiographic regions of Tennessee (Miller, 1974) (fig. 1). The climate of the area is moderate, with annual rainfall averaging 46 inches per year. The river flows from the dissected limestone highlands in northern Coffee County into Normandy Reservoir, which was completed by the Tennessee Valley Authority (TVA) in 1976 (fig. 1). The reservoir, with a capacity of 117,000 acre-feet at normal maximum headwater elevation of 875 feet, is used for flood control, water supply, water-quality enhancements, and recreation. The study area begins downstream from Normandy Dam in Bedford County, and includes Marshall, Maury, and southern Williamson Counties. The study area ends at the western Maury County line and, including southern Williamson County, is about 1,500 square miles.

Since the construction of Normandy Reservoir, the 3-day 20-year low-flow discharge in the upper Duck River as it flows through Bedford, Marshall, and Maury Counties has increased. As a result, the river has become a more reliable source of supply of water during low flows. The 3-day 20 year (3Q20) minimum flow represents the lowest mean daily flow for a consecutive 3-day period with a recurrence interval of 20 years. The period of record is based on the climatic year which extends from April 1 through March 31 and is designated by the year in which it ends (Bingham, 1985). Estimated values of 3Q20 at Shelbyville and Columbia before and after completion of the dam site are as follows (Outlaw and Weaver, 1996):

Location	Before 1976			After 1976		
	Cubic feet per second	Million gallons per day	Period of record	Cubic feet per second	Million gallons per day	Period of record
Shelbyville	53.8	34.8	1949-1976	73.5	47.5	1977-1993
Columbia	68.6	44.3	1961-1976	97.9	63.3	1977-1993

WATER USE

The study area in the upper Duck River basin is divided into three municipal water-service areas (WSA's), whose boundaries closely coincide with their respective county boundaries (fig. 1). The three WSA's are Bedford, Marshall, and Maury/southern Williamson. The Duck River is the main source of water for municipal use in the study area, supplying 91 percent of the municipal water used. Springs and wells supplement the municipal water supply from the Duck River. Ground water supplies about 9 percent of the municipal water used in the study area.

Withdrawals for municipal use in the basin increased 19 percent from 1980 to 1993 from 14.5 to 17.2 Mgal/d (Alexander and others, 1984; Tennessee Division of Water Supply files, 1990 and 1993; USGS files, 1985, 1988, and 1990) (table 1). Most of the increase has occurred in recent years. The

Table 1. Total surface- and ground-water withdrawals by water-service area

Water-service area	Withdrawals, in million gallons per day					
	1980	1985	1988	1989	1990	1993
Bedford	3.81	3.62	3.99	4.36	4.20	4.59
Marshall	2.27	2.51	2.61	2.82	2.85	2.52
Maury/ southern Williamson.	8.43	8.75	8.94	9.41	9.93	10.1
Study area totals	14.5	14.9	15.5	16.6	17.0	17.2

Maury/southern Williamson WSA, for example, increased water use 1.16 Mgal/d between 1988 and 1993, the largest increase in the study area. Total water use in the study area in the same period increased 1.70 Mgal/d. These increases in water use are mainly due to increased demands by the residential and industrial sectors (table 2). However, recent dry weather (drought of 1988) also affected water demand in the area by creating a need for additional water for outside usage in the summer (Hutson, 1993).

Municipal water use in the study area for 1993 was estimated from an inventory of public-supply systems. The inventory was conducted in 1994 in cooperation with the Duck River Development Agency. Data from 14 public-supply systems provided information on the source of supply; daily average-annual amount of water withdrawn or purchased (table 3); and, amounts distributed to residential, commercial, and industrial users; conveyance losses; and free service. Conveyance losses and free water is referred to as **public/unaccounted**. Self-supplied water for industrial and residential purposes was not inventoried. The inventory showed that:

- Municipal withdrawals in 1993 totaled 17.2 Mgal/d (table 1).
- Surface water accounted for 91 percent of the withdrawals (15.6 Mgal/d); the remaining 9 percent or 1.66 Mgal/d was withdrawn from springs and wells (table 3). Water was withdrawn downstream of Normandy Reservoir at four public-supply intakes on the Duck River (Shelbyville Water System, Bedford County Utility District, Lewisburg Water System, and Columbia Water Department).
- Residential water use accounted for 40 percent (6.86 Mgal/d); commercial use, 15 percent (2.60 Mgal/d); industrial use, 25 percent (4.34 Mgal/d); and, public/unaccounted water, 20 percent (3.41 Mgal/d) of the total withdrawals (table 2).

WATER-DEMAND SIMULATION

Water demand in the upper Duck River basin downstream of Normandy Dam was simulated to 2050 for two scenarios. Scenario 1 reflects a steady growth pattern based on present demographic and socioeconomic conditions in the Bedford, Marshall, and Maury/southern Williamson WSA's. Scenario 2 considers steady growth in the Bedford and Marshall WSA's and additional industrial and residential development in the Maury/southern Williamson WSA beginning in 2000.

The combined use of the IWR-MAIN System and the population estimates was necessary to forecast water demand to 2050. The econometric and single-coefficient requirement models of the IWR-MAIN System were used to forecast water demand to 2015 for each WSA. The IWR-MAIN System limits water-use projections to 25 years (1990-2015). To forecast water demand to 2050, the IWR-MAIN results for 2015 were combined with population estimates for 2050 in a single-coefficient requirement (SCR) model. The gross per capita water use for scenario 1 (steady growth) and for scenario 2 (steady growth with increased growth in Maury/southern Williamson WSA) was determined from estimates produced by IWR-MAIN for each WSA for 2015. The gross per capita water use for each scenario and the projected population for 2050 were multiplied to estimate water use for each scenario in 2050. The methodology used for projecting population over a long period of time was developed for this study by G.E. Schwarz. The methodology and the results for the study area are detailed in a separate section by Schwarz.

Institute for Water Resources-Municipal and Industrial Needs System

The IWR-MAIN System (herein referred to as IWR-MAIN or System) was used to estimate future municipal water demand. Econometric demand and single-coefficient requirement (usually of the unit-use type) models calculated water demand as a function of socioeconomic parameters. A value for each of these parameters was projected for the years for which water demand was estimated.

IWR-MAIN is used primarily to test assumptions and the effect various assumptions or changes would have on water use in the basin rather than as a predictive tool to generate absolute amounts of water use in the future. This fact and basic assumptions about growth, land use, population, and technology drive the results. If the assumptions are changed (for example, population decreases in the area), the model's water-demand results will change. The accuracy of the results depends on the validity of the assumptions.

Table 2. Public-supply deliveries of water to various water-use sectors by water-service area in 1993
[Values, in million gallons per day]

Water-service area	Sector: Residential	Commercial	Industrial	Public/ unaccounted
Bedford	1.75	0.72	1.08	1.04
Marshall	.96	.50	.45	.61
Maury/southern Williamson	4.15	1.38	2.81	1.76
Study area totals	6.86	2.60	4.34	3.41

Table 3. Public-supply systems and source(s) of supply in 1993
[Mgal/d, million gallons per day; WSA, water-service area; --, no transaction; gw, ground water; WS, Water System; UD, Utility District; MCBPU, Marshall County Board of Public Utilities; and, WD, Water Department]

Public-supply system	Source of supply (river mile)	Withdrawals (Mgal/d)	Purchased water (Mgal/d)
	Bedford WSA		
Shelbyville Water System	Duck River (227.0)	3.28	--
Bedford County UD #1 and #2	Duck River (202.4)	.77	--
	Shelbyville WS	--	0.00
Bell Buckle Water System	Wartrace WS	--	.16
Wartrace Water System	Cascade Spring (gw)	.54	--
Flat Creek Cooperative	Shelbyville WS	--	.09
	Marshall WSA		
Chapel Hill Water System	MCBPU #1	--	.00
	Town well (gw)	.11	--
Marshall County Board of Public Utilities.	Lewisburg WS	--	.37
	Cornersville WD	--	.02
	Chapel Hill WS	--	.03
Cornersville Water Department	Lewisburg WS	--	.11
Petersburg Water System[1]	Fayetteville WS	--	.05
Lewisburg Water System	Duck River (181.0) (metered at City Lake)	2.41	--
Henry Horton State Park[2]	MCBPU #1	--	.02
	Chapel Hill WS	--	--
	Maury/southern Williamson WSA		
Columbia Water Department	Duck River (133.7)	9.09	--
Mount Pleasant Water System	Spring (gw)	1.01	--
Spring Hill Water Department	Columbia WD	--	.23
Maury County Water System	Columbia WD	--	.69
Hillsboro and Thompson Station Utility District.	Spring Hill WD	--	.36

[1] Water purchased by this public water-supply system is not included in the study area total. Water is withdrawn from the Elk River watershed.

[2] Facility is a non-community public-water supplier. For the purposes of the model, the transfer of water from MCBPU #1 is handled as if it were a delivery to a commercial user.

Model Description

IWR-MAIN is a water-demand forecasting system that contains a range of water-use models, socioeconomic-parameter generating procedures, and data-management techniques (U.S. Army Corps of Engineers, 1988). Nonmunicipal (self-supplied) or rural water demand is not simulated by IWR-MAIN. The architecture of IWR-MAIN allows for the separation of the study area into smaller study units (spatial disaggregation) and the analysis of the smaller units by sector and by season (fig. 2).

The Duck River study area was separated in IWR-MAIN into study units that correspond with the water-service areas. The IWR-MAIN divides municipal water users within each WSA into four major sectors: residential, commercial, industrial, and public/unaccounted. Each sector is further divided into a number of categories for simulation purposes. The seasonal dimensions of the System consider any one of the elements of annual average water use, summer or winter season water use, or maximum daily water use for each sector and category.

The relation between the calibration and the simulation processes of the System is displayed graphically in figure 3. The schematic illustrates how the data modules relate to the computational modules and to the results. The base-year data are used to produce future-year data by means of internal models (computational module) that project growth for the various socioeconomic parameters. The growth in socioeconomic parameters (future data) also may be projected externally by the user and added to the model. Base (or calibration) year data are incorporated into the water-use models to simulate base (or calibration) year water use. If the base and calibration year are different years, selected future-year data also are incorporated into the model. The water-use models are calibrated by adjusting the library values to reflect local socioeconomic and climatic conditions. The libraries contain the model **constants**, parameter coefficients, and climatic values. Base-year and future-year data are used by the water-use models to calculate water demand for future years.

Version 5.1 of IWR-MAIN used in this study was prepared by Planning and Management Consultants, Ltd., in cooperation with the U.S. Army Corps of Engineers, Water Resources Support Center, Institute for Water Resources (U.S. Army Corps of Engineers, 1988). The user's manual and system description for IWR-MAIN provides additional details for much of the discussion presented in this section of the report.

Demand models

The econometric water-demand models relate socioeconomic parameters to water use for the residential, industrial, and commercial sectors (table 4). For the purposes of this study, these models were applied to only the residential sector. For the purposes of this study the residential sector was divided into two categories: metered and self-supplied. The number of housing units in each category is the variable driving the residential models. Housing value and price of water are the primary economic variables.

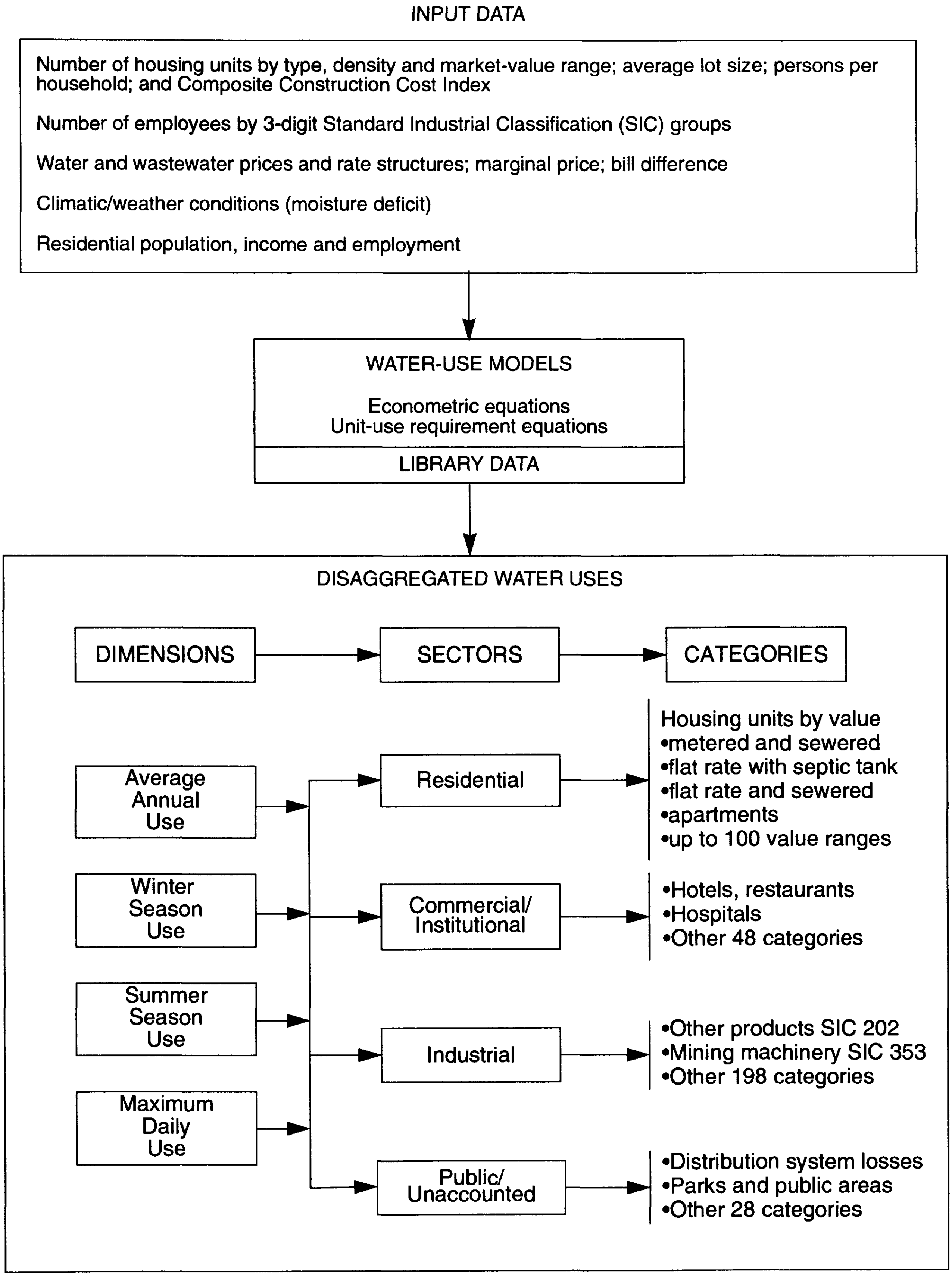
Water demand for the metered water category is calculated for each housing value range for the summer and winter seasons by applying **multiple-coefficient demand models**. Only the equation for the summer metered water category includes weather conditions as a factor for influencing water demand. Precipitation and evapotranspiration values read from the IWR-MAIN Library of Climatic Variables produce values for the moisture deficit parameter. Precipitation and evaporation values are retrieved from the library using the latitude and longitude for each WSA (table 5).

Summer water demand includes indoor and outdoor usage. Indoor usage is related to the drinking, cooking, bathing, cleaning, and similar activities inside a household. Outdoor usage is related to lawn or garden sprinkling, car washing, or other similar activity. Irrigable land and moisture-deficit are variables

Table 4. Socioeconomic parameters input to the Institute for Water Resources-Municipal and Industrial Needs System
[SIC, Standard Industrial Classification groups]

Required base-year parameters	Future-year parameters
Number of residences by type and value range.	Number of residences by type and value range.
Commercial and industrial employment by SIC.	Commercial and industrial employment by SIC.
Number of persons per household	Total employment[1]
Median household income	Median household income[1]
Resident population	Resident population[1]
Water and sewer rate structure	
Composite Construction Cost Index	
Bill difference	
Climatic conditions	
Total population	

[1] Required model input.



Modified from U.S. Army Corps of Engineers (1988)

Figure 2. Schematic describing the Institute for Water Resources-Municipal and Industrial Needs System architecture.

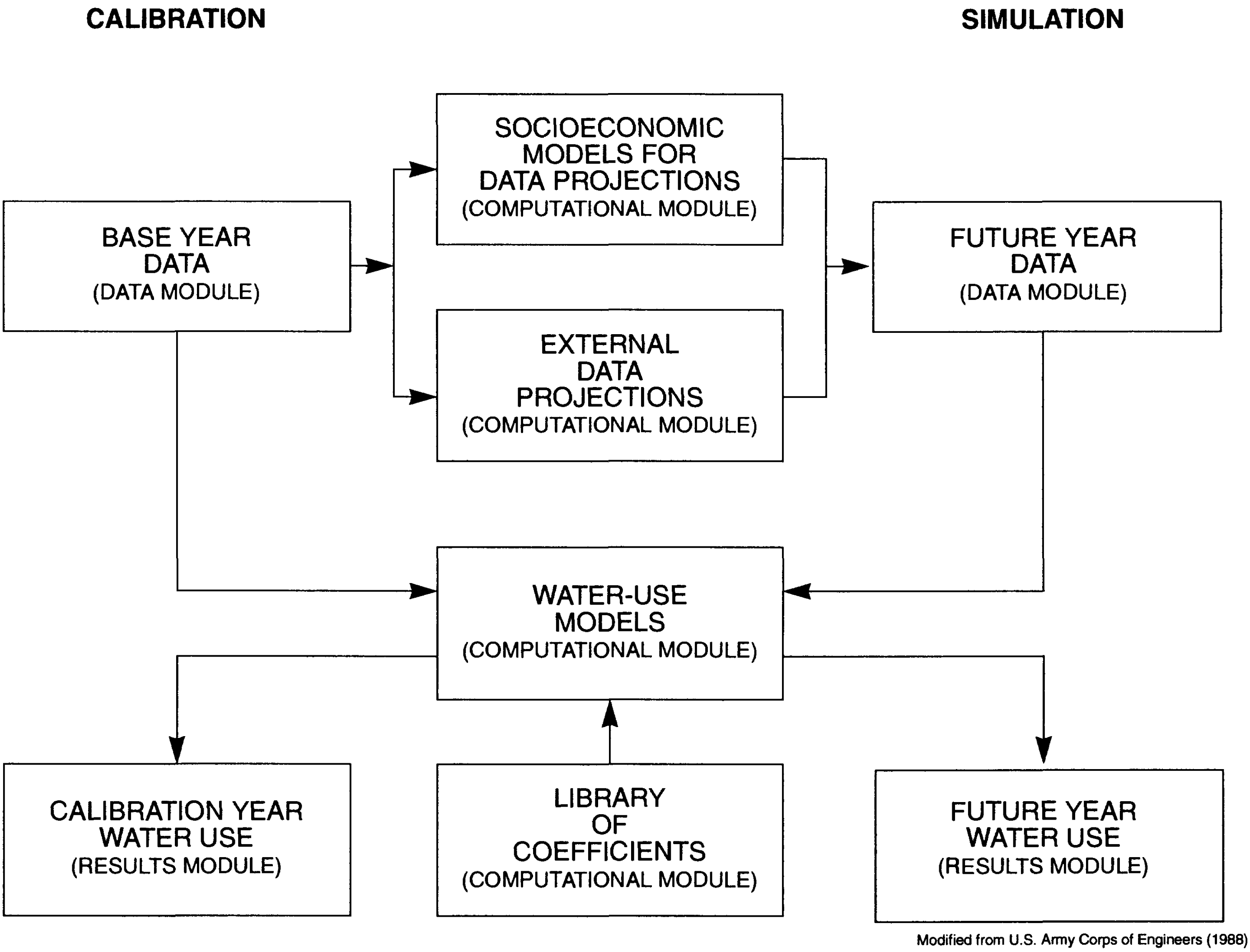


Figure 3. Schematic relating the calibration and simulation processes of the Institute for Water Resources-Municipal and Industrial Needs System to data, computational subroutine, library, and result modules.

Table 5. Climatological variables for the water-service areas

[Source: Institute for Water Resources-Municipal and Industrial Needs Library of Climatic Variables; rain and evapotranspiration measurements for the summer season (June, July, and August) based on long-term average weather data as of 1967]

Water-service area	Potential evapotranspiration: Total (inches)	Potential evapotranspiration: Maximum daily (inches)	Precipitation (inches)
Bedford	17.25	0.29	11.0
Marshall	18.5	.29	10.5
Maury/ southern Williamson.	18.5	.29	10.5

which factor into the summer water demand. The equation to determine the metered water use in the summer, indoor and outdoor, in the study area is:

$$(Q_s) = (385 + 2.876V/F_a - 285.8P_s - 4.35I_s + 157.77 \cdot B \cdot MD)N_r, \quad (1)$$

where

Q_s is summer season indoor and outdoor water usage, in gallons per day;

385 is a constant representing indoor and outdoor usage, in number of gallons per household per day;

V is average **house value, in a range of value** per 1,000 dollars;

F_a is assessment factor;

P_s is effective summer marginal price of water, in dollars per 1,000 gallons;

I_s is effective summer bill difference variable, in dollars per billing period;

B is irrigable land per dwelling unit, in acres per unit;

MD is summer-season moisture deficit, in inches; and

N_r is number of residences, in value range r.

Irrigable land is a function of **housing density** and is derived from the following equation:

$$B = 0.803 \cdot Hd^{-1.26}, \quad (2)$$

where

B is irrigable land per dwelling unit, in acres per unit, and

Hd is housing density, in number of units per acre.

Summer-season moisture deficit, MD, is calculated as follows:

$$MD = E - 0.6R, \quad (3)$$

where

E is summer-season potential evaporation, in inches, and

R is summer-season precipitation, in inches.

Winter water demand includes only indoor water usage. Winter usage is calculated as follows:

$$(Q_D) = (234 + 1.451\,V/F_a - 45.9P_a - 2.59I_a)\,N_r, \quad (4)$$

where

V, F_a, and N_r are as defined in equation (1);

Q_D is winter or indoor water use, in gallons per day;

234 is y-intercept, number of gallons per household per day;

P_a is effective annual marginal price of water, in dollars per 1,000 gallons; and

I_a is effective annual marginal price of water, in dollars per billing period.

Summer and winter season water demand are calculated. The model then aggregates the water use and produces a seasonally weighted residential rate of use for the WSA, including maximum daily and average annual. In the model, housing value acts as a proxy for income. **Marginal price** and **bill difference** variables capture the effects of change in the water-rate structure on disposable income. The model assumes that the higher the disposable income, the greater the water use.

Single-coefficient requirement models

The single-coefficient requirement (unit-use) models estimate future water demand as a product of projected WSA commercial or industrial employment and a projected value of per-employee water use. Price of water is not a factor in these models. The unit-use coefficient is assumed to be fixed through time, that is, new technology is not a factor the model recognizes. Industrial and commercial water use were estimated as follows:

$$(Q_a)_n = (C_a)_n \cdot P_n, \quad (5)$$

where

Q is water usage, in gallons per day;

$_a$ is average annual use;

$_n$ is industrial (commercial) use category;

C is industrial (commercial) water-use coefficient, in gallons per employee per day; and

P is number of employees.

Data Preparation/Model Input

Housing and employment data were prepared as input to the water-use and socioeconomic models contained in IWR-MAIN. Several assumptions (about the character of the data for the base and the calibration years, and about the structure of socioeconomic conditions in future years) were necessary to model the basin. These assumptions are detailed within the respective data sections.

The water-use models of the IWR-MAIN System utilized demographic and economic data provided externally by the user as well as parameter values generated internally by socioeconomic models in the System. Actual values of these parameters are required for a base, or beginning year, and projected values of selected parameters for specified future years. The data were developed for each WSA (Bedford, Marshall, and Maury/southern Williamson) for the residential, commercial, and industrial sectors. This spatial separation allowed the System to consider varying rates of sector growth within the study area, and to consider the effect of different price rate structures on water use. Data were prepared for the base year, 1990; calibration year, 1993; and future years, 2000 and 2015.

The socioeconomic models (herein referred to as housing or employment models) in IWR-MAIN generated future values for housing and employment. These models contain coefficients and elasticities developed from intensive statistical analyses of data sets representing a cross section of national housing and employment patterns (U.S. Army Corps of Engineers, 1988). Unlike the water-use models, these models are not calibrated by the user to reflect local socioeconomic conditions. Data for local conditions instead can be input to the System and override the parameter-generating algorithms.

For the purposes of this study, the base and calibration years are different years. Socioeconomic conditions changed significantly between 1990 and 1993, and sufficient data were available for 1993 to calibrate that year. Some base-year data are required to calibrate the water-use models in this study, and some base-year data are required by the socioeconomic models to quantify initial housing and employment conditions (table 4). At a minimum, a water-demand forecast requires the user to input total population, total employment, and **median income** for each future year. For this study, the number of housing units by type and employment statistics for several categories were projected externally and added to the System for the forecasts.

For the public/unaccounted sector and for the maximum-daily use dimension, water use was estimated externally to the System (fig. 2). Public/unaccounted water was calculated as a percentage of the total municipal water use (table 2). For the calibration year, the percentage reflects the unaccounted rate of use of water observed for the major public-water facility in each WSA; for the future years, the percentage remains constant through time for each WSA at 15 percent (water-industry average) (Verne Achtermann, water industry data base manager, American Association of Water Works, oral commun., 1995). This 15 percent is close to the default value for IWR-MAIN, 14.9 percent. For maximum-daily water use, the ratio of maximum-daily to annual average water use for 1988 (a drought year) was applied to the total water use for the calibration and future years (Hutson, 1993). This estimated amount represents maximum-daily water use under drought conditions.

Housing data

The residential module of the IWR-MAIN System was divided into two housing categories: metered and self-supplied. For each water-service area, the housing types are defined as follows:

- Metered consists of specified-occupied housing units (housing units built on less than 10 acres of land without commercial property attachment) (U.S. Department of Commerce, 1992a). These units are individually metered, but not necessarily sewered.
- The self-supplied category was principally used to manage the housing-unit data that were not included in the metered category. It includes housing units that depend on domestic wells (self-supplied) for their water; nonspecified housing units (housing units situated on 10 acres or more or housing units attached to a commercial establishment) (U.S. Department of Commerce, 1992a); and mobile homes. Self-supplied water demand is not included in the municipal totals for water demand for the basin. Self-supplied water-use models were not calibrated for this study, because no water-use data were collected to compare to a simulated water demand.

Base- and calibration-year data

Several sources of data were used in preparing base-year input values required to construct the residential water-use models (table 4). The U.S. Census Bureau provided data enumerating specified owner-occupied housing units by value range, renter-occupied units by range of contract rent, and occupied-housing units in the county and urban areas served by public-supply systems or served by sewerage (U.S. Department of Commerce, 1992a). The number of housing units by type (metered or self-supplied) and value range input to the model for the base year, 1990, resulted from a spatial analysis of these data sets. For the purposes of the modeling, it is assumed that all renter-occupied units are metered and that the renters pay the water bill. Renter-occupied units were combined with owner-occupied units by value range. Contract rent was converted to housing value using the following equation:

$$F = \frac{(1+i)^N - 1}{i(1+i)^N}, \qquad (6)$$

where

F is conversion factor,
i is 1980 present worth discounting factor, and
N is the number of months in mortgage period.

For the study area,

i is 0.006 (7.20 percent annual rate), Harry Slingerland, Credit Officer, Federal Reserve Bank, St.Louis, Missouri, written commun., 1994) and

N is 360.

Equivalent housing value expressed in 1980 dollars is as follows:

$$V = R \cdot F, \qquad (7)$$

where

F is conversion factor as defined in equation (6).

V is equivalent housing value, and

R is monthly rent.

For the calibration year (1993), estimates of the total number of occupied-housing units by WSA were provided by the Tennessee Housing Development Agency (Kimberly Clark, Senior Housing Research Analyst, written commun., 1994) (table 6). The type of data (U.S. Department of Commerce, 1992a) used to separate the housing units by type (metered or self-supplied) for 1990 were not available for 1993, therefore, the same proportion of housing unit by type as determined for the base year was used for 1993.

The housing density (fig. 2; table 4) for the central part of the City of Columbia was used as the housing density value for each WSA for the base, calibration, and future years. This density value is six units per acre (David Holderfield, Director of Grants and Planning, City of Columbia, oral commun., 1994). The model uses housing density to calculate the rate of use of water for irrigable land and, ultimately, summer demand (indoor use plus outdoor use) (equations 1, 2, and 3). In preliminary model runs, this housing density closely approximated the amount of water used on an average lot in 1993 in each WSA and, therefore, was used for each WSA to calibrate the expected summer usage.

The model is structured so that only one value for the housing density variable for the base,

Table 6. Summary of housing data input to the Institute for Water Resources–Municipal and Industrial Needs System for the residential model by water-service area and by modeling scenario

[--, no data input; scenario 1, steady growth from year to year; and scenario 2, selected higher growth in the residential sectors for the Maury/southern Williamson water-service area only]

Water-service area / Housing type	Occupied-housing units 1990	1993	2000	2015
		Scenario 1 and 2		
Bedford				
Metered	9,809	9,854	10,054	10,754
Self-supplied	--	1,704	2,107	1,951
Total	--	11,653	12,705	14,487
		Scenario 1 and 2		
Marshall				
Metered	6,333	6,552	6,695	7,584
Self-supplied	--	2,002	2,002	1,501
Total	--	8,554	9,086	10,360
		Scenario 1		
Maury/southern Williamson.				
Metered	18,732	20,931	21,381	23,147
Self-supplied	--	2,284	2,284	1,781
Total	--	23,125	24,929	29,218
		Scenario 2		
Maury/southern Williamson.				
Metered	18,732	20,931	21,382	24,833
Self-supplied	--	2,284	2,284	1,781
Total	--	23,215	26,614	32,164

calibration, and future years can be specified for each model run. This housing density value can change with time. However, changing the housing density value creates an alternative water-use scenario by changing summer demand. Model input for the pricing and the climate variables have the same limits, wherein each pricing and climate specification represents a new set of model conditions and an alternative water-use scenario.

The U.S. Census Bureau provided statistics for resident population (U.S. Department of Commerce, 1992b) for each WSA for the base year (1990). Calibration and future-year projections of population resulted from the application of a log-linear type regression model to census data. See section by G.E Schwarz in this report for an explanation of the methodology and the population projections for Bedford, Marshall, and Maury Counties.

The water and wastewater price-rate structures for each system in each WSA (data collected as part of this study) were used to specify annual and summer-season marginal price and to calculate bill difference for the base year. Rate structure information for the period 1993 was compiled and the rates were expressed in 1980 **constant dollars** (table 7).

Table 7. Marginal price and bill difference for water and wastewater, in 1980 dollars, for the metered housing category

Water-service area	Marginal price per thousand gallons	Bill difference per thousand gallons
Bedford	3.46	1.08
Marshall	3.85	.52
Maury/ southern Williamson.	1.93	4.29

The rate structure imposed by the largest public supplier in each WSA was adopted as the determining rate structure for water demand in that WSA. This technique was justified because, either this public supplier served most of the connections in the WSA, or it distributed water to other systems, influencing their rate structure. The selected systems were Columbia Water Department (Maury/southern Williamson WSA), Lewisburg Water System (Marshall WSA), and Shelbyville Water System (Bedford WSA).

Future-year data

For future years (2000 and 2015) the number of total housing units was generated externally to IWR-MAIN (fig. 2; table 4). The external method utilizes the projected resident population for each future year and the number of persons per household in 1990 (Bedford 2.59; Marshall, 2.57; and Maury, 2.62 persons per household) (U.S. Department of Commerce, 1992b) (table 8). The rate of expansion of public-supply service also was factored into future estimates of housing units.

For the calibration and future years, the number of housing units within a selected value range for a specified housing type was generated by an internal econometric housing model (table 7). In calculating the percentage of housing units for a selected value range, this housing model considered the rate of change in median income and in population from the base year to the future year (U.S. Army Corps of Engineers, 1988).

The only complete assessment of median household income in Tennessee occurs in each decennial census. The 1990 census provided base year (1990) median household income (U.S. Department of Commerce, 1992c) (table 9). For the calibration and future years, median household income was estimated using a multiplier derived from the average of the rate of change in **per capita income** in constant 1972 dollars from 1980 to 2015 as projected by the U.S. Department of Commerce, Bureau of Economic Analysis (Duanne Hackman, Regional Economic Office, written commun., 1992). The six rates of change that were projected for per capita income by the Bureau of Economic Analysis were used in conjunction with the 1989 median household income for each county as reported by the U.S. Department of Commerce, U.S. Census Bureau in 1990; and, with the 1993 median household income as reported by the Columbia Power System (Linsay Boyd, General Manager, Columbia Power System, written commun., 1994). For the purposes of the model input, the dollars are expressed as 1980 constant dollars.

Employment data

For the commercial sector, several **Standard Industrial Classification (SIC)** categories are grouped together because the water-use coefficients for the categories are similar. For industry, a more comprehensive data base (results from the inventory of the public-supply systems) and a greater range of coefficients for the categories resulted in more

Table 8. Estimated metered occupied-housing units by value range

[Scenario 1, steady growth from year to year; scenario 2, higher growth in the residential sector for the Maury/southern Williamson water-service area]

Water-service area Value range (1,000 dollars) (1980 constant dollars)	Housing units 1990	1993	2000	2015
Bedford		Scenario 1 and 2		
0.0 - 14.1	1,225	1,128	981	500
14.1 - 21.1	1,053	970	843	430
21.1 - 28.1	1,812	1,669	1,451	740
28.1 - 35.2	2,106	2,076	2,019	2,002
35.2 - 42.2	1,274	1,256	1,221	1,211
42.2 - 49.2	887	874	850	843
49.2 - 56.3	553	670	872	1,500
56.3 - 63.3	303	367	478	822
63.3 - 70.4	199	241	314	540
70.4 - 87.9	163	198	257	442
87.9 - 105.5	92	112	145	249
105.5 - 123.1	79	163	737	2,100
123.1 - 140.7	25	52	233	665
140.7 - 175.9	16	33	149	425
175.9 - 211.1	15	31	140	399
211.1 - 281.4	4	8	37	106
281.4 - 351.8	1	2	9	27
351.8 - 703.5	2	4	19	53
Total	9,809	9,854	10,754	13,054
Marshall		Scenario 1 and 2		
0.0 - 14.1	886	828	754	357
14.1 - 21.1	725	678	617	292
21.1 - 28.1	1,117	1,044	950	450
28.1 - 35.2	1,379	1,395	1,428	1,485
35.2 - 42.2	734	743	760	790
42.2 - 49.2	502	508	520	541
49.2 - 56.3	335	432	588	1,070
56.3 - 63.3	204	263	358	651
63.3 - 70.4	121	156	212	386
70.4 - 87.9	171	220	300	546
87.9 - 105.5	75	97	132	239
105.5 - 123.1	40	90	460	1,391
123.1 - 140.7	15	34	173	521
140.7 - 175.9	20	45	230	695
175.9 - 211.1	7	16	81	243
211.1 - 281.4	1	2	12	35
281.4 - 351.8	0	0	0	0
351.8 - 703.5	1	2	12	35
Total	6,333	6,552	7,802	9,975

division of the industrial categories than of the commercial categories.

Base- and calibration-year data

The employment model requires total employment data for each WSA for 5 years before the base year (1985), the base year (1990), the calibration year (1993), and each future year (2000 and 2015). The State Labor Force Summary provided total employment statistics from 1985 (5 years before the base year) to 1993 (Michael Ballard, Tennessee Department of Employment Security, written commun., 1994). For the base and calibration years, the Directory of Manufacturers (White, 1994) and the Tennessee labor force estimates from Tennessee Department of Employment Security (1990) provided a means of separating the employment

Table 8. Estimated metered occupied-housing units by value range—Continued

Water-service area Value range (1,000 dollars) (1980 constant dollars)	Housing units 1990	1993	2000	2015
Maury/southern Williamson		Scenario 1		
0.0 - 14.1	2,102	1,978	1,669	380
14.1 - 21.1	1,408	1,325	1,118	255
21.1 - 28.1	2,055	1,934	1,632	372
28.1 - 35.2	3,396	3,626	3,561	3,471
35.2 - 42.2	2,360	2,520	2,475	2,412
42.2 - 49.2	2,502	2,671	2,624	2,557
49.2 - 56.3	1,441	1,947	2,477	4,125
56.3 - 63.3	906	1,224	1,557	2,594
63.3 - 70.4	733	991	1,260	2,098
70.4 - 87.9	846	1,143	1,454	2,422
87.9 - 105.5	480	649	825	1,374
105.5 - 123.1	228	419	1,131	2,832
123.1 - 140.7	104	191	516	1,292
140.7 - 175.9	88	162	436	1,093
175.9 - 211.1	50	92	248	621
211.1 - 281.4	19	35	94	236
281.4 - 351.8	7	13	35	87
351.8 - 703.5	7	13	35	87
Total	18,732	20,931	24,473	29,849
Maury/southern Williamson 2		Scenario 2		
0.0 - 14.1	2,102	1,978	1,755	414
14.1 - 21.1	1,408	1,325	1,176	277
21.1 - 28.1	2,055	1,934	1,716	405
28.1 - 35.2	3,396	3,626	3,813	3,811
35.2 - 42.2	2,360	2,520	2,650	2,648
42.2 - 49.2	2,502	2,671	2,809	2,808
49.2 - 56.3	1,441	1,947	2,629	4,498
56.3 - 63.3	906	1,224	1,653	2,828
63.3 - 70.4	733	991	1,337	2,288
70.4 - 87.9	846	1,143	1,543	2,641
87.9 - 105.5	480	649	876	1,498
105.5 - 123.1	228	419	1,304	3,235
123.1 - 140.7	104	191	595	1,476
140.7 - 175.9	88	162	503	1,249
175.9 - 211.1	50	92	286	710
211.1 - 281.4	19	35	109	270
281.4 - 351.8	7	13	40	99
351.8 - 703.5	7	13	40	99
Total	18,732	20,931	26,254	32,920

Table 9. Median household income

Water-service area	Median household income expressed in 1980 constant dollars 1990	1993	2000	2015
Bedford	14,887	16,077	18,090	22,263
Marshall	15,039	16,417	18,473	22,734
Maury/ southern Williamson.	16,542	18,065	20,128	24,311

statistics for 2-digit SIC classifications into statistics for three-digit SIC categories.

Future-year data

Projected rates of growth by two-digit SIC group from year 1993 inventoried to 2005 were prepared by the Tennessee Labor Market Unit, Research and Statistics Unit, Tennessee Department of Employment and Security for the Middle Tennessee Substate areas (Sarah Caldwell, written commun., Tennessee

Department of Employment Security, 1994). Individual county statistics cannot be officially disclosed because of the confidentiality of the information. Employment projections for the study area, therefore, were based on the growth projected for the substate area.

For a scenario of steady growth in the commercial and industrial sectors in each WSA, the model input for the total number of employees for the study area is as follows: year 1985, 34,659; 1990, 45,310; 1993, 51,445; 2000, 56,453; and 2015, 67,853. For a scenario of higher growth in selected industrial categories for the Maury/southern Williamson WSA for years 2000 and 2015, the estimated number of employees for the study area is 58,803 and 70,203, respectively.

Model Calibration

Model calibration consists of using actual socioeconomic data for a period of time and simulating water use for various sectors of the municipal system such as residential, commercial, and industrial customers. Actual water-use data are compared to simulated results to determine calibration accuracy. Two major steps comprise the calibration process:

1. Initial simulation using all the default parameters of IWR-MAIN; and
2. Analysis of the pattern of errors resulting when the simulated water demand is compared to the actual values, then adjusting equations as needed.

The year 1993 was selected for calibration because water use was inventoried for the public-supply systems and industries in the study area for that year. The water-use inventory provided a guide to adjusting the residential, commercial, and industrial constants and coefficients of the water-use models. Industrial activity and the resulting water-use patterns were sufficiently different between 1990 and 1993 so that a model using the 1993 calibration data is more reliable for estimating future use.

For this study, IWR-MAIN estimates of the residential-water demand exhibit systematic errors of predicting actual water use in the calibration year. This pattern of prediction error indicates that residential water use in the study area is characterized by a lower (or higher) base use than that observed in the data that were used to derive the IWR-MAIN demand models. The simulated residential water demand for the Bedford and Marshall WSA's was lower and for the Maury/southern Williamson WSA, water demand was higher (table 10).

The winter and summer model constants representing gallons of water per household per day were adjusted to calibrate the seasonal models (table 11). The coefficients for the effective marginal price variable for summer and winter usage also were adjusted using regional (State of Kentucky) coefficients for the Bedford and Marshall WSA's (Eva Opitz, Director of Research, Planning and Management Consultants, Ltd., oral commun., 1994). The combined water and wastewater rates in these two WSA's exceeded the rates in the national data sets that were used to develop the System's price coefficients. The coefficient applied to the effective summer marginal price of water was changed from 285.5 to 160.9 as defined in equation (1). The effect in the equation is, for each 1.00 dollar increase in the price, water use is reduced 160.9 gallons for the summer season for each residential unit.

The coefficient applied to the effective annual marginal price of water was changed from 45.9 to 37.55 as defined in equation (4). The effect in the equation is, that for each 1.00 dollar increase in the price, water use is reduced 37.55 gallons for the winter season for each residential unit. Changing the elasticity reduces the effect of the rate of change on the amount of water used.

Summer and annual water usage for each WSA was calibrated to yield ratios of 1.06 (Bedford), 1.03 (Marshall), and 1.07 (Maury/southern Williamson) of summer to annual use to agree with the summer (June-August) use to annual use ratio for the largest public-supplies in each WSA in 1993. These systems closely mirrored the seasonal water use of the other major water systems in the respective WSA.

For the industrial and commercial single-coefficient requirement models (water-use per employee) the results of the initial calibration revealed the need to adjust IWR-MAIN default coefficients from the Library of Coefficients (U.S. Army Corps of Engineers, 1988). The commercial sector required the greatest adjustments. Per employee rates of use for the largest utility customers were verified using the data compiled from the inventory of 1993 usage. These included industries with high employment, high projected rate of growth, or large quantity users. Changes to the coefficients reflect local per employee use for a specific SIC or represent average employee use for combined SIC categories.

Model Reliability

The constants and coefficients in the water-use models were generally reliable in estimating residential

Table 10. Observed and simulated average annual water demand for metered housing for 1993

[Mgal/d, million gallons per day]

Water-service area	Observed value (Mgal/d)	Simulated water demand without adjustments (Mgal/d)	Simulated water demand with adjustments (Mgal/d)
Bedford	1.75	1.24	1.75
Marshall	.96	.93	.96
Maury/ southern Williamson.	4.15	4.21	4.15

Table 11. Model and calibration constants for the metered housing models

[WSA, water-service area]

Metered housing models	Model constant	Calibration constant		
		Bedford WSA	Marshall WSA	Maury/ southern Williamson WSA
Winter season	234	247	234	218
Summer season	385	499	503	486

water demand. The adjustment to the constant (y-intercept) for the residential equations and the use of regional coefficients may create some concern about reliability of the model. The alternative is to develop extensive local data sets not currently available to calculate a local set of coefficients. The developers of the model recommend adjusting the equations by the methods described rather than attempting to develop local data sets with limited resources (Eva M. Opitz, Director of Research, Planning and Management Consultants, Ltd, oral commun., 1994).

The housing model calculates the percentage of change in the number of housing units within a value range based on the change in median income and the change in population from the base year to the future year. This assumed correlation is problematic, in part, because the actual correlation between housing units and median income is uncertain. Additionally, this correlation is weak because the projected median income is based on projected average per capita income (table 9). These two measures of income are not the same. However, per capita income was used in this study because it is the best of the available projected income data. In Maury County, for example, the percentage of change in projected median income from 1993 to 2015 is 35 percent ($18,065 to $24,311) (table 9). Although the change in the dollar amount is small ($6,246), the percentage of change is large enough so that for the housing value ranges greater than $100,000, the shift is significant (table 8). For the base year (1990) about 3 percent (503 units) of the metered housing units are valued at greater than $100,000; for 2015, 21 percent (6,248 units) of the metered housing units exceed this value. In the residential equations in the water demand model, as housing value increases, water use increases (U.S. Army Corps of Engineers, 1988). Per capita use values illustrate this effect. For the Maury/southern Williamson WSA, per capita use increases 43 percent from 1993 to 2015 (from 67 to 96 gal/d) (table 12). Overall comparisons of the simulated residential water demand for the basin, however, are acceptable and represent the specified model assumptions.

As with any model, the degree of certainty lessens the further out in time that the projections are made. Projecting 25 years involves assuming many political, environmental, economic, and technical factors will not shift radically. For this study, no assumptions about conservation or severe restrictions or a permitting system were evaluated.

Table 12. Per capita use for the residential sector

[Scenario 1, steady growth; scenario 2 , higher growth in the residential sector in the Maury/southern Williamson water-service area; --, no data]

Water-service area	Per capita use, in gallons per person per day 1993 (actual)	2000 (simulated)	2015 (simulated)
		Scenario 1 and 2	
Bedford	57	66	87
Marshall	42	54	81
		Scenario 1	
Maury/ southern Williamson.	67	79	96
		Scenario 2	
Maury/ southern Williamson.	--	79	98

Model Results

The IWR-MAIN models that were calibrated for the Bedford, Marshall, Maury/southern Williamson water-service areas were used to simulate water demand for years 2000 and 2015. The estimates for each water-service area are aggregated to yield basin totals for the upper Duck River. The results of the simulation for scenario 1 (table 13) show that:

- Simulated average water demand in the basin could increase 10 percent by 2000 and 57 percent by 2015.
- Residential water demand could increase 24 percent by 2000 and 90 percent by 2015.
- Commercial water demand could increase 23 percent by 2000 and 58 percent by 2015.
- Industrial water demand could increase 13 percent by 2000 and 41 percent by 2015.

The results of the simulation for scenario 2 show that average water demand could increase 16 percent by 2000 and 71 percent by 2015.

Single-Coefficient Requirement Model

Water demand for the Bedford, Marshall, and Maury/southern Williamson WSA's for years 2025, 2035, and 2050 was estimated using a single-coefficient requirement model expressed as

$$Q = C \cdot P, \qquad (8)$$

where

Q is average annual water use, in gallons per day;

C is a water-use coefficient (gross per capita use); and

P is population.

This single-coefficient requirement model was devised to extend the forecast to 2050. The mathematical structure of the IWR-MAIN System limits projections of the socioeconomic parameters to 25 years, for example, 1990 to 2015.

Gross per capita use for each WSA for year 2015 was multiplied by a projected population to calculate water demand (table 13). The gross per capita values for 2015 were derived from sector estimates of water demand produced by the IWR-MAIN System. For the Bedford WSA, the gross per capita value is 205; Marshall, 176; Maury/southern Williamson scenario one, 225; and Maury/southern Williamson scenario two, 219 gallons per person per day.

Population projections for Bedford, Marshall and Maury counties for 2025, 2035, and 2050 are based on published census data for the years 1960 to 1992. Population data for census years 1960, 1970, 1980, and 1990 are based on census count data. Data for intercensus years 1961-68, 1971-78, 1981-88, and 1991-1992 are estimated by the Bureau of Census on the basis of ancillary data, such as birth and death rates and migration statistics. No population data are available for the years preceding the decennial census for 1969, 1979, and 1989. Population projections for southern Williamson County for a steady and higher rate of growth for 2025, 2035, and 2050 were provided by the Williamson County Planning Commission (Joseph Horne, Director, written commun., 1994).

Table 13. Simulated water demand for Bedford, Marshall, and Maury/southern Williamson water-service areas for 1993, 2000, 2015, 2025, 2035, and 2050

[WSA, water-service area; --, not simulated]

Sector	1993	2000	2015	2025	2035	2050	Percent increase from 1993
	in million gallons per day						
Bedford WSA scenario 1 and 2							
Residential	1.75	2.2	3.3	--	--	--	--
Commercial	.72	1.0	1.3	--	--	--	--
Industrial	1.08	1.2	1.3	--	--	--	--
Public/unaccounted	1.04	.8	1.0	--	--	--	--
Total water demand	4.59	5.2	6.9	7.6	8.2	9.4	105
Maximum daily use	6.75	7.8	10	11	12	14	107
Marshall WSA scenario 1 and 2							
Residential	0.96	1.3	2.1	--	--	--	--
Commercial	.50	.6	.8	--	--	--	--
Industrial	.45	.5	.9	--	--	--	--
Public/unaccounted	.61	.4	.7	--	--	--	--
Total water demand	2.52	2.8	4.5	4.8	5.2	6.0	138
Maximum daily use	3.83	4.2	6.8	7.2	7.8	9.0	135
Maury/southern Williamson WSA scenario 1							
Residential	4.15	5.0	7.3	--	--	--	--
Commercial	1.38	1.6	2.0	--	--	--	--
Industrial	2.81	3.2	3.9	--	--	--	--
Public/unaccounted	1.76	1.7	2.3	--	--	--	--
Total water demand	10.1	12	16	17	19	22	118
Maximum daily use	15.6	18	24	26	29	33	112
Maury/southern Williamson WSA scenario 2							
Residential	4.15	5.5	8.1	--	--	--	--
Commercial	1.38	1.6	2.0	--	--	--	--
Industrial	2.81	3.4	4.2	--	--	--	--
Public/unaccounted	1.47	1.9	2.5	--	--	--	--
Total water demand	10.1	12	17	20	23	27	167
Maximum daily use	15.6	19	26	30	35	40	156
Upper Duck River study area scenario 1							
Residential	6.86	8.5	13.0	--	--	--	--
Commercial	2.60	3.2	4.1	--	--	--	--
Industrial	4.34	4.9	6.1	--	--	--	--
Public/unaccounted	2.44	2.9	4.0	--	--	--	--
Total water demand	17.2	19	27	30	33	38	121
Maximum daily use	25.8	29	40	45	50	57	121
Upper Duck River study area scenario 2							
Residential	6.86	9.0	14.0	--	--	--	--
Commercial	2.60	3.2	4.1	--	--	--	--
Industrial	4.34	5.1	6.4	--	--	--	--
Public/unaccounted	2.43	3.1	4.2	--	--	--	--
Total water demand	17.2	20	29	33	36	43	150
Maximum daily use	25.8	30	44	50	54	64	148

Regression analysis was used to relate the population data to time. The choice of a functional form of regression analysis for relating population to time is important because the projections are carried out for a long period of time (58 years, 1993 to 2050). Three types of regression models were evaluated. When graphed, the Box-Cox model yields projections that show a large negative curvature so that out-of-sample projections quickly go to infinity. The large negative curvature is partially a consequence of using ancillary data to estimate population between census years. Special cases of the Box-Cox model (the linear and log-linear models) yield more reasonable and similar projections. In the final analysis, the projections generated by the log-linear model were used (see section by G.E. Schwarz).

The population projection for each WSA was modified to reflect the number of persons (households) served by public supply. The percentage of the population served for 2025, 2035, and 2050 is the same percentage as determined for 2015. For Bedford, 90 percent; Marshall, 94 percent; and Maury/southern Williamson WSA, 97 percent. For scenario two, additional population is added to the log-linear projection of population for Maury County to account for additional industrial growth.

The results of the simulation for the upper Duck River basin study area for 2050 (table 13) show that:

- For scenario one (steady growth), average water demand could increase 121 percent to 38 Mgal/d.
- For scenario two (higher growth in selected sectors for the Maury/southern Williamson WSA), average water demand could increase 150 percent to 43 Mgal/d.

SUMMARY

Municipal water use in the upper Duck River basin study area increased 19 percent from 1980 to 1993 (14.5 Mgal/d to 17.2 Mgal/d). Socioeconomic data for the area suggest that water demand will continue to increase in response to residential, industrial, and commercial development. Officials from the Duck River Development Agency are concerned whether the capacity of the river can meet future water demands. In an attempt to address this concern, an investigation was conducted during 1994 by the USGS in cooperation with the Duck River Development Agency to determine potential future water demands. Methods used in this study can be applied nationwide to estimate water use and project water demand.

The study area includes about 1,500 square miles of the Duck River basin, extending from Normandy Dam to the western Maury County line. Included in the study area are Bedford County, Marshall County, and Maury/southern Williamson Counties. The 3-day 20-year low-flow discharge of the upper Duck River has increased as a result of controlled releases since the construction of Normandy Reservoir.

Public-water systems withdrawing from the Duck River supplied an average of 15.6 Mgal/d (about 91 percent of the total water demand) in 1993 to utilities in the Bedford, Marshall, and Maury/southern Williamson water-service areas (WSA's). The balance (about 9 percent or 1.66 Mgal/d) was supplied from springs and wells. The public-water systems delivered water for residential, commercial, and industrial uses to the cities of Lewisburg, Shelbyville, Columbia, Mount Pleasant, and several smaller communities in the basin.

Water demand in the basin for the year 2050 was estimated using the Institute for Water Resources-Municipal And Industrial Needs System (IWR-MAIN) and a single-coefficient requirement model. Water demand was projected based on steady growth throughout the study area (scenario 1) and increased growth starting in 2000 in the Maury/southern Williamson WSA (scenario 2).

The IWR-MAIN was calibrated to 1993 conditions and was used to estimate water demand for 2015. Gross per capita use was derived from the water-demand estimate for 2015 for each water-service area. For the Bedford WSA, the gross per capita use is 205; Marshall, 176; Maury/southern Williamson scenario 1, 225; and Maury/southern Williamson scenario 2, 219 gallons per person per day.

Census data for each WSA were used in a log-linear regression analysis to project population in each WSA to 2050. A single-coefficient model was devised to combine water use with projected population and extend the water-use forecast to 2050. Gross per capita use for each WSA for year 2015 was multiplied by a projected population to calculate water demand. Water demand for each water-service area was combined to yield a total for the basin.

Results from the water-use models indicate that water demand for the upper Duck River study area for the years indicated could increase as follows:

Year	Scenario 1		Scenario 2	
	Total demand (Mgal/d)	Percent increase from 1993	Total demand (Mgal/d)	Percent increase from 1993
1993	17.2	--	17.2	--
2015	27	57	29	69
2050	38	121	43	150

Methodology Used to Develop Population Forecasts for Bedford, Marshall, and Maury Counties, Tennessee, From 1993 Through 2050

by

Gregory E. Schwarz

Methodology Used to Develop Population Forecasts for Bedford, Marshall, and Maury Counties, Tennessee, From 1993 Through 2050

By Gregory E. Schwarz

This section describes the methods used to develop population forecasts for Bedford, Marshall, and Maury Counties, Tennessee, from 1993 through 2050. Three approaches for estimating the relation between population and time are compared. These approaches include the Box-Cox, the log-linear, and the linear regression methods. The Box-Cox model allows for a general nonlinear relation between population and time and includes, as special cases, the linear and log-linear models. The results from the three approaches indicate that the best model for forecasting population into the distant future is the log-linear model.

The Data

The forecasts are based on published census data for the years 1960 to 1992. Population data for census years 1960, 1970, 1980, and 1990 are based on census count information. Data for the intercensus years from 1961 through 1968, 1971 through 1978, 1981 through 1988, and 1991 through 1992 are estimated by the census on the basis of ancillary data. No population data are available for the years 1969, 1979, and 1989. Census estimates of population for the three counties between 1960 and 1992 are presented in table 14.

Method of Estimation

Regression analysis was used to relate the population data to time. Because the forecast is carried out over a long period (58 years), the choice of a functional form relating population to time is critical. To facilitate the estimation of a correct functional form, a Box-Cox transformation (Davidson and MacKinnon, 1993, p. 483-507) of the population variable is used. The functional form with the Box-Cox transformation applied to the dependent variable is

$$\frac{P_t^{\lambda} - 1}{\lambda} = \beta_0 + \beta_1 T_t + e_t, \qquad (1)$$

where P_t is the population for observation t; T_t is the year corresponding to observation t; e_t is an unobserved error term, assumed independent and normally distributed with mean of 0 and variance of σ^2_e; and λ, β_0, and β_1 are constants to be estimated.

There are two approaches to estimating the coefficients λ, β_0, and β_1 in equation 1—maximum likelihood (ML) and nonlinear instrumental variables (NLIV). In practice, ML estimates can be obtained by rescaling the population variable. This is accomplished by dividing population by the geometric mean of population for the sample and estimating the resulting equation using nonlinear least squares (NLLS). The resulting estimate of λ is identical to the ML estimate, although its estimated standard error is not (the coefficients β_0 and β_1 in the rescaled model are transformed by the scale change and, thus, must be retransformed to obtain ML estimates). To calculate the standard error of the ML estimate of λ, we use the method of double-length artificial regression (DLR) (Davidson and MacKinnon, 1993, p. 498-500). An alternative approach to ML (or NLLS) estimation is to estimate the rescaled version of the model using NLIV (Amemiya, 1985, p. 249-252). Under this approach, DLR is not needed because the standard error of λ is consistently estimated.

The form of equation 1 permits a number of commonly used functional forms. The linear and log-linear models are simply restricted versions of the equation. For the linear model, λ is set equal to 1, and

Table 14. Populations of Bedford, Marshall, and Maury Counties, Tennessee, 1960-92

[N.A., not available. Sources: U.S. Bureau of the Census, (1967), (1968), (1969), (1973), (1974), (1975), (1976), (1978), (1979), (1988a), (1988b), (1990), and (1994)]

Year	Number		
	Bedford County	Marshall County	Maury County
1960	23,223	16,914	41,789
1961	23,534	16,534	41,862
1962	23,412	16,614	43,356
1963	23,923	17,501	43,770
1964	24,282	17,722	44,160
1965	24,394	17,973	45,019
1966	25,052	17,774	45,275
1967	25,383	18,091	45,753
1968	26,581	18,489	44,807
1969	N.A.	N.A.	N.A.
1970	25,039	17,319	44,028
1971	24,700	17,400	43,600
1972	25,100	17,500	44,200
1973	25,600	17,700	44,800
1974	25,700	17,900	45,200
1975	25,800	18,100	45,200
1976	26,000	18,000	46,400
1977	25,900	18,300	47,500
1978	26,390	18,312	47,895
1979	N.A.	N.A.	N.A.
1980	27,916	19,698	51,095
1981	27,900	19,800	51,600
1982	28,200	19,800	51,600
1983	28,200	19,700	51,500
1984	28,600	20,400	52,300
1985	28,900	20,600	53,300
1986	29,200	20,800	54,100
1987	29,300	21,200	55,000
1988	29,500	21,300	55,300
1989	N.A.	N.A.	N.A.
1990	30,411	21,539	54,812
1991	31,053	22,469	58,131
1992	31,738	22,974	59,740

the intercept term is used to absorb the -1 appearing in the left-hand side of the equation. The estimated model is

$$P_t = \beta_0 + \beta_1 T_t + e_t. \quad (2)$$

This model can be estimated by ordinary least squares.

Conversely, the log-linear model can be obtained by taking the limit of equation 1 as λ goes to infinity to obtain

$$\ln(P_t) = \beta_0 + \beta_1 T_t + e_t, \quad (3)$$

where $\ln(\cdot)$ denotes the natural logarithm. To understand this derivation, take the limit of $(P_t^\lambda - 1)/\lambda$ as λ goes to zero using L'hospital's rule from elementary calculus. This model, too, can be estimated by ordinary least squares.

Method of Forecasting

Population forecasts from the Box-Cox model were estimated by the equation

$$\hat{P}_t = \left[\hat{\lambda}\left(\hat{\beta}_0 + \hat{\beta}_1 T_t\right) + 1\right]^{1/\hat{\lambda}}, \quad (4)$$

where $\hat{P}_t$ is the forecast of population for period t; and $\hat{\lambda}, \hat{\beta}_0$, and $\hat{\beta}_1$ are ML or NLIV estimates of λ, β_0, and β_1. This form of the forecast equation is technically incorrect because it does not fully account for the effects of random error arising from the e_t term and error in the parameter estimates. However, the forecast should be relevant for indicating the general performance of the Box-Cox transformation.

Projections for the linear and log-linear models are based on the following equations

$$\text{Linear model: } \hat{P}_t = \hat{\beta}_0 + \hat{\beta}_1 T_t \quad (5)$$

and

$$\text{Log-Linear model: } \hat{P}_t = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 T_t + \frac{\hat{\sigma}_t^2}{2}\right), \quad (6)$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated values of β_0 and β_1. The term $\hat{\sigma}_t^2$, appearing in equation 6, is the variance of the forecast error. This variance is observation-dependent and is given by

$$\hat{\sigma}_t^2 = \hat{\sigma}_e^2\left(1 + \frac{V(T_s) + (T_t - T_s)^2}{NV(T_s)}\right), \qquad (7)$$

where $\hat{\sigma}_e^2$ is the estimated variance of the regression residual; N is the sample size; $V(T_s)$ is the variance of T for in-sample observations (that is, observations used to estimate the regression); and T_s is the average of the in-sample T's. Note that the variance given in equation 7 is for out-of-sample forecasts only. For in-sample forecasts, the variance is equal to $2\hat{\sigma}_e^2$ minus the quantity expressed in the righthand side of the equals sign in equation 7. Note also that the forecast equation for the log-linear model given in equation 6 is biased in small samples due to the use of an estimated value for σ_e^2 (as opposed to a known value). However, given the small value of this variance in practice (see below), the magnitude of the bias is small and can reasonably be ignored.

Confidence intervals for the linear-model forecast are given by

Confidence Interval Upper Bound:

$$\hat{P}_t + t_{\alpha/2}(N-2)\sqrt{\hat{\sigma}_t^2}$$

and (8)

Confidence Interval Lower Bound:

$$\hat{P}_t - t_{\alpha/2}(N-2)\sqrt{\hat{\sigma}_t^2},$$

where $\hat{P}_t$ is forecasted population as given by equation 5; $t_{\alpha/2}(N-2)$ is the critical level for a t statistic with N - 2 degrees of freedom at significance level α; and $\hat{\sigma}_t^2$ is the forecast variance as given in equation 7 using the $\hat{\sigma}_e^2$ estimated from the linear model.

The confidence interval for the log-linear model is derived in a more complicated manner. Observation-dependent bounds L_t and U_t are needed such that the difference $U_t - L_t$ is minimized subject to the constraint that the probability of the event $L_t \leq P_t \leq U_t$ equals $1 - \alpha$. Due to the skewed nature of the log-normal distribution for the error term, the forecasted value will not necessarily lie halfway between the lower and upper bounds. For the case where σ_e^2 is known, the lower and upper bounds satisfy the relations

Confidence Interval Lower Bound:

$$\hat{P}_t \exp\left(-\sqrt{\hat{\sigma}_t^2}\left(Z_t + \frac{5}{2}\sqrt{\hat{\sigma}_t^2}\right)\right)$$

and (9)

Confidence Interval Upper Bound:

$$\hat{P}_t \exp\left(\sqrt{\hat{\sigma}_t^2}\left(Z_t - \frac{1}{2}\sqrt{\hat{\sigma}_t^2}\right)\right),$$

where $\hat{\sigma}_t^2$ is given in equation 7 and Z_t solves the equation

$$\frac{1}{\sqrt{2\pi}}\int_{-z_t - 2\sqrt{\hat{\sigma}_t^2}}^{z_t} e^{-\varepsilon^2/2}\, d\varepsilon = 1 - \alpha. \qquad (10)$$

A derivation of these equations is given in Appendix A.

Model Estimation and Population Forecasts

This section presents the results from estimating the Box-Cox, linear, and log-linear models. The estimated models are used to generate population forecasts for Bedford, Marshall, and Maury Counties, Tennessee.

Box-Cox Model

Table 15 gives ML and NLIV estimates of λ and associated standard errors for Bedford, Marshall, and Maury Counties. Instruments used for the NLIV procedure are a constant, time, and time-squared. Estimated values of λ are consistently negative and, in most cases, significantly different from 0. All the estimates are significantly different from 1.0, implying systematic rejection of the linear model.

Population forecasts for the ML and NLIV models for the three counties are shown in figures 4 to 6. For comparison, the population projections for the linear and log-linear models also are shown (see below for details on these projections), as well as projections for 1995, 2000, 2005, 2008, and 2010 from a study by the University of Tennessee (Economic and Community Staff, 1992). The figures show clear problems with the Box-Cox model. All projections, regardless of the

Table 15. Maximum likelihood and nonlinear instrumental variables estimates of the Box-Cox transformation parameter

[λ, Box-Cox transformation parameter. Asymptotic standard errors in parentheses]

County	Maximum likelihood estimate of λ	Nonlinear instrumental variables estimate of λ
Bedford	-1.809[a]	-3.088[a b]
	(0.978)	(1.212)
Marshall	-2.900[a b]	-7.507[a b]
	(1.266)	(2.390)
Maury	-2.931[a b]	-6.195[a b]
	(1.095)	(1.852)

[a] Significantly different from 1 at 5-percent level of significance.
[b] Significantly different from 0 at 5-percent level of significance.

Table 16. Maximum likelihood and nonlinear instrumental variables estimates of the Box-Cox transformation parameter using census year data only (N = 4)

[λ, Box-Cox transformation parameter. Asymptotic standard errors in parentheses]

County	Maximum likelihood estimate of λ	Nonlinear instrumental variables estimate of λ
Bedford	-0.633	-0.679
	(1.177)	(1.690)
Marshall	-2.886	-7.261
	(4.656)	(11.081)
Maury	-0.326	-1.413
	(3.637)	(6.302)

estimation method, reach an asymptote early in the forecast horizon. A possible cause of this problem is the use of intercensus year data to estimate the model. Unlike census year data, which are determined from a full count of the population, intercensus year data are estimated from ancillary data such as birth rates, death rates, and so forth. The use of ancillary data implies the intercensus year population estimates are not as accurate as the census year estimates, with larger errors occurring in the years just before a census year. This is demonstrated in the data as large changes in population between a census year and the preceding year (see the graphs labeled "Actual data" in figures 4 to 6). It is possible that in trying to fit the large curvature in the data between census years, the Box-Cox model generates large negative estimates for λ.

To assess the importance of the intercensus year estimates, the Box-Cox model was reestimated using only the four census year population estimates. The resulting estimates of λ are presented in table 16, and population projections are shown in figures 7 to 9. With the exception of Marshall County, the estimates of λ are much less negative. None of the estimates are significantly different from 0 or from 1, which is probably due to the smaller estimated values of λ and the fewer number of data points entering the regressions. With the exception of Marshall County, the projections are much more stable and do not attain an asymptote within the forecast period. Also, the Box-Cox forecasts do not differ significantly from those of the log-linear model.

Linear and Log-Linear Models

The unrealistically large curvatures generated by the Box-Cox model suggest that attention should be focused on the linear and log-linear models. The negative values for λ obtained from the models estimated using only census year data tend to support a model with curvature, such as the log-linear model. Table 17 presents results for the linear and log-linear models using all the data. Results based only on census year data are reported in table 18. All the results show the time trend variable to be significantly different from 0, and the coefficients of correlation (R^2) indicate good regression fits.

Projections and 95-percent confidence intervals for Bedford, Marshall, and Maury Counties, using the linear and log-linear models and estimated using all the data, are given in figures 10 to 12. Projections and 95-percent confidence intervals for models estimated from only census year data are given in figures 13 to 15. Generally, the forecasts of the linear and log-linear models are very similar. All show an upper trend in population with a characteristic widening of the confidence intervals for forecasts further into the future. With the exception of later year forecasts for Bedford County, the confidence interval of the linear model overlaps the confidence interval of the log-linear

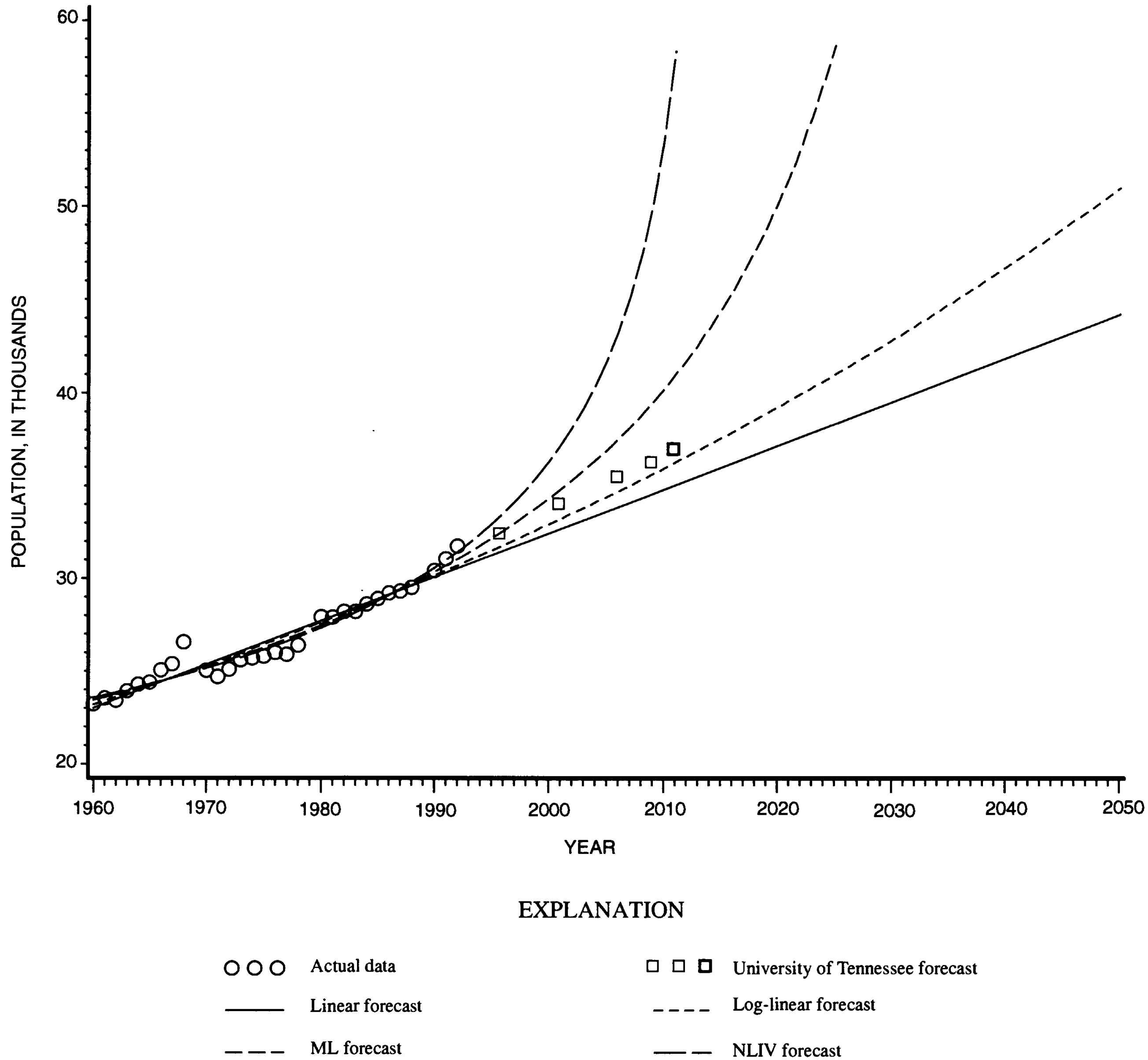


Figure 4. Actual and projected populations for Bedford County, Tennessee, for the linear, log-linear, maximum likelihood (ML), and nonlinear instrumental variables (NLIV) models.

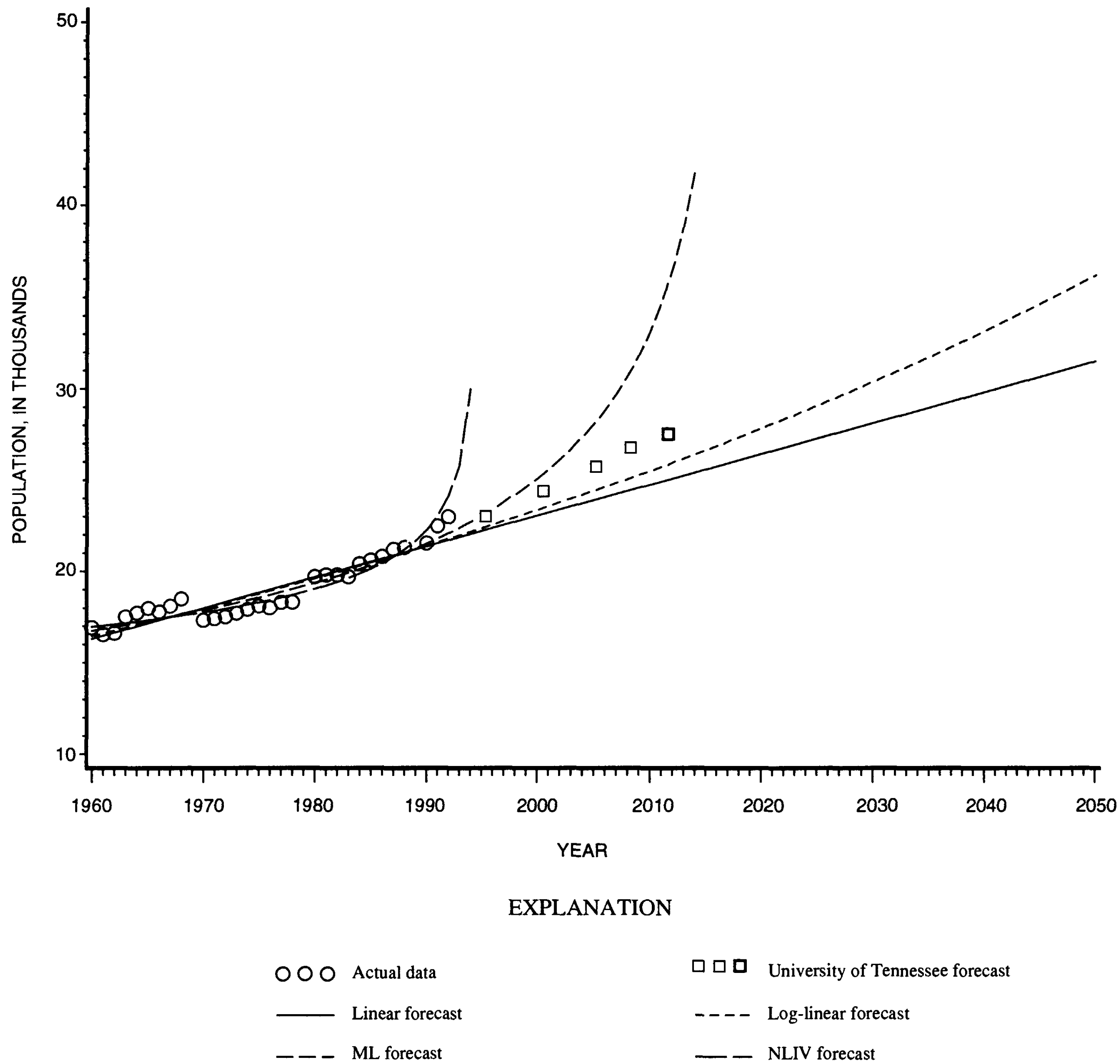


Figure 5. Actual and projected populations for Marshall County, Tennessee, for the linear, log-linear, maximum likelihood (ML), and nonlinear instrumental variables (NLIV) models.

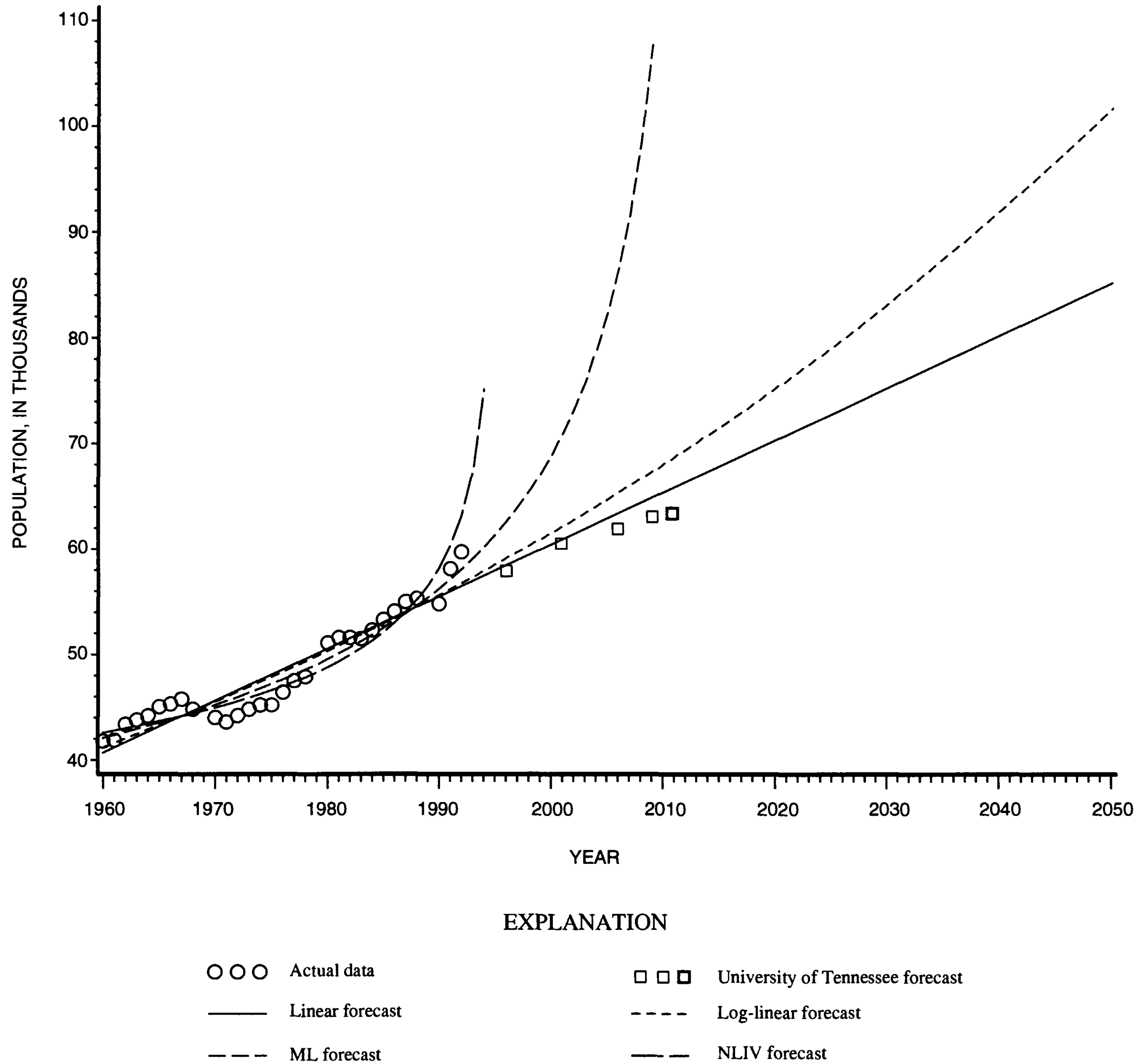


Figure 6. Actual and projected populations for Maury County, Tennessee, for the linear, log-linear, maximum likelihood (ML), and nonlinear instrumental variables (NLIV) models.

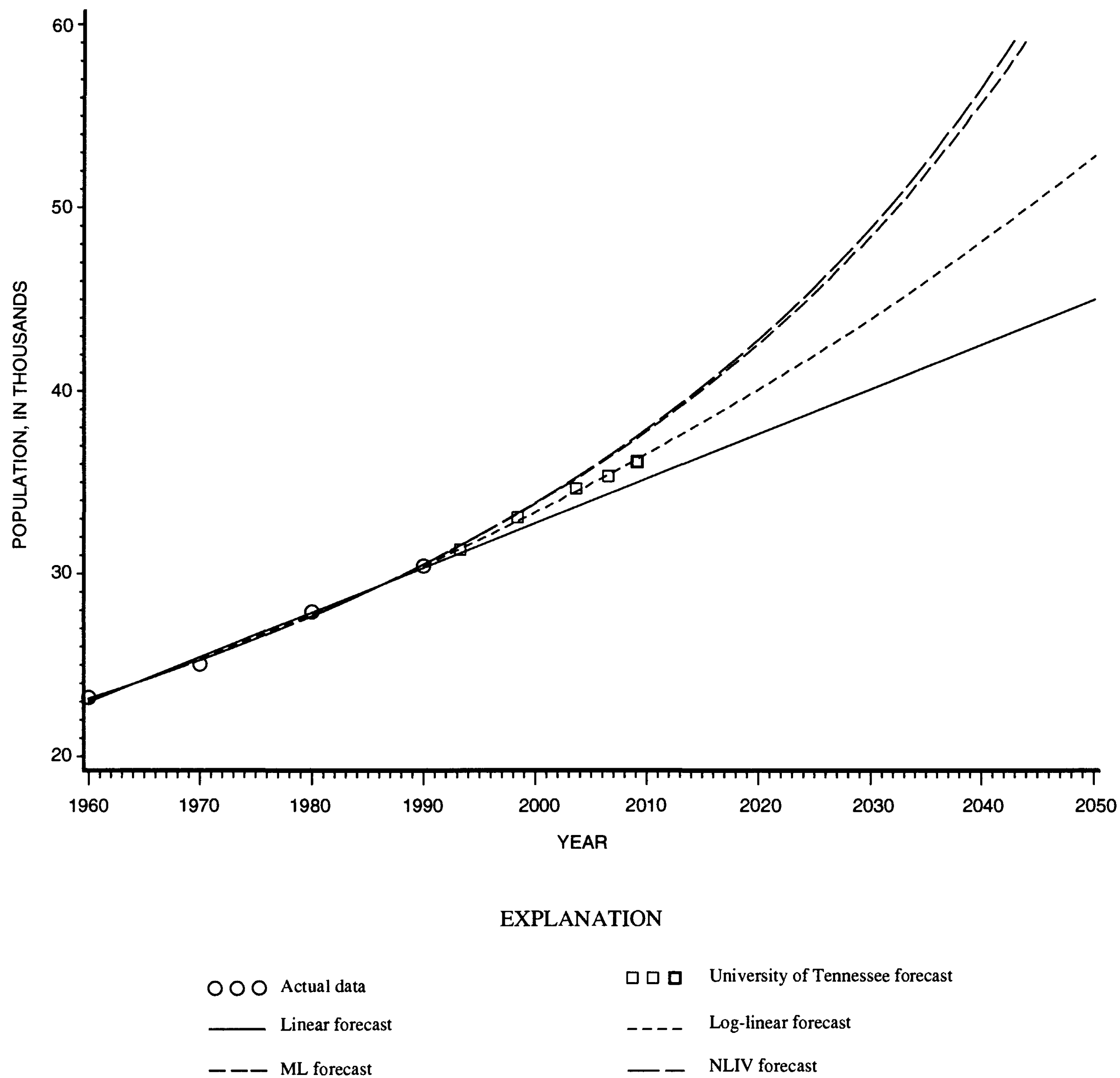


Figure 7. Actual and projected populations for Bedford County, Tennessee, using census-year data only for the linear, log-linear, maximum likelihood (ML), and nonlinear instrumental variables (NLIV) models.

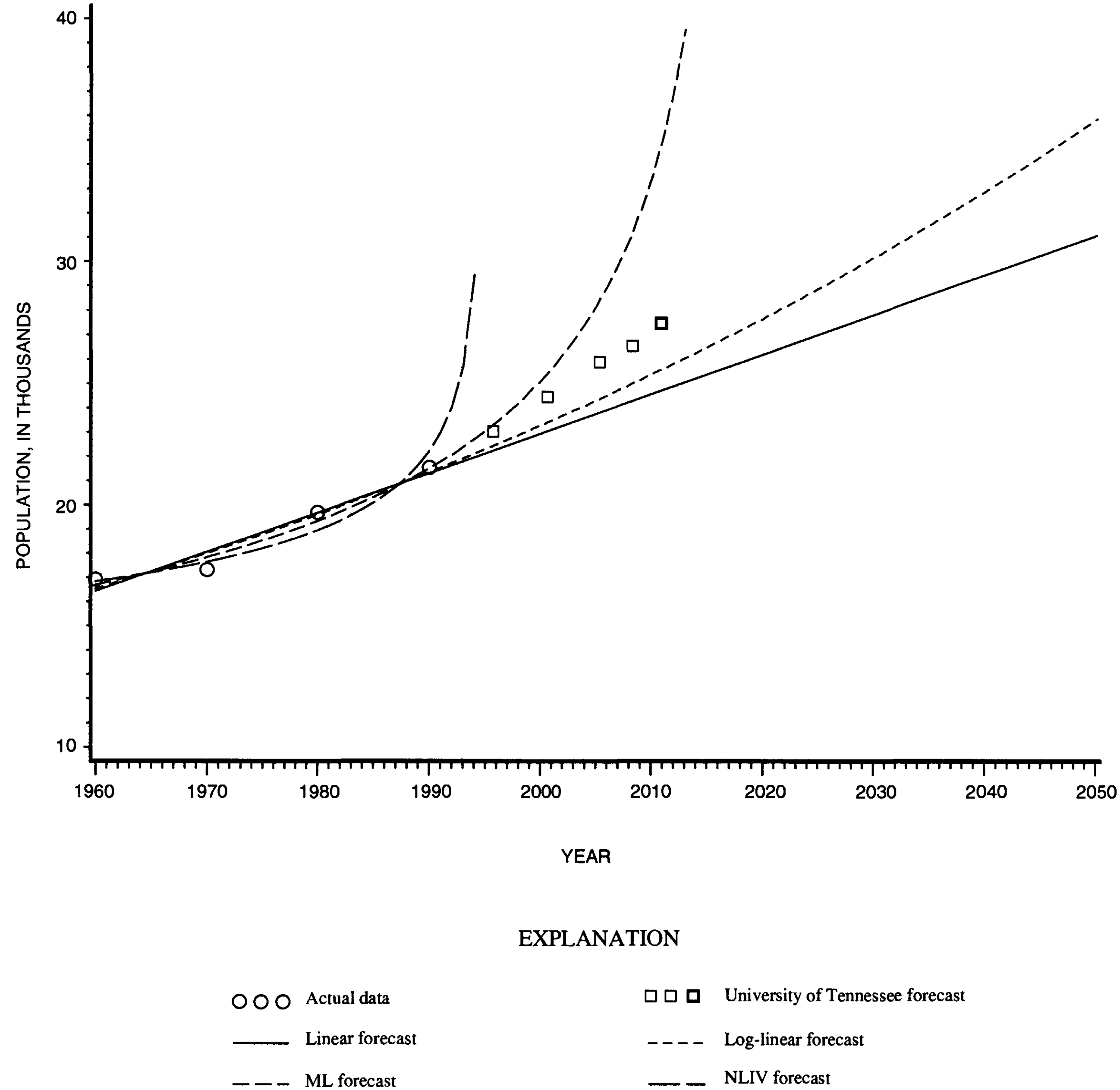


Figure 8. Actual and projected populations for Marshall County, Tennessee, using census-year data only for the linear, log-linear, maximum likelihood (ML), and nonlinear instrumental variables (NLIV) models.

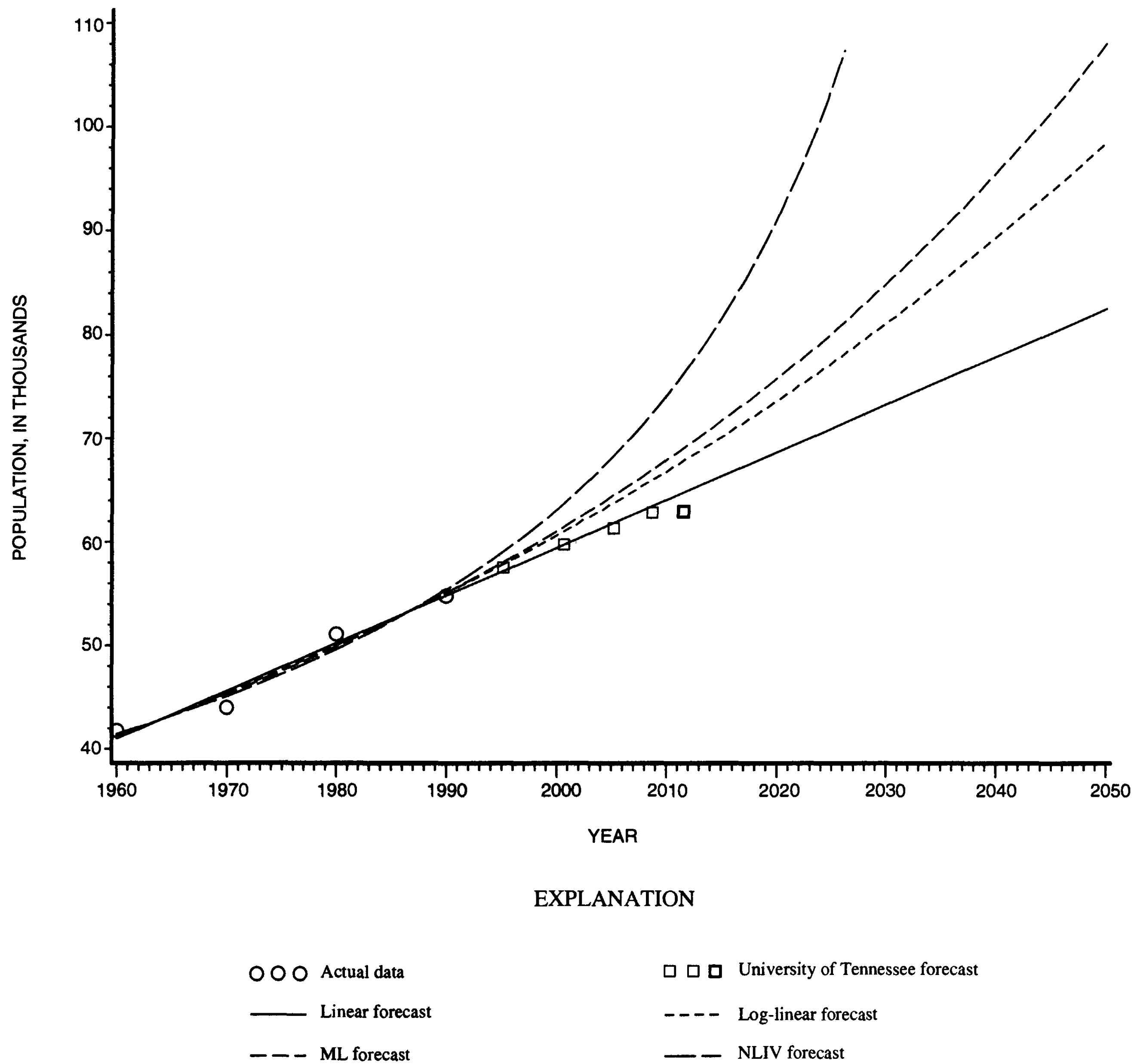


Figure 9. Actual and projected populations for Maury County, Tennessee, using census-year data only for the linear, log-linear, maximum likelihood (ML), and nonlinear instrumental variables (NLIV) models.

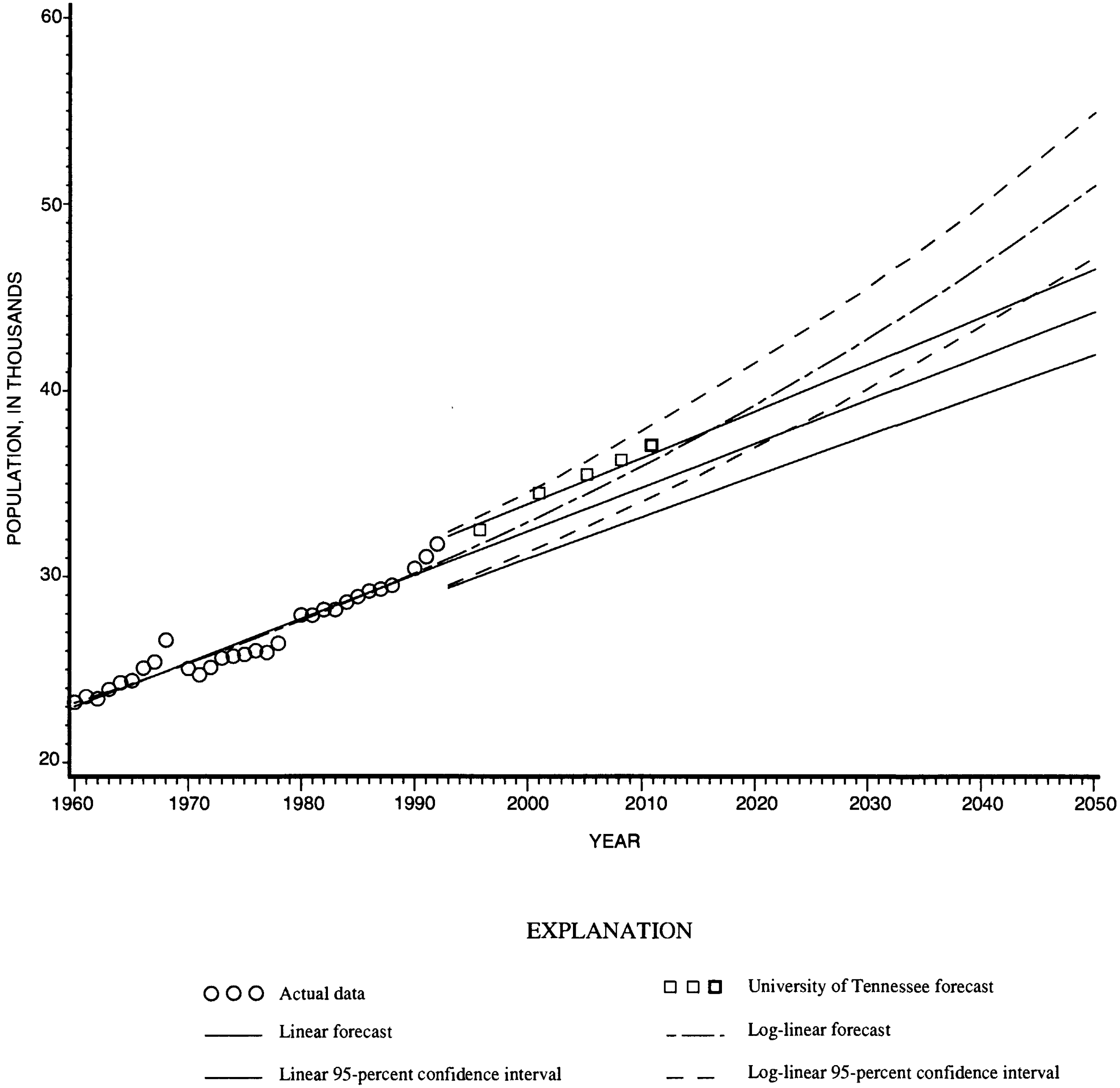


Figure 10. Actual and projected populations for Bedford County, Tennessee, for the linear and log-linear models, with confidence intervals.

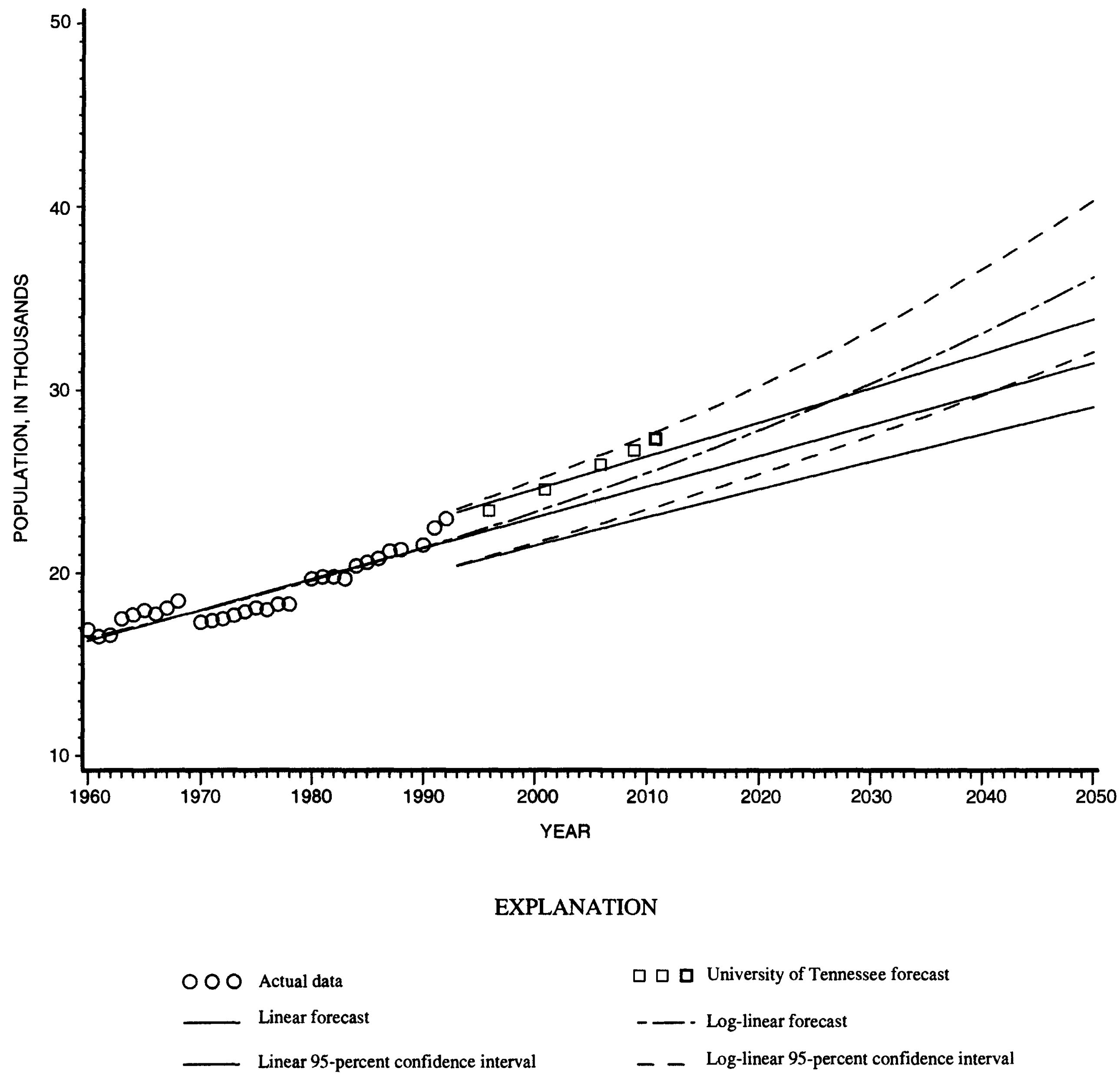


Figure 11. Actual and projected populations for Marshall County, Tennessee, for the linear and log-linear models, with confidence intervals.

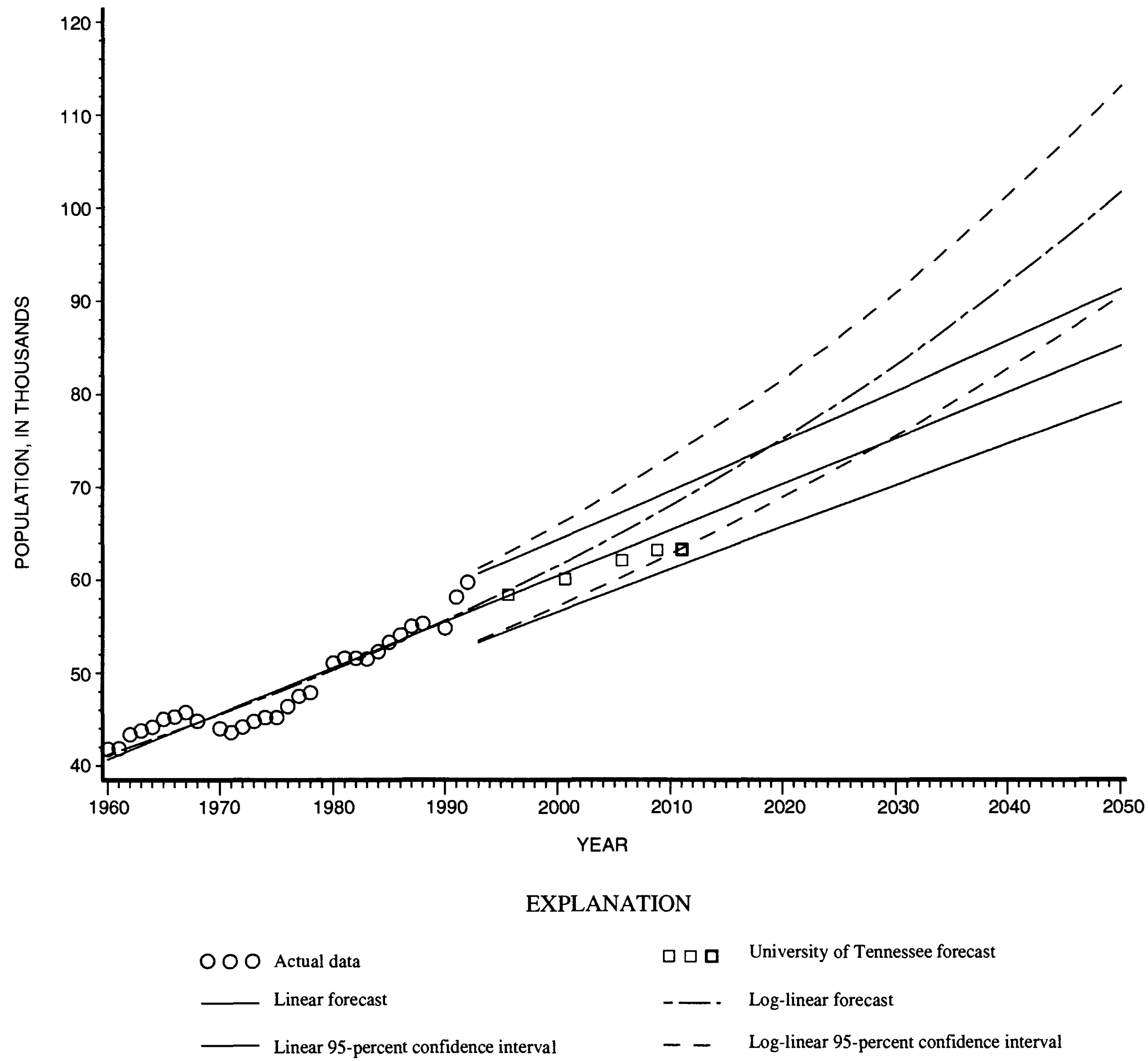


Figure 12. Actual and projected populations for Maury County, Tennessee, for the linear and log-linear models, with confidence intervals.

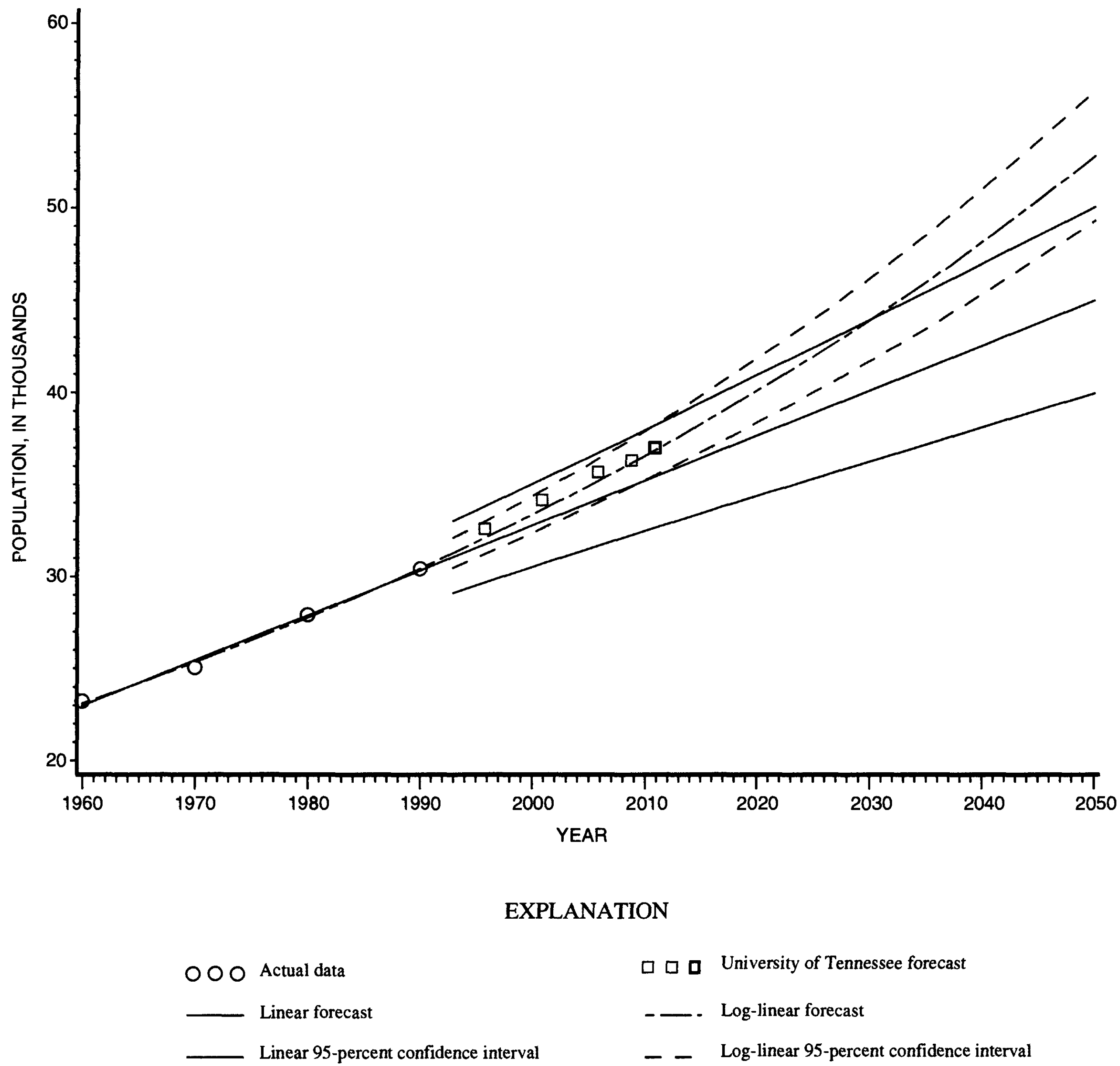


Figure 13. Actual and projected populations for Bedford County, Tennessee, using census-year data only for the linear and log-linear models, with confidence intervals.

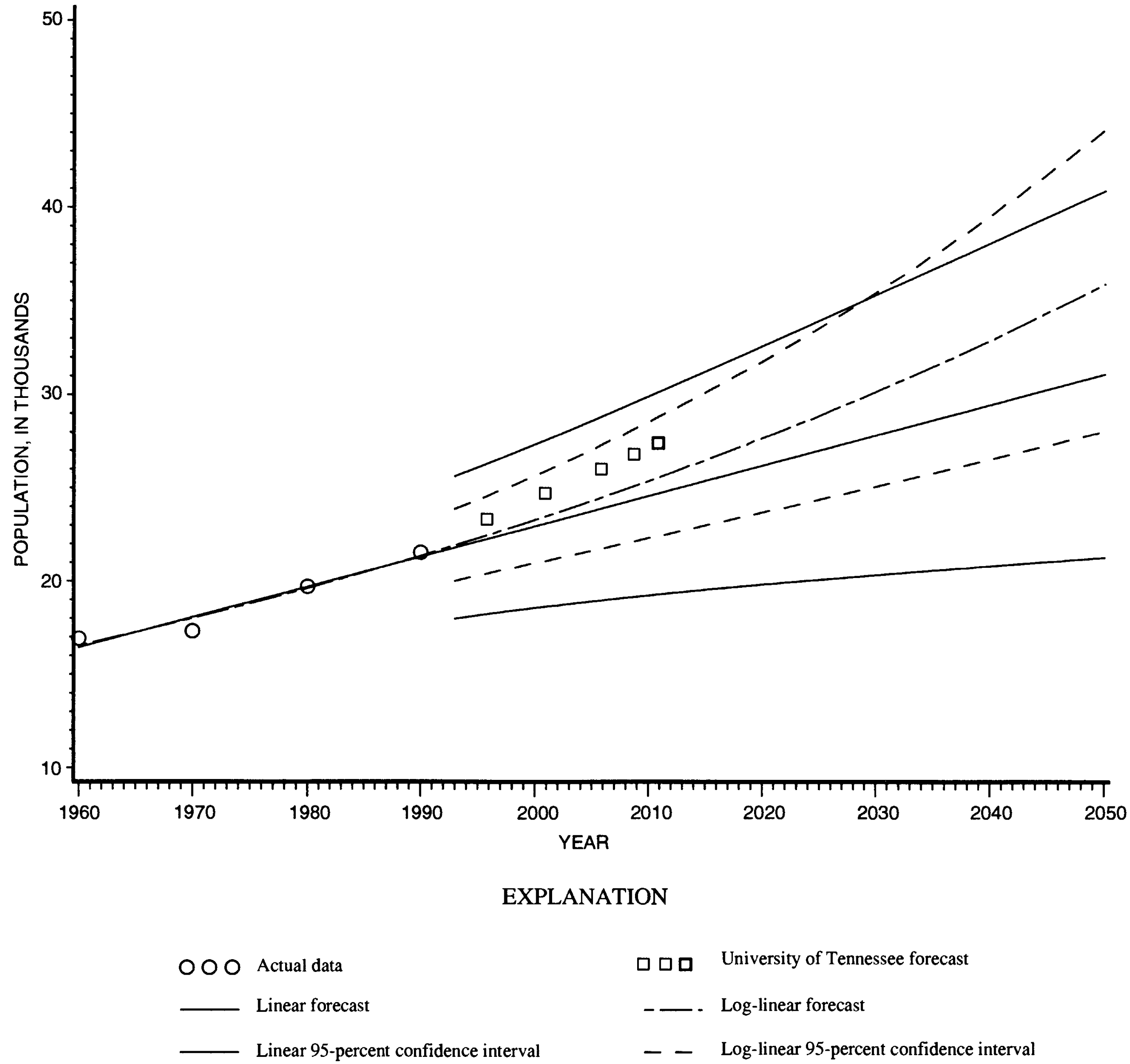


Figure 14. Actual and projected populations for Marshall County, Tennessee, using census-year data only for the linear and log-linear models, with confidence intervals.

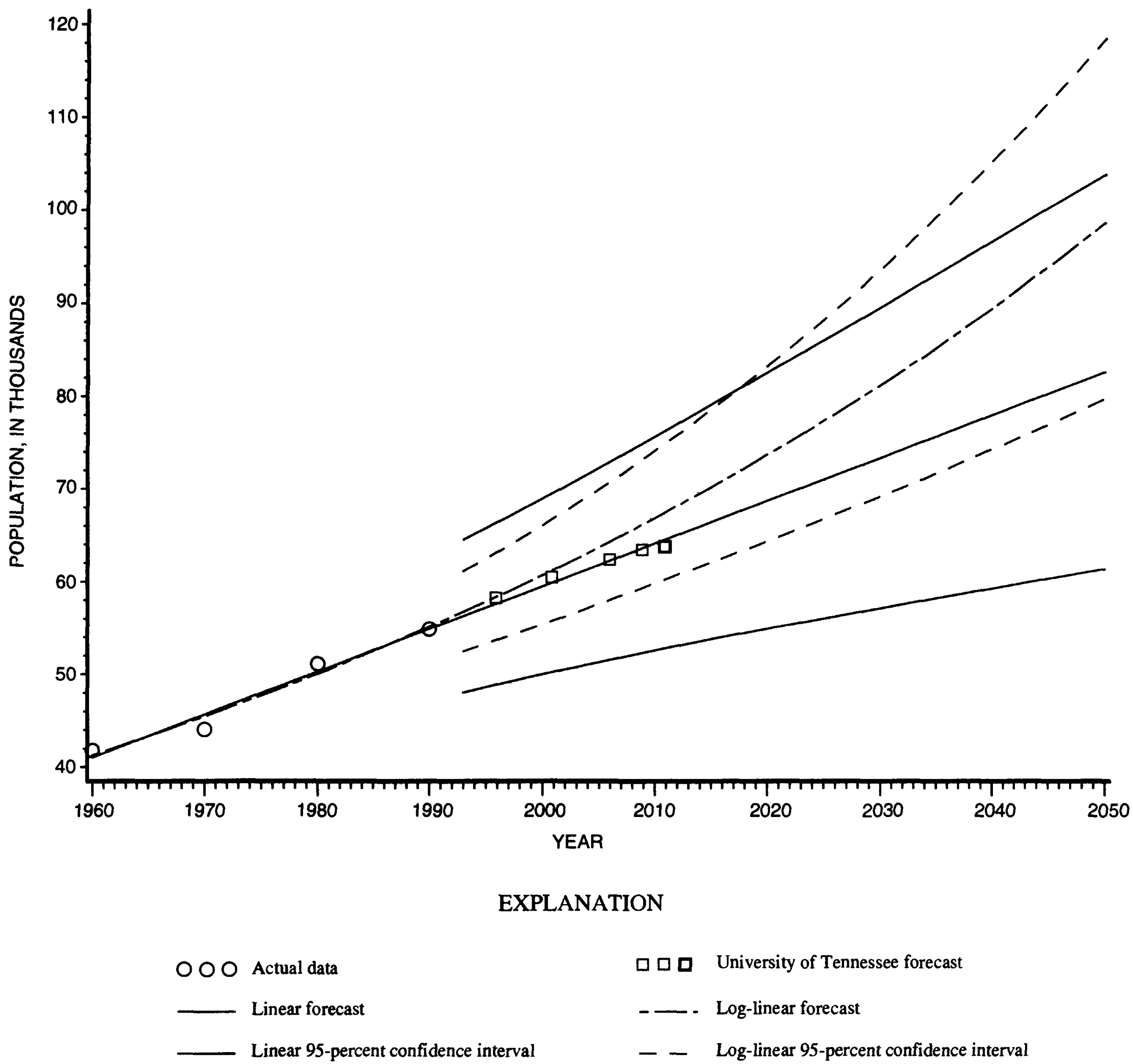


Figure 15. Actual and projected populations for Maury County, Tennessee, using census-year data only for the linear and log-linear models, with confidence intervals.

Table 17. Parameter estimates for the linear and log-linear models for Bedford, Marshall, and Maury Counties, Tennessee

[Standard errors in parentheses; β_0 and β_1, estimated parameters; R^2, coefficient of correlation]

	Linear model			Log-linear model		
County	β_0	β_1	R^2	β_0	β_1	R^2
Bedford	23.001[a]	0.235[a]	0.930	3.143[a]	0.00875[a]	0.937
	(0.224)	(0.012)		(0.008)	(0.00043)	
Marshall	16.299[a]	0.169[a]	0.861	2.800[a]	0.00874[a]	0.871
	(0.236)	(0.013)		(0.012)	(0.00064)	
Maury	40.679[a]	0.494[a]	0.893	3.718[a]	0.01003[a]	0.903
	(0.595)	(0.032)		(0.011)	(0.00062)	

[a] Significantly different from 0 at the 5-percent level of significance.

Table 18. Parameter estimates for the linear and log-linear models for Bedford, Marshall, and Maury Counties, Tennessee, using census-year data only (N = 4)

[Standard errors in parentheses; β_0 and β_1, estimated parameters; R^2, coefficient of correlation]

	Linear model			Log-linear model		
County	β_0	β_1	R^2	β_0	β_1	R^2
Bedford	22.981[a]	0.244[a]	0.993	3.140[a]	0.00918[a]	0.996
	(0.277)	(0.015)		(0.008)	(0.00043)	
Marshall	16.429[a]	0.163[a]	0.941	2.805[a]	0.00854[a]	0.945
	(0.539)	(0.029)		(0.027)	(0.00146)	
Maury	41.011[a]	0.461[a]	0.965	3.719[a]	0.00963[a]	0.966
	(1.167)	(0.062)		(0.024)	(0.00127)	

[a] Significantly different from 0 at the 5-percent level of significance.

model. The forecast from the University of Tennessee study (Economic and Community Staff, 1992) generally lies within the confidence interval for the log-linear model. The smaller confidence interval for the log-linear model projections is misleading because the method for estimating the confidence interval assumes σ_e^2 is known. A corrected confidence interval would include additional spread due to inherent uncertainty in the estimated value of the variance term. This correction would be particularly large in the case of regressions estimated with census-year data because of the few number of observations ($N = 4$) used to estimate the variance.

Tables 19 to 21 list population projections and confidence intervals for the three counties for the case where the model is linear and all available data are used in the regression. Tables 22 to 24 give projections and confidence intervals for the log-linear model, also using all available data. With the exception of the last 5 years for Bedford County, the confidence intervals for the linear models overlap with those for the log-linear models.

Table 19. Projected populations and 95-percent confidence intervals for Bedford County, Tennessee, 1993-2050, results from the linear model

Year	Projected population	95-percent confidence interval	
		Lower bound	Upper bound
1993	30,772	29,370	32,173
1994	31,007	29,598	32,416
1995	31,242	29,825	32,660
1996	31,478	30,052	32,904
1997	31,713	30,278	33,149
1998	31,949	30,504	33,394
1999	32,184	30,730	33,639
2000	32,420	30,955	33,884
2001	32,655	31,180	34,130
2002	32,891	31,405	34,377
2003	33,126	31,629	34,623
2004	33,362	31,853	34,870
2005	33,597	32,077	35,118
2006	33,833	32,300	35,365
2007	34,068	32,523	35,613
2008	34,304	32,746	35,862
2009	34,539	32,968	36,110
2010	34,775	33,190	36,359
2011	35,010	33,412	36,608
2012	35,246	33,634	36,857
2013	35,481	33,855	37,107
2014	35,717	34,076	37,357
2015	35,952	34,297	37,607
2016	36,188	34,518	37,857
2017	36,423	34,738	38,108
2018	36,659	34,958	38,359
2019	36,894	35,178	38,610
2020	37,129	35,398	38,861
2021	37,365	35,617	39,113
2022	37,600	35,836	39,365
2023	37,836	36,055	39,616
2024	38,071	36,274	39,869
2025	38,307	36,493	40,121
2026	38,542	36,711	40,373
2027	38,778	36,930	40,626
2028	39,013	37,148	40,879
2029	39,249	37,366	41,132
2030	39,484	37,584	41,385
2031	39,720	37,801	41,638
2032	39,955	38,019	41,892
2033	40,191	38,236	42,146
2034	40,426	38,453	42,399
2035	40,662	38,670	42,653
2036	40,897	38,887	42,907
2037	41,133	39,104	43,162
2038	41,368	39,320	43,416
2039	41,604	39,537	43,671
2040	41,839	39,753	43,925
2041	42,075	39,969	44,180
2042	42,310	40,185	44,435
2043	42,546	40,401	44,690
2044	42,781	40,617	44,945
2045	43,017	40,833	45,200
2046	43,252	41,049	45,455
2047	43,487	41,264	45,711
2048	43,723	41,480	45,966
2049	43,958	41,695	46,222
2050	44,194	41,910	46,477

Table 20. Projected populations and 95-percent confidence intervals for Marshall County, Tennessee, 1993-2050, results from the linear model

Year	Projected population	95-percent confidence interval	
		Lower bound	Upper bound
1993	21,866	20,393	23,339
1994	22,035	20,554	23,516
1995	22,203	20,714	23,693
1996	22,372	20,873	23,871
1997	22,541	21,032	24,049
1998	22,710	21,191	24,228
1999	22,878	21,350	24,407
2000	23,047	21,508	24,586
2001	23,216	21,665	24,766
2002	23,384	21,823	24,946
2003	23,553	21,980	25,127
2004	23,722	22,136	25,307
2005	23,890	22,292	25,489
2006	24,059	22,448	25,670
2007	24,228	22,604	25,852
2008	24,397	22,759	26,034
2009	24,565	22,914	26,216
2010	24,734	23,069	26,399
2011	24,903	23,223	26,582
2012	25,071	23,377	26,765
2013	25,240	23,531	26,949
2014	25,409	23,685	27,133
2015	25,578	23,838	27,317
2016	25,746	23,991	27,501
2017	25,915	24,144	27,686
2018	26,084	24,296	27,871
2019	26,252	24,449	28,056
2020	26,421	24,601	28,241
2021	26,590	24,753	28,427
2022	26,758	24,904	28,613
2023	26,927	25,056	28,799
2024	27,096	25,207	28,985
2025	27,265	25,358	29,171
2026	27,433	25,509	29,358
2027	27,602	25,659	29,545
2028	27,771	25,810	29,731
2029	27,939	25,960	29,919
2030	28,108	26,110	30,106
2031	28,277	26,260	30,293
2032	28,446	26,410	30,481
2033	28,614	26,560	30,669
2034	28,783	26,709	30,857
2035	28,952	26,858	31,045
2036	29,120	27,007	31,233
2037	29,289	27,156	31,422
2038	29,458	27,305	31,610
2039	29,626	27,454	31,799
2040	29,795	27,603	31,988
2041	29,964	27,751	32,177
2042	30,133	27,899	32,366
2043	30,301	28,048	32,555
2044	30,470	28,196	32,744
2045	30,639	28,344	32,934
2046	30,807	28,492	33,123
2047	30,976	28,639	33,313
2048	31,145	28,787	33,503
2049	31,313	28,935	33,692
2050	31,482	29,082	33,882

Table 21. Projected populations and 95-percent confidence intervals for Maury County, Tennessee, 1993-2050, results from the linear model

Year	Projected population	95-percent confidence interval Lower bound	95-percent confidence interval Upper bound
1993	56,985	53,266	60,704
1994	57,479	53,739	61,219
1995	57,973	54,211	61,735
1996	58,467	54,682	62,252
1997	58,961	55,152	62,771
1998	59,456	55,621	63,290
1999	59,950	56,090	63,810
2000	60,444	56,557	64,331
2001	60,938	57,023	64,853
2002	61,432	57,488	65,376
2003	61,926	57,953	65,900
2004	62,420	58,416	66,424
2005	62,914	58,879	66,950
2006	63,409	59,341	67,476
2007	63,903	59,802	68,004
2008	64,397	60,262	68,532
2009	64,891	60,722	69,060
2010	65,385	61,180	69,590
2011	65,879	61,638	70,120
2012	66,373	62,096	70,651
2013	66,868	62,552	71,183
2014	67,362	63,008	71,715
2015	67,856	63,463	72,248
2016	68,350	63,918	72,782
2017	68,844	64,372	73,316
2018	69,338	64,825	73,851
2019	69,832	65,278	74,387
2020	70,326	65,730	74,923
2021	70,821	66,182	75,460
2022	71,315	66,633	75,997
2023	71,809	67,083	76,535
2024	72,303	67,533	77,073
2025	72,797	67,983	77,612
2026	73,291	68,432	78,151
2027	73,785	68,880	78,691
2028	74,280	69,328	79,231
2029	74,774	69,776	79,772
2030	75,268	70,223	80,313
2031	75,762	70,669	80,854
2032	76,256	71,116	81,396
2033	76,750	71,562	81,939
2034	77,244	72,007	82,482
2035	77,738	72,452	83,025
2036	78,233	72,897	83,568
2037	78,727	73,341	84,112
2038	79,221	73,785	84,656
2039	79,715	74,229	85,201
2040	80,209	74,672	85,746
2041	80,703	75,116	86,291
2042	81,197	75,558	86,837
2043	81,692	76,001	87,382
2044	82,186	76,443	87,929
2045	82,680	76,885	88,475
2046	83,174	77,326	89,022
2047	83,668	77,767	89,569
2048	84,162	78,208	90,116
2049	84,656	78,649	90,664
2050	85,150	79,090	91,211

Table 22. Projected populations and 95-percent confidence intervals for Bedford County, Tennessee, 1993-2050, results from the log-linear model

Year	Projected population	95-percent confidence interval Lower bound	95-percent confidence interval Upper bound
1993	30,951	29,508	32,408
1994	31,223	29,760	32,701
1995	31,497	30,012	32,998
1996	31,774	30,267	33,297
1997	32,053	30,524	33,599
1998	32,335	30,782	33,905
1999	32,619	31,042	34,214
2000	32,906	31,304	34,526
2001	33,195	31,567	34,841
2002	33,487	31,833	35,160
2003	33,781	32,100	35,482
2004	34,078	32,369	35,807
2005	34,378	32,640	36,135
2006	34,680	32,913	36,467
2007	34,985	33,188	36,802
2008	35,293	33,465	37,141
2009	35,603	33,744	37,483
2010	35,916	34,025	37,829
2011	36,232	34,308	38,178
2012	36,550	34,593	38,531
2013	36,872	34,880	38,888
2014	37,196	35,169	39,248
2015	37,523	35,460	39,611
2016	37,853	35,753	39,979
2017	38,186	36,048	40,350
2018	38,521	36,345	40,725
2019	38,860	36,645	41,103
2020	39,202	36,947	41,486
2021	39,547	37,251	41,872
2022	39,894	37,557	42,262
2023	40,245	37,865	42,656
2024	40,599	38,176	43,054
2025	40,956	38,489	43,456
2026	41,316	38,805	43,862
2027	41,680	39,122	44,272
2028	42,046	39,442	44,686
2029	42,416	39,765	45,104
2030	42,789	40,089	45,527
2031	43,165	40,417	45,953
2032	43,545	40,746	46,384
2033	43,928	41,078	46,819
2034	44,315	41,413	47,258
2035	44,704	41,750	47,702
2036	45,098	42,090	48,150
2037	45,494	42,432	48,603
2038	45,894	42,777	49,060
2039	46,298	43,124	49,521
2040	46,706	43,474	49,987
2041	47,116	43,827	50,457
2042	47,531	44,182	50,933
2043	47,949	44,540	51,412
2044	48,371	44,901	51,897
2045	48,796	45,264	52,386
2046	49,226	45,630	52,880
2047	49,659	45,999	53,379
2048	50,096	46,371	53,882
2049	50,537	46,746	54,391
2050	50,981	47,123	54,905

Table 23. Projected populations and 95-percent confidence intervals for Marshall County, Tennessee, 1993-2050, results from the log-linear model

Year	Projected population	95-percent confidence interval Lower bound	95-percent confidence interval Upper bound
1993	21,963	20,438	23,512
1994	22,156	20,609	23,728
1995	22,351	20,781	23,946
1996	22,547	20,954	24,166
1997	22,745	21,128	24,389
1998	22,945	21,303	24,614
1999	23,147	21,479	24,842
2000	23,350	21,656	25,072
2001	23,555	21,834	25,305
2002	23,762	22,013	25,540
2003	23,971	22,194	25,778
2004	24,182	22,375	26,019
2005	24,394	22,557	26,262
2006	24,609	22,741	26,508
2007	24,825	22,926	26,757
2008	25,043	23,111	27,008
2009	25,263	23,298	27,262
2010	25,485	23,487	27,519
2011	25,709	23,676	27,779
2012	25,935	23,866	28,041
2013	26,163	24,058	28,307
2014	26,393	24,251	28,575
2015	26,626	24,445	28,846
2016	26,860	24,640	29,120
2017	27,096	24,837	29,397
2018	27,334	25,035	29,677
2019	27,574	25,234	29,960
2020	27,817	25,434	30,245
2021	28,062	25,636	30,534
2022	28,308	25,839	30,826
2023	28,557	26,043	31,121
2024	28,809	26,249	31,419
2025	29,062	26,456	31,721
2026	29,318	26,664	32,025
2027	29,576	26,874	32,333
2028	29,836	27,085	32,643
2029	30,098	27,298	32,957
2030	30,363	27,512	33,275
2031	30,630	27,727	33,595
2032	30,900	27,944	33,919
2033	31,172	28,162	34,247
2034	31,446	28,382	34,577
2035	31,723	28,603	34,912
2036	32,002	28,826	35,249
2037	32,284	29,050	35,590
2038	32,568	29,276	35,935
2039	32,854	29,503	36,283
2040	33,144	29,732	36,635
2041	33,435	29,962	36,990
2042	33,730	30,194	37,349
2043	34,027	30,428	37,711
2044	34,326	30,663	38,078
2045	34,628	30,900	38,448
2046	34,933	31,138	38,822
2047	35,241	31,378	39,199
2048	35,551	31,620	39,581
2049	35,864	31,864	39,966
2050	36,180	32,109	40,356

Table 24. Projected populations and 95-percent confidence intervals for Maury County, Tennessee, 1993-2050, results from the log-linear model

Year	Projected population	95-percent confidence interval Lower bound	95-percent confidence interval Upper bound
1993	57,345	53,466	61,283
1994	57,924	53,984	61,923
1995	58,508	54,505	62,572
1996	59,098	55,030	63,228
1997	59,694	55,559	63,892
1998	60,295	56,092	64,564
1999	60,904	56,629	65,245
2000	61,518	57,170	65,934
2001	62,138	57,716	66,631
2002	62,765	58,265	67,337
2003	63,398	58,819	68,051
2004	64,037	59,377	68,773
2005	64,683	59,939	69,505
2006	65,336	60,506	70,245
2007	65,995	61,077	70,995
2008	66,661	61,653	71,753
2009	67,333	62,233	72,520
2010	68,012	62,817	73,297
2011	68,699	63,406	74,082
2012	69,392	64,000	74,877
2013	70,092	64,598	75,682
2014	70,799	65,201	76,496
2015	71,514	65,809	77,320
2016	72,235	66,422	78,153
2017	72,964	67,040	78,996
2018	73,701	67,662	79,850
2019	74,444	68,290	80,713
2020	75,196	68,922	81,586
2021	75,955	69,560	82,470
2022	76,721	70,203	83,364
2023	77,496	70,850	84,269
2024	78,278	71,504	85,184
2025	79,068	72,162	86,109
2026	79,866	72,826	87,046
2027	80,673	73,495	87,993
2028	81,487	74,170	88,952
2029	82,310	74,850	89,921
2030	83,141	75,536	90,902
2031	83,980	76,227	91,894
2032	84,828	76,924	92,898
2033	85,685	77,627	93,913
2034	86,550	78,335	94,940
2035	87,424	79,050	95,979
2036	88,307	79,770	97,030
2037	89,199	80,496	98,093
2038	90,100	81,228	99,168
2039	91,010	81,967	100,255
2040	91,929	82,711	101,355
2041	92,857	83,462	102,467
2042	93,795	84,219	103,593
2043	94,743	84,982	104,731
2044	95,700	85,752	105,882
2045	96,667	86,528	107,046
2046	97,643	87,310	108,223
2047	98,630	88,100	109,414
2048	99,626	88,895	110,619
2049	100,633	89,698	111,837
2050	101,650	90,507	113,069

Summary and Conclusions

Three types of models are used to estimate the relation between population and time for Bedford, Marshall, and Maury Counties. The Box-Cox model yields projections with a large negative curvature—so large that out-of-sample projections quickly go to infinity. The large curvature is partially a consequence of using ancillary data to estimate population between census years. More reasonable and similar projections are obtained for the linear and log-linear models. Generally, the confidence intervals around the forecasts show some degree of overlap between these two models. Given the curvature indicated by the Box-Cox model, which includes the linear and log-linear models as special cases, the log-linear model appears to best forecast populations of Bedford, Marshall, and Maury Counties into the future.

SELECTED REFERENCES

Alexander, F.M., Keck, L.A., Conn, L.G., and Wentz, S.J., 1984, Drought-related impacts on municipal and major self-supplied industrial water withdrawals in Tennessee—Part B: U.S. Geological Survey Water-Resources Investigations Report 84-4074, 398 p.

Amemiya, Takeshi, 1985, Advanced econometrics: Cambridge, Harvard University Press, 521 p.

Bingham, R.H., 1985, Low flows and flow duration of Tennessee streams through 1981: U.S. Geological Survey Water-Resources Investigations Report 84-4347, 325 p.

Boland, J.J., Dziegielewski, B., Baumann, D.D., Opitz, E.M., 1984, Influence of price and rate structures on municipal and industrial water use: Fort Belvoir, Virginia, U.S. Army Corps of Engineers Institute for Water Resources, Contract Report 84-Industrial Water Use: Fort Belvoir, Virginia, U.S. Army Corps of Engineers Institute for Water Resources, Contract Report 84-c-2, 187 p.

Burchett, C.R., 1977, Water resources of the upper Duck River basin, central Tennessee: Tennessee Division of Water Resources, Tennessee Water Resources Series no. 12, 103 p.

Davidson, Russell, and MacKinnon, J.G., 1993, Estimation and inference in econometrics: New York, Oxford University Press, 874 p.

Economic and Community Staff, South Central Tennessee Development District, comp., 1992, Population estimates for the State of Tennessee, *in* 1994 data profile: Knoxville, Tennessee, University of Tennessee, 25 p.

Howe, C.W., Linaweaver, F.P., Jr., 1967, The impact of price on residential water demand and its relation to system design and price structure: Water Resources Research 3 (1st Quarter), p. 12-32.

Hutson, S.S., 1993, Water availability, use, and estimated future water demand in the upper Duck River basin, middle Tennessee: U.S. Geological Survey Water-Resources Investigations Report 92-4179, 39 p.

Kindler, J., and Russell, C.S., 1984, Modelling water demands: Orlando, Florida, Academic Press (Harcourt, Brace, Jovanovich, Inc), 248 p.

Miller, R.A., 1974, The geologic history of Tennessee: Tennessee Division of Geology Bulletin 74, 63 p.

Office of Management and Budget, 1987, Standard industrial classification manual 1987: 705 p.

Outlaw, G.S., and Weaver, J.D., 1996, Flow duration and low flows of Tennessee streams through 1992: U.S. Geological Survey Water-Resources Investigations Report 95-4293, 245 p.

Tennessee Department of Employment Security, 1990, Annual averages Tennessee labor force estimates 1985-1989: Nashville, Tennessee, Labor Market Information Unit, Research and Statistics Division, Tennessee Department of Employment Security, 87 p.

———1991, Annual averages labor force and nonagricultural employment estimates 1986-1990: Nashville, Tennessee, Labor Market Information Research and Statistics Division, Tennessee Department of Employment Security, 87 p.

U.S. Army Corps of Engineers, 1988, IWR-MAIN water use forecasting system, version 5.1: Fort Belvoir, Virginia, Water Resources Support Center, Institute for Water Resources, 324 p.

U.S. Bureau of the Census, 1967, Estimates of the population of states, July 1, 1966, *in* Current Population Reports, Population Estimates: Bureau of the Census, Series P-25, No. 380.

———1968, Estimates of the population of states, July 1, 1966 and July 1, 1967, *in* Current Population Reports, Population Estimates: Bureau of the Census, Series P-25, No. 414.

———1969, Estimates of the population of states, July 1, 1967 and July 1, 1968, *in* Current Population Reports, Population Estimates: Bureau of the Census, Series P-25, No. 436 and 437.

———1973, Estimates of the population of Tennessee counties, July 1, 1971 and July 1, 1972, *in* Current Population Reports, Population Estimates: Bureau of the Census, Series P-26, No. 47.

———1974, Estimates of the population of Tennessee counties and metropolitan areas, July 1, 1972 and 1973, *in* Current Population Reports, Population Estimates: Bureau of the Census, Series P-26, No. 83.

———1975, Estimates of the population of Tennessee counties and metropolitan areas, July 1, 1973 and 1974, *in*

Current Population Reports, Population Estimates: Bureau of the Census, Series P-26, No. 133.

———1976, Estimates of the population of Tennessee counties and metropolitan areas, July 1, 1974 and 1975, *in* Current Population Reports, Population Estimates: Bureau of the Census, Series P-26, No. 75-42.

———1978, Estimates of the population of Tennessee counties and metropolitan areas, July 1, 1976 (revised) and 1977 (provisional), *in* Current Population Reports, Population Estimates: Bureau of the Census, Series P-26, No. 77-42.

———1979, Estimates of the population of Tennessee counties and metropolitan areas, July 1, 1977 (revised) and 1978 (provisional), *in* Current Population Reports, Population Estimates: Bureau of the Census, Series P-26, No. 78-42.

———1988a, Estimates of the Population of Tennessee counties and metropolitan areas, July 1, 1981, to 1985, *in* Current Population Reports, Local Population Estimates: Bureau of the Census, Series P-26, No. 85-TN-C.

———1988b, South, 1986 Population and 1985 per capita income estimates for counties and incorporated places, *in* Current Population Reports, Local Population Estimates: Bureau of the Census, Series P-26, No. 86-S-SC.

———1990, South, 1988 Population and 1987 per capita income estimates for counties and incorporated places, *in* Current Population Reports, Local Population Estimates: Bureau of the Census, Series P-26, No. 88-S-SC.

———1994, Population estimates for counties and metropolitan areas, July 1, 1991, *in* Current Population Reports, Population Estimates and Projections: Bureau of the Census, P25-1108.

U.S. Department of Commerce, 1988, Climatological data annual summary, Tennessee 1988: Asheville, North Carolina, National Oceanic and Atmospheric Administration, 17 p.

———1989, Climatological data annual summary, Tennessee 1989: Asheville, North Carolina, National Oceanic and Atmospheric Administration, 17 p.

———1992a, 1990 Census of housing, general housing characteristics, Tennessee: U.S. Government Printing Office, 270 p.

———1992b, 1990 Census of population, general population characteristics, Tennessee: U.S. Government Printing Office, 353 p.

———1992c, 1990 Census of population and housing, summary social, economic, and housing characteristics, Tennessee: U.S. Government Printing Office, 203 p.

Vickers, B.B., ed., 1989, Tennessee statistical abstract 1989, Center for Business and Economic Research: Knoxville, Tennessee, University of Tennessee, 755 p.

White, J.L., ed., 1994, 1994 Directory of Tennessee manufacturers: geographic section, Nashville, Tennessee: M. Lee Smith Publishers & Printers, section B, 297 p.

Appendix A. Derivation of the Confidence Intervals for the Log-Linear Model

This appendix derives the relations necessary to determine minimal confidence limits for the log-linear model. The derivation assumes that the standard error of the residuals is known.

We seek to minimize the difference $U_t - L_t$ subject to the constraint

$$Prob\,(L_t \le P_t \le U_t) = 1 - \alpha, \tag{11}$$

where L_t is the lower bound on the confidence interval; U_t is the upper bound; and α is the significance level (for a 95-percent confidence interval, α equals 0.05). The probability expressed in equation 11 is conditioned on the information contained in the sample used to estimate the log-linear model.

It is convenient at this point to adopt some matrix notation. Let x_t represent the row vector $[1\ T_t]$; let X_s represent the N by 2 matrix consisting of ones in the first column and the sample values of T in the second column; let β be the 2 by 1 vector $[\beta_0\ \beta_1]'$, with $\hat{\beta}$ being the β vector evaluated at the sample estimates; and let e_s correspond to the N vector of residuals for the sample values of P. The population P_t appearing in equation 11 can then be written as

$$\begin{aligned} P_t &= \exp\,(x_t\beta + e_t) \\ &= \exp\left(x_t\hat{\beta} + e_t - x_t\,(\hat{\beta} - \beta)\right) \\ &= \exp\left(x_t\hat{\beta} + e_t - x_t\,(X'_s X_s)^{-1} X'_s\, e_s\right). \end{aligned} \tag{12}$$

The term $e_t - x_t(X'_sX_s)^{-1}X'_se_s$ represents the forecast error and is given by the sum of independent, normally distributed random variables with a mean of 0. As such, it is itself normally distributed with mean of 0 and variance equal to the variance of the forecast error $\hat{\sigma}_t^2$ (see equation 7). Let u_t represent this forecast error. By substituting equation 12 into equation 11, taking natural logarithms, and dividing by the square root of $\hat{\sigma}_t^2$ we obtain

$$Prob\left(\frac{\ln\,(L_t) - x_t\hat{\beta}}{\sqrt{\hat{\sigma}_t^2}} \le \frac{u_t}{\sqrt{\hat{\sigma}_t^2}} \le \frac{\ln\,(U_t) - x_t\hat{\beta}}{\sqrt{\hat{\sigma}_t^2}}\right) = 1 - \alpha. \tag{13}$$

If σ_e^2 is known, then the term $u_t/\sqrt{\hat{\sigma}_t^2}$ is normally distributed with a mean of 0 and variance of 1. Define $Y_t(L_t) = \left(\ln(L_t) - x_t\hat{\beta}\right)/\sqrt{\hat{\sigma}_t^2}$ and $Z_t(U_t) = \left(\ln(U_t) - x_t\hat{\beta}\right)/\sqrt{\hat{\sigma}_t^2}$. The minimization problem can be stated as

$$\min_{w.r.t.\,L_t, U_t} \quad \mathcal{L} = U_t - L_t - \lambda\left(\frac{1}{\sqrt{2\pi}}\int_{Y_t(L_t)}^{Z_t(U_t)} e^{-\varepsilon^2/2}\,d\varepsilon - (1 - \alpha)\right). \tag{14}$$

The first-order conditions for minimization (in addition to equation 11) are

$$1 = \frac{\lambda \exp\left(-Z_t^2 (U_t)/2\right)}{U_t \sqrt{2\pi \hat{\sigma}_t^2}} \tag{15}$$

and

$$1 = \frac{\lambda \exp\left(-Y_t^2 (L_t)/2\right)}{L_t \sqrt{2\pi \hat{\sigma}_t^2}}. \tag{16}$$

These conditions define a unique solution for L_t and U_t. Since the second-order conditions for a minimum are satisfied, the solution obtained is a global minimum.

Dividing equation 15 by equation 16 and taking natural logarithms, we obtain

$$\frac{Y_t^2 (L_t) - Z_t^2 (U_t)}{2} = \ln (U_t) - \ln (L_t) \tag{17}$$

$$= (Z_t (U_t) - Y_t (L_t)) \sqrt{\hat{\sigma}_t^2}.$$

Upon simplifying equation 17, we obtain

$$Z_t + Y_t = -2\sqrt{\hat{\sigma}_t^2}. \tag{18}$$

Thus, the solution to the minimization problem is

$$L_t = \exp\left(x_t \hat{\beta} - \left(Z_t + 2\sqrt{\hat{\sigma}_t^2}\right)\sqrt{\hat{\sigma}_t^2}\right)$$

and (19)

$$U_t = \exp\left(x_t \hat{\beta} + Z_t \sqrt{\hat{\sigma}_t^2}\right),$$

where Z_t solves the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-Z_t - 2\sqrt{\hat{\sigma}_t^2}}^{Z_t} e^{-\varepsilon^2/2} d\varepsilon = 1 - \alpha. \tag{20}$$

A solution for Z_t can be obtained using the Newton method. Note that if α is less than 0.5, then Z_t must be positive (if Z_t is less than zero, then the probability defined by the left-hand side of equation 20 would be less than 0.5). Because the first derivative of the left-hand side of equation 20 is positive and the second derivative is negative for all positive values of Z_t, the Newton method converges quickly and uniformly on the solution if the starting value for Z_t is 0.

Appendix B. SAS Computer Code Used to Generate the Results

The following is the SAS computer code used to generate the tables and figures appearing in the text.

```
/* This program estimates the lambda coefficient for the simple one-sided Box-Cox transformation of the
   dependent variable. The lambda coefficient and standard error correspond to the maximum likelihood esti-
   mates (the beta coefficients on the explanatory variables are wrong (as are the standard errors)). See my
   notes for transforming these values into the correct values. The method of estimation is maximum likeli-
   hood, using the concentrated log likelihood function and rescaling the dependent variable by its geometric
   mean. */

%let tol = 1E-10 ;
%let cutoff = 2.25 ;

%let pop1 = bedpop ;
%let pop2 = marpop ;
%let pop3 = maupop ;

%let tpop1 = tbedpop ;
%let tpop2 = tmarpop ;
%let tpop3 = tmaupop ;

%let apop1 = abedpop ;
%let apop2 = amarpop ;
%let apop3 = amaupop ;

%let lpop1 = lbedpop ;
%let lpop2 = lmarpop ;
%let lpop3 = lmaupop ;

%let ppop1 = pbedpop ;
%let ppop2 = pmarpop ;
%let ppop3 = pmaupop ;

%let ctitle11 = Results for Bedford County ;
%let ctitle21 = Results for Marshall County ;
%let ctitle31 = Results for Maury County ;

%let ctitle12 = Results for Bedford County (census years only);
%let ctitle22 = Results for Marshall County (census years only);
%let ctitle32 = Results for Maury County (census years only);

%let headl11 = 'Bedford' / 'County, 1993-2050' / / '[Results from the linear model]' ;
%let headl21 = 'Marshall' / 'County, 1993-2050' / / '[Results from the linear model]' ;
%let headl31 = 'Maury' / 'County, 1993-2050' / / '[Results from the linear model]' ;

%let headl12 = 'Bedford' / 'County, 1993-2050' / / '[Results from the linear model using census year data]' ;
%let headl22 = 'Marshall' / 'County, 1993-2050' / / '[Results from the linear model using census year data]' ;
%let headl32 = 'Maury' / 'County, 1993-2050' / / '[Results from the linear model using census year data]' ;
```

```
%let headll11 = 'Bedford' / 'County, 1993-2050' / / '[Results from the log-linear model]' ;
%let headll21 = 'Marshall' / 'County, 1993-2050' / / '[Results from the log-linear model]' ;
%let headll31 = 'Maury' / 'County, 1993-2050' / / '[Results from the log-linear model]' ;

%let headll12 = 'Bedford' / 'County, 1993-2050' / / '[Results from the log-linear model using census year data]' ;
%let headll22 = 'Marshall' / 'County, 1993-2050' / / '[Results from the log-linear model using census year data]' ;
%let headll32 = 'Maury' / 'County, 1993-2050' / / '[Results from the log-linear model using census year data]' ;

libname dir '';
filename indata '~/tenn/popdat.dat';
filename indata2 '~/tenn/proj.dat';

/* Read the data */

data one ;
infile indata ;
input year time bedpop marpop maupop ;
bedpop = bedpop/1000 ;
marpop = marpop/1000 ;
maupop = maupop/1000 ;
time = time - 60 ;
dep = 0 ;

/* if year ^= 1960 and year ^= 1970 and year ^= 1980 and year ^= 1990 then do ;*/

if year = 1969 or year = 1979 or year = 1989 then do ;
 bedpop = . ;
 marpop = . ;
 maupop = . ;
 dep = . ;
end ;
lbedpop = log(bedpop) ;
lmarpop = log(marpop) ;
lmaupop = log(maupop) ;
time2 = time*time ;

data two ;
do i = 93 to 150 ;
 time = i - 60 ;
 year = 1900 + time + 60 ;
 output ;
end ;

data three ;
infile indata2 ;
input year time pbedpop pmarpop pmaupop ;
pbedpop = pbedpop/1000 ;
pmarpop = pmarpop/1000 ;
pmaupop = pmaupop/1000 ;
```

```
/* Calculate the ln of the geometric mean */

proc means data = one ;
var lbedpop lmarpop lmaupop ;
output out = meandat mean = mlbedp mlmarp mlmaup ;

data one;
if _n_ = 1 then set meandat ;
set one two ;

/* Rescale the dependent variable */

tbedpop = bedpop*exp(-mlbedp) ;
tmarpop = marpop*exp(-mlmarp) ;
tmaupop = maupop*exp(-mlmaup) ;

/* Calculate the geometric mean */

abedpop = exp(mlbedp) ;
amarpop = exp(mlmarp) ;
amaupop = exp(mlmaup) ;

data one1 ; set one ;

/* Set up data for the Census Year regressions */

data one2 ; set one ;
if year <= 1992 and year ^= 1960 and year ^= 1970
  and year ^= 1980 and year ^= 1990 then do ;
bedpop = . ;
marpop = . ;
maupop = . ;
tbedpop = . ;
tmarpop = . ;
tmaupop = . ;
lbedpop = . ;
lmarpop = . ;
lmaupop = . ;
pbedpop = . ;
pmarpop = . ;
pmaupop = . ;
end ;

%macro estimate ;

%do iter = 1 %to 2 ;

%do cnty = 1 %to 3 ;

option dquote ;
```

```
filename resout1 "~/tenn/linres&cnty&iter..dat" ;
filename resout2 "~/tenn/llnres&cnty&iter..dat" ;

/*  Set up dataset of actual data (for display purposes) */

title1 &&&ctitle&cnty&iter ;
title2 ;
title3 ;

data outdat0 ; set one&iter ;
  pophat = &&&pop&cnty ;
  if pophat ^= . ;
  disp = 0 ;
keep pophat year disp ;

data outdat5 ; set three ;
  pophat = &&&ppop&cnty ;
  disp = 5 ;
keep pophat year disp ;

/*  Here's a linear model  */

proc reg data = one&iter ;
model &&&pop&cnty = time ;
output out = outdat1 p = pophat l95 = cilo u95 = cihi ;

data outdat1 ; set outdat1 ;
disp = 1 ;
if year <= 1992 then do ;
  cilo = . ;
  cihi = . ;
end ;
keep pophat year disp cilo cihi ;

data results ; set outdat1 ;
  if year > 1992 ;
  pophat = pophat*1000 ;
  cilo = cilo*1000 ;
  cihi = cihi*1000 ;
  file resout1 ;
  if year = 1993 then put
   'Table . Projected populations and 95-percent confidence intervals for '
&&&headl&cnty&iter / / / /
   '    95-percent          95-percent' /
   '    confidence          confidence' /
   '    interval            interval' /
   '    Projected           Projected' / /
   'Year   population   Lower bound  Upper bound  Year  population  Lower bound
Upper bound' / / / @ ;
```

```
if year <= 2021 then
  put year '' pophat comma7.0 '  ' cilo comma7.0 '' cihi comma7.0  '' ;
else
  put year '' pophat comma7.0 '  ' cilo comma7.0 '' cihi comma7.0 ;

/* Here's a log-linear model */

proc reg data = one&iter ;
model &&&lpop&cnty = time ;
output out = outdat2 p = pophat stdi = sdfore stdr = sdres ;

/* These statements calculate the upper and lower bounds of the 95% confidence interval for the log linear
model, assuming the standard deviation of the forecast is known. This allows us to simplify the estimation and
use the normal distribution.  The assumption also simplifies the estimation of the un-logged forecast.  */

data outdat2 ; set outdat2 ;
disp = 2 ;
dif = 1 ;
if year > 1992 then do ;
  do until(dif < &tol) ;
    if x = . then x = 0 ;
    pdif = .95 - (probnorm(x + 2*sdfore) - probnorm(-x)) ;
    dfx = ((2*3.141592654)**(-.5))*(exp(-((x+2*sdfore)**2)/2) + exp(-(x**2)/2)) ;
    xn = x + (pdif/dfx) ;
    dif = abs(xn - x) ;
    x = xn ;
  end ;
  cilo = exp(pophat - sdfore*(x + 2*sdfore)) ;
  cihi = exp(pophat + sdfore*x) ;
  pophat = exp(pophat + ((sdfore**2)/2)) ;
end ;
else do ;
  cilo = . ;
  cihi = . ;
  if sdres = . then sdres = sdfore ;
  pophat = exp(pophat + ((sdres**2)/2)) ;
end ;
keep pophat year disp cilo cihi ;

data results ; set outdat2 ;
  if year > 1992 ;
  pophat = pophat*1000 ;
  cilo = cilo*1000 ;
  cihi = cihi*1000 ;
  file resout2 ;
  if year = 1993 then put
   'Table   .  Projected  populations  and  95-percent  confidence  intervals  for  '
&&&headll&cnty&iter / / / /
   '     95-percent          95-percent' /
```

```
   '    confidence          confidence' /
   '    interval             interval' /
   '    Projected            Projected' / /
   'Year     population  Lower bound  Upper bound    Year   population    Lower bound
Upper bound' / / / @ ;

if year <= 2021 then
  put year '' pophat comma7.0 '  ' cilo comma7.0 '' cihi comma7.0  '' ;
else
  put year '' pophat comma7.0 '  ' cilo comma7.0 '' cihi comma7.0 ;

/* These are the ML estimates for lambda  */

proc model data = one&iter ;
dep = (b0 + b1*time) - (((&&&tpop&cnty**lambda)-1)/lambda) ;
lmbda = lambda ;
beta0 = (&&&apop&cnty**lmbda)*(b0 + (1/lmbda)) - (1/lmbda) ;
beta1 = b1*(&&&apop&cnty**lmbda) ;
e = (&&&apop&cnty**lmbda)*(-((b0 + b1*time) - (((&&&tpop&cnty**lmbda)-1)/lmbda))) ;
pophat = &&&apop&cnty*((lmbda*(b0 + b1*time) + 1)**(1/lmbda)) ;
blambda = (((&&&pop&cnty**lmbda)*(lmbda*&&&lpop&cnty - 1)) + 1)/(lmbda**2) ;
lnpop = &&&lpop&cnty ;
fit dep start = (lambda 2 b0 0 b1 0) / maxiter = 200
  out = outdat3 ;
outvars pophat year time lmbda beta0 beta1 e blambda lnpop &&&apop&cnty ;
run ;

data outdat3 ; set outdat3 ;
disp = 3 ;
if pophat > &cutoff*&&&apop&cnty then pophat = . ;
drop _type_ &&&apop&cnty ;

data correct ; set outdat3 ;
  if e ^= . ;

proc means data = correct ;
  var e ;
  output out = sigma n = nobs uss = sse ;

data temp ;
  if _n_ = 1 then set sigma ;
  set correct ;
  if _n_ = 1 ;
  sigma = (sse/nobs)**.5 ;
  keep lmbda beta0 beta1 sigma ;

proc print data = temp ;

data correct ;
  if _n_ = 1 then set sigma ;
```

```
 set correct ;
 sigma = (sse/nobs)**.5 ;
 lr = e/sigma ;
 lr2 = 1 ;
 r1 = 1/sigma ;
 r12 = 0 ;
 r2 = time/sigma ;
 r22 = 0 ;
 r3 = -blambda/sigma ;
 r32 = lnpop ;
 r4 = e/(sigma**2) ;
 r42 = -1/sigma ;

data correct2 ; set correct ;
 lr = lr2 ;
 r1 = r12 ;
 r2 = r22 ;
 r3 = r32 ;
 r4 = r42 ;

data correct ; set correct correct2 ;
keep lr r1 r2 r3 r4 ;

proc model data = correct ;
 lr = r1*b0 + r2*b1 + r3*lambda + r4*sigma ;
 fit lr start = (b0 0 b1 0 lambda 0 sigma 0) ;
run ;

/* Try an instrumental variables estimator */

proc model data = one&iter ;

endogenous dep &&&pop&cnty ;
exogenous time ;
instruments _exog_ time2 ;

dep = (b0 + b1*time) - (((&&&tpop&cnty**lambda)-1)/lambda) ;
lmbda = lambda ;
beta0 = (&&&apop&cnty**lmbda)*(b0 + (1/lmbda)) - (1/lmbda) ;
beta1 = b1*(&&&apop&cnty**lmbda) ;
e = (&&&apop&cnty**lmbda)*(-((b0 + b1*time) - (((&&&pop&cnty**lmbda)-1)/lmbda))) ;
pophat = &&&apop&cnty*((lmbda*(b0 + b1*time) + 1)**(1/lmbda)) ;
fit dep start = (lambda 2 b0 0 b1 0) / n2sls maxiter = 200
 out = outdat4 ;
outvars pophat year lmbda beta0 beta1 e &&&apop&cnty ;
run ;

data outdat4 ; set outdat4 ;
disp = 4 ;
if pophat > &cutoff*&&&apop&cnty then pophat = . ;
```

```
drop _type_ &&&apop&cnty ;

data temp ; set outdat4 ;
 if e ^= . ;
keep lmbda beta0 beta1 e ;

proc means data = temp ;
 var e ;
 output out = sigma n = nobs uss = sse ;

data temp ;
 if _n_ = 1 then set sigma ;
 set temp ;
 if _n_ = 1 ;
 sigma = (sse/nobs)**.5 ;
keep lmbda beta0 beta1 sigma ;

proc print data = temp ;

data outdat ; set outdat0 outdat1 outdat2 outdat3 outdat4 outdat5 ;
keep pophat year disp ;

data outdatb ; set outdat0 outdat1 outdat2 outdat5 ;
keep pophat year disp ;

data outdat6 outdat7 ; set outdat1 ;
pophat = cilo ;
disp = 6 ;
output outdat6 ;
pophat = cihi ;
output outdat7 ;
keep pophat year disp ;

data outdat8 outdat9 ; set outdat2 ;
pophat = cilo ;
disp = 7 ;
output outdat8 ;
pophat = cihi ;
output outdat9 ;
keep pophat year disp ;

data outdatb ; set outdatb outdat6 outdat7 outdat8 outdat9 ;
proc sort ; by disp ;

option dquote ;
filename gsafile1 "~/tenn/graph.1&cnty&iter";
filename gsafile2 "~/tenn/graph.2&cnty&iter";
filename gsafil1a "~/tenn/graphwp.1&cnty&iter";
filename gsafil2a "~/tenn/graphwp.2&cnty&iter";

/* Set up commands for graphing the ML, NLIV, Linear and Log-linear models */
```

```
goptions reset=global
      device=pslmono
      interpol=join
      gaccess='sasgaedt'
      gsfmode=replace
      rotate=landscape
      gprolog='25210A'X ;

axis1
  color=BLACK
  width=2.0
  label=(font=centx 'YEAR')
  value=(font=centx height=1)
  ;
axis2
  color=BLACK
  width=2.0
  label=(justify=r a=90 font=centx 'POPULATION, IN THOUSANDS')
  value=(font=centx height=1)
  ;

symbol1
  v = plus
  interpol=none ;
symbol2
  interpol = join
  line = 1 ;
symbol3
  interpol = join
  line = 2 ;
symbol4
  interpol = join
  line = 3 ;
symbol5
  interpol = join
  line = 4 ;
symbol6
  v = star
  interpol = none ;

proc gplot data=OUTDAT ;
  plot pophat*year = disp /
    haxis=axis1
    vaxis=axis2
    nolegend ;

goptions gsfname = gsafile1 ;

run ;
```

```
goptions device=hpljgll
      gsfmode=replace
      rotate=landscape
      gprolog='';

axis1
  color=BLACK
  width=2.0
  label=(font=centx 'YEAR')
  value=(font=centx height=1)
  ;
axis2
  color=BLACK
  width=2.0
  label=(justify=r a=90 font=centx 'POPULATION, IN THOUSANDS')
  value=(font=centx height=1)
  ;

symbol1
  v = plus
  interpol=none ;
symbol2
  interpol = join
  line = 1 ;
symbol3
  interpol = join
  line = 2 ;
symbol4
  interpol = join
  line = 3 ;
symbol5
  interpol = join
  line = 4 ;
symbol6
  v = star
  interpol = none ;

proc gplot data=OUTDAT ;
  plot pophat*year = disp /
    haxis=axis1
    vaxis=axis2
    nolegend ;

goptions gsfname = gsafil1a ;

run ;

/* Set up code for graphing linear and log-linear models with confidence intervals */
```

```
goptions reset=global
       device=pslmono
       interpol=join
       gaccess='sasgaedt'
       gsfmode=replace
       rotate=landscape
       gprolog='25210A'X ;

axis1
  color=BLACK
  width=2.0
  label=(font=centx 'YEAR')
  value=(font=centx height=1)
  ;
axis2
  color=BLACK
  width=2.0
  label=(justify=r a=90 font=centx 'POPULATION, IN THOUSANDS')
  value=(font=centx height=1)
  ;

symbol1
  v = plus
  interpol=none ;
symbol2
  interpol = join
  line = 1 ;
symbol3
  interpol = join
  line = 8 ;
symbol4
  v = star
  interpol = none ;
symbol5
  line = 2 ;
symbol6
  line = 20 ;

proc gplot data=OUTDATB ;
  plot pophat*year = disp /
    skipmiss
    haxis=axis1
    vaxis=axis2
    nolegend ;

goptions gsfname=gsafile2 ;

run ;

goptions device=hpljgll
```

```
        gsfmode=replace
        rotate=landscape
        gprolog='';

axis1
  color=BLACK
  width=2.0
  label=(font=centx 'YEAR')
  value=(font=centx height=1)
  ;
axis2
  color=BLACK
  width=2.0
  label=(justify=r a=90 font=centx 'POPULATION, IN THOUSANDS')
  value=(font=centx height=1)
  ;

symbol1
  v = plus
  interpol=none ;
symbol2
  interpol = join
  line = 1 ;
symbol3
  interpol = join
  line = 8 ;
symbol4
  v = star
  interpol = none ;
symbol5
  line = 2 ;
symbol6
  line = 20 ;

proc gplot data=OUTDATB ;
  plot pophat*year = disp /
    skipmiss
    haxis=axis1
    vaxis=axis2
    nolegend ;

goptions gsfname=gsafil2a ;

run ;

%end ;

%end ;

%mend estimate ;

%estimate ;

^Z
```