

USE OF SAS-ETS AND THE BLS-CENSUS DATA FERRETT FOR THE COMPREHENSIVE EVERGLADES RESTORATION PROGRAM

Richard A. March, South Florida Water Management District, West Palm Beach, Florida

ABSTRACT

This paper describes the proposed use of SAS/ACCESS, the BLS-Census Data FERRETT and SAS/ETS in the analysis of socioeconomic and environmental justice data for the Comprehensive Everglades Restoration Plan <http://www.evergladesplan.org>,

a co-operative plan between the South Florida Water District and the Jacksonville District of the U. S. Army Corps of Engineers. Particular attention will be paid to the use of the DataFerrett developed jointly by the U. S. Bureau of the Census and the U. S. Bureau of Labor Statistics <http://ferret.bls.census.gov>

The data ferret is an electronic data review and extraction tool designed to facilitate review and extraction of socioeconomic data from massive BLS and Census Bureau databases. The latest release of the DataFerrett is available in a Beta release in JAVA for use on the World Wide Web. One form in which the DataFerret can output data is as a SAS data set. A SAS data set will be extracted and analyzed using crosstabs and frequencies. An additional analysis relating shrimp harvest data from National Marine Fisheries Service to South Florida Water Management District operational policies using PROC PDLREG in SAS/ETS will also be presented

INTRODUCTION

In this paper, four major subject areas will be discussed: (1) an overview of the Federal statistical resources available online including the Census-BLS Data FERRETT and other resources available online; (2) an overview of data warehousing, with particular emphasis on data warehousing at the South Florida Water Management District and SAS applications at the District ; (3) an overview of the Comprehensive Everglades Restoration Program (CERP) and how SAS is being and will be used in the CERP; (4) an application of SAS-ETS using PROC PDLREG; and (5) summary and conclusions.

FEDERAL STATISTICAL DATA RESOURCES AVAILABLE ONLINE

The BLS-Census Data FERRETT (Federal Electronic Review Retrieval Extraction and Tabulation Tool) is one of a set of data extraction tools available over the Internet for accessing data from large Federal statistical databases. Major participants involved in the Data FERRETT include: (1) the Bureau of Labor Statistics of the U. S. Department of Labor; (2) the Demographic Surveys Division of the

Survey Modernization Programming Branch of the Bureau of the Census of the U. S. Department of Commerce; and (3) the Division of Public Health, Surveillance, and Informatics of the Centers for Disease Control.

At the present time, data from the following data sets are available on the FERRETT system:

Current Population Survey (CPS)

Basic Monthly Survey, January 1994 to present;
1992-2000 March Supplement;
1996, 1998 and 2000 February Displaced Worker Supplement;
1996, 1998 and 2000 February Job Tenure Supplement;
1995 May Race and Ethnicity Supplement;
1995, 1997, and 1999 February Contingent Worker Supplement;
1994-1998 October School Enrollment Supplement;
1994, 1996 and 1998 November Voting and Registration Supplement;
1999 April Food Security Supplement;
1998 August Food Security Supplement;
1997 April Food Security Supplement;
1996 September Food Security Supplement;
1995 April Food Security Supplement;
2000 August and 1998 December Internet and Computer Use Supplement;
1997 October Computer Ownership/Internet Supplement;
1994 October Computer Ownership/Uses Supplement;
1998 June Fertility and Birth Expectations Supplement;
1995 June Fertility and Marital History Supplement;
1997 May Work Schedules Supplement;
1995, 1997, 1999 Veterans Supplement;
1997, 1999 American Housing Survey (AHS) - National Survey
1998 American Housing Survey (AHS) - Metropolitan Sample
Survey of Income and Program Participation (SIPP)
1992 10 wave longitudinal;
1993 9 wave longitudinal;
1996 Panel Waves 1-12 Core;
1996 Panel Wave 1 Topical Modules;
1996 Panel Wave 2 Topical Modules;
1996 Panel Wave 3 Topical Modules;
1996 Panel Wave 11 Adult Disability Topical Module (FTP only);
1997 Survey of Program Dynamics (SPD);
1993 National Health Interview Survey (NHIS from the National Center for Health Statistics).
1988-1994 National Health and Nutrition Examination Survey III (NHANES from the National Center for Health Statistics).

1994 Mortality - Underlying Cause-of-Death (from the National Center for Health Statistics).
 1996 National Ambulatory Care Survey (NAMCS from the National Center for Health Statistics).
 1996 National Hospital Ambulatory Care Survey (NHAMCS from the National Center for Health Statistics).

Future data to be placed on the Data FERRETT include:

Population Estimates
 County Level Poverty estimates
 Economic Census
 Fedstats Quick Facts/Mapstats Database
 -EPA Data -Crime Data
 -Labor Data -U. S. Statistical Abstract
 FBI Uniform Crime Data by County
 Economic Census
 National Hospital Discharge Survey
 Multicause of Mortality

The data release schedule for Census 2000 data products is presented at the web site below:
<http://www.census.gov/population/www/censusdata/c2kproducts.html>

In general, first Census 2000 data, Congressional redistricting data, were released in March, 2001. Place and census tract data, including selected population and housing characteristics, is scheduled for release in May and June, 2001. Other Census 2000 data will be released between now and 2003. There is an increasing reliance on the Internet and CD-ROM and DVD media for disseminating Census data. Data not yet on the FERRETT system can be extracted through the Census Quickfacts (<http://quickfacts.census.gov/qfd/>) system.

Other Census data access tools include: Censtats, Map Stats, TIGER Maps, US Gazetteer, 1990 Decennial Lookup, and Data Extraction System. MAPSTATS is an interactive tool which links the Census Quickfacts database to State and County maps.

Quickfacts also has a non-graphic link that allows the user to extract historic census counts for all counties within an individual state. These data can be downloaded to ASCII files for import into SAS. Another site with a good link between its graphical data and its time-series tabular data is the CLIMVIS (Climatic Visualization) site maintained by the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (<http://www.ncdc.noaa.gov/onlineprod/drought/xmgr.html>)

Although the South Florida Water Management District has access to the NCDC data through DBHYDRO, for online acquisition of data for a single climatic region, it is generally just as easy to access that data from the web-site above, unless one requires greater temporal resolution than one-month or greater spatial resolution than the NCDC climate division (there are seven in Florida).

The Census Bureau Website indicates that: "QuickFacts includes Census 2000 counts for all persons plus race groups, Hispanic origin, and persons under 18. Other characteristics asked of everyone will be added later in the summer of 2001. Characteristics asked of a sample of households, like income and education, will become available in 2002. New statistics will appear in American FactFinder shortly before they appear in QuickFacts."
<http://quickfacts.census.gov/qfd/faq.html>

An example Census Quickfacts extraction using Florida State totals and Broward County totals is given in Table I below:

Table I : Sample Florida and Broward County Census Data Extracted using Census Quickfacts

People QuickFacts	Broward County	Florida
Population, 2000	1,623,018	15,982,378
Population, percent change, 1990 to 2000	29.3%	23.5%
White persons, percent, 2000 (a)	70.6%	78.0%
Black or African American persons, percent, 2000 (a)	20.5%	14.6%
American Indian and Alaska Native persons, percent, 2000 (a)	0.2%	0.3%
Asian persons, percent, 2000 (a)	2.3%	1.7%
Native Hawaiian and Other Pacific Islander, percent, 2000 (a)	0.1%	0.1%
Persons reporting some other race, percent, 2000 (a)	3.0%	3.0%
Persons reporting two or more races, percent, 2000	3.4%	2.4%
Persons under 18 years old, percent, 2000	23.6%	22.8%
Persons of Hispanic or Latino origin, percent, 2000 (b)	16.7%	16.8%
High school graduates, persons 25 years and over, 1990	690,696	6,616,094
College graduates, persons 25 years and over, 1990	168,799	1,624,405
Homeownership rate, 1990	68.0%	67.2%
Single family homes, number 1990	277,639	3,368,567
Households, 1990	527,860	5,138,360
Persons per household, 1990	2.35	2.46
Family households, 1990	337,284	3,541,308
Median household money income, 1997 model-based estimate	\$37,832	\$32,877
Persons below poverty, percent, 1997 model-based estimate	11.7%	14.4%
Children below poverty, percent, 1997 model-based estimate	17.5%	21.8%

DATA WAREHOUSING

Similar data down to the census tract level is available through American Factfinder. (<http://factfinder.census.gov/servlet/>)

In a press release dated May 29, 1997 the Census Bureau and the Bureau of Labor Statistics announced the release of the Federal Electronic Research and Review Extraction Tool or FERRET. Cavan Capps of the Bureau of the Census and John Bosley of the Bureau of Labor Statistics, along with their staffs are the primary developers of the Data FERRETT. The second t in FERRETT was added when the FERRETT system was integrated into the Data Web (<http://www.thedataweb.org/>)

On July 31, 2000 SAS announced that it had entered into an agreement with the Census Bureau giving the Bureau unlimited access to SAS technology solutions, including any new products created by SAS during the five-year term of the contract. The SAS news release stated, "Besides strengthening the existing relationship between SAS and the Census Bureau, the contract is also the first of its kind for the Census Bureau that is broad enough to include any and all products from a technology vendor." Data can be accessed directly from the Census Bureau's web site using the 1990 Decennial Lookup, Data Extraction System, and the FERRETT system. The sas transport dataset that comes from FERRETT is created with the export engine and proc copy. Since the sas transport dataset from FERRETT is created with the xport engine and proc copy, only xport & proc copy can be used to convert the transport dataset to a regular SAS dataset. Listed below is the SAS code needed to convert the transport dataset to a regular SAS dataset.

```
libname trans xport "c:\temp\q16363.trn";
libname new "c:\temp";
proc copy in=trans out=new;
run;
```

(The directory and filename listed in the SAS code above are used for example; the above code and description are drawn from the FERRET FAQ website, http://ferret.bls.census.gov/ferret_faq.html).

Other Social Science Data Extraction Tools

A number of other Data Extraction Tools are currently available for accessing 1990 Census of Population Data and other Census Reports. A fairly complete list of "Social Science Data Extractors" is available from the Center for Demography and Ecology at the University of Wisconsin-Madison (<http://www.ssc.wisc.edu/cde/datalib/extract.htm>)

One of the best of these is the Basic Tables: 1990 Demographic Profile Generator from the Missouri State Census Data Center

<http://mcdc2.missouri.edu/websas/xtabs3menus/mo/>

"This application is entirely menu driven and requires no code input from the user. The reports generated by this application are all done "on the fly" using a SAS(r) program which is dynamically invoked. This application generates a single 1990 "Basic Tables" (demographic profile) report for any of the supported geographic units, including census tract, block group, city (no size limit), 5-digit ZIP code, state, county or metro area for anywhere in the United States."

Another useful Website for Federal statistical data is Fedstats (<http://www.fedstats.gov/>). FedStats provides information for Federal agencies reporting expenditures of at least \$500,000 per year in one or more statistical activities including:

1. planning of statistical surveys and studies, including project design, sample design and selection, and design of questionnaires, forms, or other techniques of observation and data collection
2. training of statisticians, interviewers, or processing personnel
3. collection, processing, or tabulation of statistical data for publication, dissemination, research, analysis, or program management and evaluation
4. publication or dissemination of statistical data and studies
5. methodological testing or statistical research
6. data analysis
7. forecasts or projections that are published or otherwise made available for government-wide or public use
8. statistical tabulation, dissemination, or publication of data collected by others
9. construction of secondary data series or development of models that are an integral part of generating statistical series or forecasts
10. management or coordination of statistical operations
11. statistical consulting or training

Among the major Federal Agencies with data retrievable on-line through FEDSTATS are:

Bureau of Economic Analysis
Bureau of Justice Statistics
Bureau of Labor Statistics
Bureau of the Census
Bureau of Transportation Statistics
Energy Information Administration
Economic Research Service in the Department of Agriculture
Environmental Protection Agency
Internal Revenue Service, Statistics of Income Division
National Agricultural Statistics Service, U. S. Department of Agriculture
National Center for Education Statistics
National Center for Health Statistics
Science Resources Studies, National Science Foundation
Social Security Administration, Office of Research, Evaluation, and Statistics
Office of Management and Budget

Although not listed as principal statistical agencies by Fedstats, Federal Agencies producing statistical data important in the Comprehensive Everglades Plan include the National Marine Fisheries Service, the U. S. Geological Survey, and the U. S. Army Corps of Engineers, the South Florida Water Management District's principal Federal partner on the Comprehensive Everglades Restoration Plan. The statistical data produced by these agencies are of less interest to the lay user but are critical to the success of the Comprehensive Everglades Restoration Plan. An example using the data extraction capabilities of the NOAA Fisheries Economics website <http://www.st.nmfs.gov/st1/commercial/index.html> and SAS/ETS is presented in a later section of this paper. A useful source of historical census information is the United States Historical Census Browser., maintained at the Fisher Library at the University of Virginia .

(<http://fisher.lib.virginia.edu/census/>) This site is maintained in co-operation with the Inter-University Consortium for Political and Social Research, a unit of the Institute for Social Research at the University of Michigan. (<http://www.icpsr.umich.edu/>) Commercial websites, such as Economagic (<http://www.economagic.com>), offer a variety of time-series data sets for downloading. This site is described: "This page is meant to be a comprehensive site of free, easily available economic time series data useful for economic research, in particular economic forecasting. This site (set of web pages) was started in 1996 to help students in an Applied Forecasting class. The idea was to give students easy access to large amounts of data, and to be able to quickly get charts of that data. This is also useful during class, so that when we use the computer and overhead projector facility in class, we can quickly retrieve series and do manipulations in class. At this time, there are more than 100,000 time series for which data and custom charts can be retrieved."

Data Warehouses and Data Warehousing Tools Used at the South Florida Water Management District

According to Webopedia (<http://www.webopedia.com/>) a Data Warehouse is "A collection of data designed to support management decision making. Data warehouses contain a wide variety of data that present a coherent picture of business conditions at a single point in time. Development of a data warehouse includes development of systems to extract data from operating systems plus installation of a warehouse database system that provides managers flexible access to the data. The term data warehousing generally refers to combine (*sic.*) many different databases across an entire enterprise."

Bill Inmon, in *Building the Data Warehouse*, defines a data warehouse as: "... a subject-oriented, integrated non-volatile, time-variant collection of data in support of management decisions." (Bill Inmon, quoted in Marc Demarest, "A Data Warehouse Evaluation Model," reprinted from *Oracle Technical Journal*, October, 1995, v. 1, No. 1, p. 29)

<http://www.hevanet.com/demarest/marc/oracle7.html>

Recently, Inmon has suggested the need to move beyond the data warehouse to what he calls "the exploration warehouse." An exploration warehouse is a copy of some or all of the enterprise data warehouse designed specifically for exploration. The exploration warehouse contains detailed data and historical data copied from the enterprise data warehouse. The exploration warehouse is created directly from the enterprise data warehouse.

The exploration warehouse can be created and recreated very quickly should the data miner decide that data is needed in a different manner or that different data is needed. The data miner can use the exploration warehouse as seen fit with no consideration for the performance impact on other users. The data miner is the sole user of the exploration warehouse so there is no conflict with resource utilization with other warehouse analysts.

Cavan Capps of the U. S. Bureau of the Census, one of the primary developers of the Data FERRETT and the Data Web has stated: "In the past, data have been system-specific. Data sets have been inextricably bound

to the hardware and software systems that house them, and to the agencies that administer them. Access to data has depended upon access to the machines or the software systems that house the data. Recent technological developments make it possible to break this bind and allow for wider access to data, regardless of differences in underlying hardware and software, and regardless of organizational boundaries." (Cavan Capps U.S. Bureau of the Census The Data Web and FERRETT: Innovations in Integrating Distributed Federal, State, and Local Data,")

<http://www.spc.uchicago.edu/datalib/ia2000/prog/capps.txt>
The South Florida Water Management District answered the question, "What is a Data Warehouse?" as follows: "Data Warehouse describes a collection of operational data from many sources that has been summarized and reconciled and is presented to the decision-maker in insightful ways. The stored information derives from day-to-day operational data and has been combined, formatted and optimized in ways conducive to analytical thinking with the purpose of providing organizational decision-makers with timely information necessary to effectively make critical business solutions. The decision-maker looks at the gathered information in many ways in order to understand the critical underlying issues and trends facing the business. Analytical thinking happens through spontaneous probing, by asking questions until a problem is understood. The benefit of the Data Warehouse comes in the ability to make effective decisions from it. Cost of the Data Warehouse is reflected in its design, implementation, execution, refinement and maintenance. What end users notice most is both the quality and reliability of the data and the usefulness of the tools used to view it."

<http://cobweb2.sfwmd.gov/iwebB501/pln/proj/warehouse/warehouse.htm>

The data warehouse, as envisioned by District staff and management would include, but not be limited to, Projects, GIS Data, Surface Water Modeling Data, Groundwater Modeling Data, Census Data, Budget Information, and Personnel Information. Each of these areas has its own set of independent data management, modeling, analysis, and presentation tools. One of the major concepts being advanced in the CERP Data Management Plan (DMP), currently under development is a "data model." The CERP team in the kick-off statement for development of the DMP stated: "Our goal is to create a Data Model. Data modeling is the process of defining the grammar, vocabulary, and content to be used to represent information in a database system." DataModel.org has defined a Data Model as: "Basically, a datamodel is any method of visualizing the informational needs of a system and typically takes the form of an ERD (Entity Relationship Diagram)." (<http://www.datamodel.org>) The SAS/EIS software and SAS/MDDB Data Server software have not yet been implemented at SFWMD.

At the present time, the primary "corporate" hydrologic and water database in use at the South Florida Water Management District is an ORACLE database known as DBHYDRO. A menu-driven Oracle Forms 4.5 application HYDRO_PREP is used to define the data to be retrieved from DBHYDRO. Separate databases are maintained by different organizational units within the agency for GIS Data (in ARC-INFO), Real Estate Data (Land Acquisition Management Information System (LAMIS) in ORACLE),

DATA WAREHOUSING

Permit Data Regulatory Database ((REGDB) in ORACLE), ROSS Human Resources and Payroll System. Recently, the Water Management and the Corps of Engineers made the decision to use the Oracle 8i System as the Central CERP Data Warehouse (the CERP Zone).

"In the 1990s SAS caught up with the rest of the computer world going from a programming style language, to a graphic and menu oriented, user friendly package. The program was also rewritten in C. (Presentation by R. Thomas James to South Florida Water Management District Unix Users' Group, August 28, 1995)." In March, 1999 the South Florida Water Management District completed its IT Strategic Plan, FY 2000-2003. The District IT Strategic Plan has several Key Findings relating to Data Warehousing and Integration, including:

"The District lacks an organized information structure that permits easy access to all District data, information, and knowledge and provides integrated tools to analyze and synthesize data. Currently District data is housed in isolated systems that are generally not accessible to one another or by end users. Many different tools on many different platforms are used to analyze and synthesize these data. The resulting reports are often generated in incompatible formats that cannot be easily distributed. "Information in each operational system is often not readily available to District Employees, as well as to the public. As a consequence, data is often duplicated and extra systems are created. System users must know and use multiple processes to access and update the same data . .

.. "It is acknowledged that a certain level of compatibility is necessary to ensure that each tool can meet the needs of all District users. However, needs also exist that must be met by very specific solutions. Therefore, a level of 'service utility' needs to be defined. These utilities include network interfaces/protocols, database systems, operating systems and platform standards. Beyond that point, customization can occur to meet the specific needs without diverging from the standards that allow for the integration. Significant evidence of integration includes

Ease of data sharing

Compatibility with the defined service utility layers of the infrastructure.

Web interface. . . .

District employees constantly deal with integration issues when they exchange information with other organizations. Integration issues occur when the District or the other agency:

Have a different version of the software

Utilize a different vendor's software

Utilize different media types

Utilize different data standards"

(South Florida Water Management District, I. T. Strategic Plan, pp. Key Findings-6-11).

The I. T. Strategic Plan has identified a strategy, "Encourage adherence to industry standard software, utilize a lowest common denominator standard when using dissimilar types for exchanging, and develop a knowledge base of exchange methodologies among organizations and publish it on the District's Web (South Florida Water Management District, I. T. Strategic Plan, p. Key Findings-11). Even though the main Data Warehouse for CERP will be an Oracle 9i System much of the data will be exchanged with SAS applications. Through SAS/ACCESS Interface to Relational Databases.

The Comprehensive Everglades Restoration Program

The Comprehensive Everglades Restoration Program is a joint program between the U.// S. Army Corps of Engineers and the South Florida Water Management District. The Comprehensive Everglades Restoration Program is authorized under Section 601 of the Water Resources Development Act of 2000 (WRDA 2000). This act provides, in part,

"(A) IN GENERAL- Except as modified by this section, the Plan is approved as a framework for modifications and operational changes to the Central and Southern Florida Project that are needed to restore, preserve, and protect the South Florida ecosystem while providing for other water-related needs of the region, including water supply and flood protection. The Plan shall be implemented to ensure the protection of water quality in, the reduction of the loss of fresh water from, and the improvement of the environment of the South Florida ecosystem and to achieve and maintain the benefits to the natural system and human environment described in the Plan, and required pursuant to this section, for as long as the project is authorized."

On May 12, 2000 a Design Agreement between the Department of the Army and the South Florida Water Management District for the Design of Elements of the Comprehensive Plan for the Everglades and South Florida Ecosystem Restoration Project (CERP) was signed. This Design Agreement, among other things, mandates the preparation of a Master Program Management Plan (MPMP). This Master Program Management Plan contains an Appendix F, discussing Programmatic Activities. Programmatic activities are defined as "... activities and tasks that are not linked to a specific project, but involve or affect more than one project or the entire restoration program. These activities include Restoration Coordination and Verification (RECOVER), public outreach, socioeconomic and environmental justice studies, program management and technical co-ordination activities." (Master Program Management Plan (http://www.evergladesplan.org/pm/docs/mpmp_08182000/MPMP%20Appendix%20F%20-%20Final%20000818.pdf))

Environmental justice, a particular concern of the CERP, is mandated by Executive Order 12898 signed by President Clinton on February 11, 1994. This Order states, in part, "To the greatest extent practicable and permitted by law, and consistent with the principles set forth in the report on the National Performance Review, each Federal agency shall make achieving environmental justice part of its mission by identifying and addressing, as appropriate, disproportionately high and adverse human health or environmental effects of its programs, policies, and activities on minority populations and low-income populations." Census data, particularly small-area studies, play an important part in targeting CERP-related activities to identify and address environmental justice issues.

The Water Resources Development Act of 2000, the Federal Legislation authorizing much of the work under the CERP specifically addresses environmental justice concerns as shown below.

OUTREACH AND ASSISTANCE-

(1) SMALL BUSINESS CONCERNS OWNED AND OPERATED BY SOCIALLY AND ECONOMICALLY DISADVANTAGED INDIVIDUALS- In executing the Plan, the Secretary shall ensure that small business concerns owned and controlled by socially and economically disadvantaged individuals are provided opportunities to participate under section 15(g) of the Small Business Act (15 U.S.C. 644(g)).

(2) COMMUNITY OUTREACH AND EDUCATION-

(A) IN GENERAL- The Secretary shall ensure that impacts on socially and economically disadvantaged individuals, including individuals with limited English proficiency, and communities are considered during implementation of the Plan, and that such individuals have opportunities to review and comment on its implementation.

(B) PROVISION OF OPPORTUNITIES- The Secretary shall ensure, to the maximum extent practicable, that public outreach and educational opportunities are provided, during implementation of the Plan, to the individuals of South Florida, including individuals with limited English proficiency, and in particular for socially and economically disadvantaged communities"

Example Using SAS/ETS and NMFS Data-

Data on monthly West Coast of Florida pink shrimp landings were obtained as an ASCII file from Guy Davenport of the National Marine Fisheries Service Miami Office. Data covering the period May, 1966 to August, 1998 was analyzed. The NMFS data extraction tool could not be used to generate the data set because monthly landings before 1990 are not on the system. The NMFS ASCII file was imported, using the import Wizard into SAS, and a SAS data set "shrimppounds.sas" was set up. The NMFS data was augmented with monthly flow data (maximum flow and average flow data for the Taylor Slough Bridge located at the mouth of the Caloosahatchee River, near Fort Myers) from the District's DBHYDRO database. Harvest data was obtained by water bodies; however, for the purposes of this presentation only the aggregate West Coast of Florida data is presented. A polynomial distributed lag model using the iterated Yule-Walker (method= ITYW) autocorrelation correction method is estimated. The iterated Yule-Walker method is described as: follows:

Let ϕ represent the vector of autoregressive parameters
 $\phi = (\phi_1, \phi_2, \phi_3, \dots, \phi_n)'$
 and let the variance matrix of the error vector
 $v = (v_1, \dots, v_N)'$ be Σ
 $E(vv') = \Sigma = \sigma^2 V$.

If the vector of autoregressive parameters ϕ is known, the matrix V can be computed from the autoregressive parameters. Σ is then $\sigma^2 V$. Given Σ , the efficient estimates of regression parameters can be computed using generalized least squares (GLS). The GLS estimates then yield the unbiased estimate of the variance, σ^2 .

The Yule-Walker method alternates estimation of β using generalized least squares with estimation of ϕ using the Yule-Walker equations applied to the sample autocorrelation function. The YW method starts by forming the OLS estimate of β .

Next, ϕ is estimated from the sample autocorrelation function of the OLS residuals using the Yule-Walker equations. Then V is estimated from the estimate of ϕ , and β is estimated from V and the OLS estimate of σ^2 . The autocorrelation corrected estimates of the regression parameters β are then computed by GLS using the estimated Σ matrix. These are the Yule-Walker estimates.

If the ITER option is specified, the Yule-Walker residuals are used to form a new sample autocorrelation function, the new autocorrelation function is used to form a new estimate of ϕ and V , and the GLS estimates are recomputed using the new variance matrix. This alternation of estimates continues until either the maximum change in the ϕ -hat estimate between iterations is less than the value specified by the CONVERGE= option or the maximum number of allowed iterations is reached. This produces the iterated Yule-Walker estimates. Iteration of the estimates may not yield much improvement.

The Yule-Walker equations, solved to obtain ϕ -hat and a preliminary estimate of σ^2 , are $R \phi\text{-hat} = -r$.

This discussion, along with a discussion of alternative autocorrelation correction methods, is in the on-line SAS/ETS User's Manual at the link below.

<http://calcul.si.uji.es/Programes/SAS/ets/chap8/sect21.htm>

Because pink shrimp harvest is, by definition, non-negative, the regression was forced through the origin using the noint option, as shown in the SAS code below.

```
proc pdlreg data=shrimppounds;
  model all=sinterm costerm real_price
  aveflow(12,2,1) maxflow(12,2,1)/noint NLAG=12
  METHOD=ITYW;
  output out=shrimp P=predict R=resid;
run;
```

The variables included in the above model statement are:
 All = monthly pink shrimp landings in the West Coast of Florida (Gulf) region, pounds;
 Time= a time-trend variable equal to 1 in May, 1966 and increasing one unit per month thereafter;
 Sinterm= $\text{SINE}((2\pi/12) \cdot \text{Time})$;
 Costerm= $\text{COSINE}((2\pi/12) \cdot \text{Time})$;
 Real_price = (Pink shrimp revenue/Pink shrimp landings)/CPI(1982-84 base), (\$/pound)
 The CPI data was retrieved from the BLS web site:
<http://stats.bls.gov/cpihome.htm>

Aveflow= average monthly flow (cfs) through the S-79 water control structure, as reported by the South Florida Water Management District;
 Maxflow= peak flow (cfs) during a particular month through the S-79 water control structure, as reported by the South Florida Water Management District

DATA WAREHOUSING

The pink shrimp landings and revenue data will periodically be updated from the NMFS web site: http://www.st.nmfs.gov/st1/commercial/landings/monthly_landings.html

"Florida is divided into three areas (East coast, West coast, and inland waters) in our surveys. Florida east coast and inland water landings are summarized with South Atlantic states and Florida west coast landings are summarized when Gulf landings are requested." <http://www.st.nmfs.gov/webplcomm/plsql/webst1.ft.help.species>

The terms sinterm and costerm are included to capture the seasonality of pink shrimp landings. The underlying theory is based on spectral analysis, which is explained in the SAS/ETS User's Guide under PROC SPECTRA at the link below.

<http://calcul.si.uji.es/Programes/SAS/ets/chap17/sect1.htm>

The model estimation results are shown below.

The SAS System 13:29 Thursday, April 12, 2001

The PDLREG Procedure

Dependent Variable All

Ordinary Least Squares Estimates

SSE	3.67472E14	DFE	367
MSE	1.00129E12	Root MSE	1000643
SBC	11501.0457	AIC	11465.6794
Regress R-Square	0.8199	Total R-Square	0.8199
Durbin-Watson	0.6403		

NOTE: No intercept term is used. R-squares are redefined.

Variable	DF	Standard Estimate	Approx Error t Value	Pr > t
sinterm	1	618053	86585 7.14	<.0001
costerm	1	-1079908	85167 -12.68	<.0001
real_price	1	301966	43690 6.91	<.0001
Aveflow**0	1	-182.6894	41.9764 -4.35	<.0001
Aveflow**1	1	-140.5207	47.8809 -2.93	0.0035
Aveflow**2	1	-5.4397	47.4891 -0.11	0.9089
Maxflow**0	1	146.7356	23.3197 6.29	<.0001
Maxflow**1	1	70.6262	30.1774 2.34	0.0198
Maxflow**2	1	16.9913	30.4533 0.56	0.5772

Yule-Walker Estimates

SSE	1.65816E14	DFE	355
MSE	4.67087E11	Root MSE	683438
SBC	11280.5936	AIC	11194.1427
Regress R-Square	0.3123	Total R-Square	0.9187

NOTE: No intercept term is used. R-squares are redefined.

Autoregressive parameters assumed given.

Variable	DF	Standard Estimate	Approx Error t Value	Pr > t
TIME	1	7070	1606 4.40	<.0001

costerm	1	-1040981	122991 -8.46	<.0001
sinterm	1	619130	125400 4.94	<.0001
real_price	1	47085	81588 0.58	0.5642
Aveflow**0	1	8.5493	61.0774 0.14	0.8888
Aveflow**1	1	-87.2570	45.5037 -2.14	0.0333
Aveflow**2	0	0		
Maxflow**0	1	4.2912	36.5003 0.12	0.9065
Maxflow**1	1	53.0982	28.0952 1.89	0.0596
Maxflow**2	1	6.5254	13.3478 0.49	0.6252

Additional output from PROC PDLREG and simulated and forecast values are available from the author upon request.

The model estimated above can be used to model simulated impacts to West Coast of Florida pink shrimp landings from changes in exogenous variables including water flows, real price, and seasonality. A number of the scenarios being examined for the Comprehensive Everglades Restoration Plan will alter flows through the S-79 water control structure. The results show that there is a strong seasonal pattern to pink shrimp landings, and that increases in average flows tend to increase pink shrimp landings, while increases in maximum daily flows, representing flushes of fresh water, tend to decrease landings. All of these effects take place with a lagged effect. It is hypothesized that this reflects the spawning and growth cycles cycle of pink shrimp. Increases in real prices are associated with increased landings; this relationship needs to be further investigated through an econometric model that examines both supply-related factors and demand-related factors.

Conclusions

The BLS and Census Bureau Data FERRETT system and other data extraction tools are useful in generating SAS data sets (and ASCII files for import into SAS and other statistical applications) for statistical and econometric analysis. The Data FERRETT is better suited, at the present time, to extracting cross-sectional and recent data; BLS and the Census Bureau have other data extraction tools better suited to generating long-time series (pre-1990). Other Federal, state, and regional agencies, such as NMFS and the South Florida Water Management District, generally have usable databases, although these agencies, particularly the Water Management District, have made less progress in making their data readily accessible than have BLS and the Census Bureau.

A major challenge to all agencies in building data warehouses will be appropriately integrating spatially disaggregated data (such as TIGER) files and time-series data sets with a long period of record and/or a short time-step. A number of sites, including National Climatic Data Center, are developing point-and-click data access tools, like MAPSTATS, which are integrated with a geographical information system.