

## **Forecasting Exogenous Variables Using SAS PROC NLIN**

**Galip Feyzioglu, Public Utilities Commission of Ohio  
David Wang, Public Utilities Commission of Ohio**

### **INTRODUCTION**

It is a matter of public record that long term forecast trajectories characterizing the dynamic behavior of many accounting time series in terms of conventional economic forecasting models have diverged from the *ex post facto* realizations of the accounting time series in question since 1973. Peak loads, kWh sales, revenues generated from sales, fuel costs, construction costs, the consumer price index (CPI), service area population, service area income, etc. are all accounting time series. Without reliable forecasts of such magnitudes, the rational design and implementation of economic plans or policies, including generating system expansion plans or demand management policies or procedures, are *ex ante* exercises in wishful thinking that may turn out to be costly errors *ex post facto*.

Conventional models of growth in market demand project trajectories of growth which are deductively equivalent to linear or nonlinear transformations of a set of linear or exponential trends. In practice, most conventional forecast trajectories are explicitly expressed as projecting growth at a constant percentage rate per year. Constant percentage rate of growth trajectories, or exponential growth trajectories, are divergent trajectories. That is, as time increases, the size of the trajectory grows without an upper bound. The characterization of growth in finite systems in terms of divergent growth trajectories is bound to lead to a divergence between projections and observations sooner or later. A clear demonstration of this fact is provided in exhibit 1 which contrasts the consolidated ten year peak load growth forecasts of the U.S. electric utilities, produced between 1974 and 1989, against the actual growth in U.S. peak load.

The methodology developed by PUCO, Division of Forecasting, to analyze the forecast demographic, economic and energy systems is based upon a system theoretical framework within which the Von Bertalanffy class of "S" curve models (PUCO, 1991a & 1991b) are applied to the characterization of time path trajectories of accounting time series. Constraining growth in accounting time series to a class of "S" curve functions is, in part, a matter of logical necessity. This is because no finite system can grow indefinitely at a constant percentage rate of growth. All growth on a finite planet has to converge to some finite upper bound. All "S" curve growth functions do converge to finite upper bounds. The magnitude of an "S" curve growth function's upper bound is generally an explicit parameter in the mathematical formulation of the function. The merits and the comparative forecasting performance of the PUCO methodology may be observed in exhibit 2.

### **THE GENERALIZED VON BERTALANFFY GROWTH MODEL**

Analyses of historical accounting time series indicate that many macroeconomic accounting magnitudes, such as population, gross national product, and consumer price index, as well as many microeconomic magnitudes, such as per customer sales and per customer revenues, follow rather smooth S-shaped growth patterns over most of the course of their respective life cycles. The smooth growth patterns are altered, however, at a limited number of significant points such as the onset or the end of the Great Depression, World War II, and the OPEC price increases.

A general methodology for measuring and incorporating the impact of historical or future catastrophes on the life cycle trajectories of empirical economic systems was developed by Feyzioglu (Feyzioglu, 1983). This methodology identifies and quantifies the impact of systemic catastrophes in terms of a finite set of points in the domain of time at which the relevant time path trajectories are nondifferentiable and the corresponding first order time derivatives are discontinuous.

In order to simulate the growth patterns in empirical economic systems, the Von Bertalanffy growth model has been revised to incorporate historical and future calamities or breakthroughs that have influenced or may influence such a system over time (Feyzioglu and Choueiki, 1989). The result is called the Generalized Von Bertalanffy growth model. It is a nonlinear spline function of the form:

$W(t) = [\alpha^{1-m} - \beta e^{-\gamma t}]^{\frac{1}{1-m}}$ , where  $\alpha$  is the limiting size of growth,  $\beta = \alpha^{1-m} - W(0)^{1-m}$  is a measure of the difference between the limiting size of the growth and the initial size of growth, and

$PI = \alpha \cdot m^{\frac{1}{1-m}}$ , where  $PI$  is the ordinate of the point of inflection and  $m > 1$ .  $f(t) = \sum_{c=1}^l k_c t_c$  is a linear

spline function if  $l > 1$ , indicating the existence of  $l - 1$  points of nondifferentiability along the functional time path trajectory. The  $t_c$ ,  $c = 1, \dots, l$ , are transformations of the domain of time indicating the duration of the  $c$ th systemic "catastrophe" at a given point in the domain of time.  $k_c$  measures the marginal impact on the time rate of change of  $f(t)$  of the  $c$ th systemic

"catastrophe", where  $\partial f(t)/\partial t = \sum_{c=1}^l k_c Dum_c$ , and  $Dum_c = 0$  if time is less than or equal to the time of

incidence of the  $c$ th systemic "catastrophe", else  $Dum_c = 1$ . We have found that the model identification and estimation phases are greatly enhanced if the parameters to be estimated are specified as  $PI$ ,  $W(0)$ ,  $\{W(t_c), c=1, 2, \dots, l\}$ , where  $W(t_c)$  is the magnitude of the time path equation at the time of incidence of the  $c$ th systemic catastrophe; and  $\alpha$  and  $k_c$  are then calculated using definitional identities.

Finally in the case of the Gompertz model, which is the limiting form of the Von Bertalanffy model as  $m \rightarrow 1$ , the parameter  $m$  does not occur and the functional form of the model is given by  $W(t) = \alpha \cdot e^{(-\beta \cdot e^{-\gamma t})}$ .

## THE SAS PROGRAM

Modular programs in SAS macro language\* which apply PROC NLIN\* to obtain weighted least squares estimates of parameters in generalized Von Bertalanffy or Generalized Gompertz growth models, with up to 18 points of nondifferentiability, were developed by Galip Feyzioglu. The statements within each program are arranged so that a number of smaller macros are nested within one large macro called EXOGVONB which executes the entire program (For an example of the SAS code of the program, refer to exhibit 3). The user selects the type of model to be used (Von Bertalanffy or Gompertz), the location of the discontinuities within the time path trajectory, the point of inflection, (and  $m$  in the case of Von Bertalanffy), and the theoretical percent error. The program calculates the partial derivatives of all of the parameters that are estimated within the NLIN procedure. It also defines the parameter dependencies and the parameter bounds to be used by the Gauss-Newton iterative search method, and generates forecasted values with 95% confidence levels.

The program is designed to investigate the magnitude, timing and existence of systemic catastrophies in a model identification phase. The output of the model includes tests of observed errors and observed error ratios against expected error limits and expected error ratios.

## AN APPLICATION OF THE GENERALIZED GOMPERTZ MODEL TO THE U.S. POPULATION GROWTH TRAJECTORY

A three-variable, i.e. two catastrophe, Generalized Gompertz growth model is utilized to describe the U.S. population growth trajectory within a standard error of the estimate of .45% of observations. The two catastrophes, occur in 1929 and 1946. Clearly, the technical use of the term catastrophe, i.e. a point of nondifferentiability in the time path trajectory is not the same as the historical usage of the term. In 1929, the technical catastrophe is obviously the onset of the Great Depression. In 1946, the technical catastrophe is the end of WWII. The high in employment rates during the Depression, and the intensive overseas military involvement during WWII, along with restrictive immigration policies, limited the formation of new families and kept the size of the existing families down through voluntary or involuntary restrictions on new births. Whereas at the end of WWII U.S. emerged as the victorious leader of a new world order, and the economic lessons learned from the Depression and the war, helped drop the U.S. economy into a relatively uninterrupted growth trajectory. The economic and the military constraints having been eliminated, both family formation rates, and new birth rates rebounded from the slump of the Depression and war years. In the absence of another Great Depression or another World War, we project the prevailing trajectory into 2010 without further modification. Some of the program outputs are presented in exhibits 4-5.

### REFERENCES

- Feyzioglu, G., (1983). The Analysis and Assessment of Time Variant Linear Trends in Annual Economic Data Series with an Application to Energy Forecasting for the State of Ohio, Ph.D. Dissertation, The Ohio State University, Columbus, Ohio.
- Feyzioglu, G. and Choueiki, H. M., An S-Curve Growth Model Applied to U.S. Population, Per Capita Peak Load Demand, Per Capita Capability, and Reserve Margin Trajectories from 1920 to 2008, IFAC/IFORS/IAEE, International Symposium on Energy Systems, Management and Economics, October 1989, Tokyo, Japan.
- PUCO, (1991a). Ohio Long Term Forecast of Energy Requirements 1990-2010, Feyzioglu, G. (Ed.), Columbus, Ohio.
- PUCO, (1991b). Technical Appendix to Ohio Long Term Forecast of Energy Requirements 1990-2010, Feyzioglu, G. & Choueiki, H., (Eds.), Columbus, Ohio.

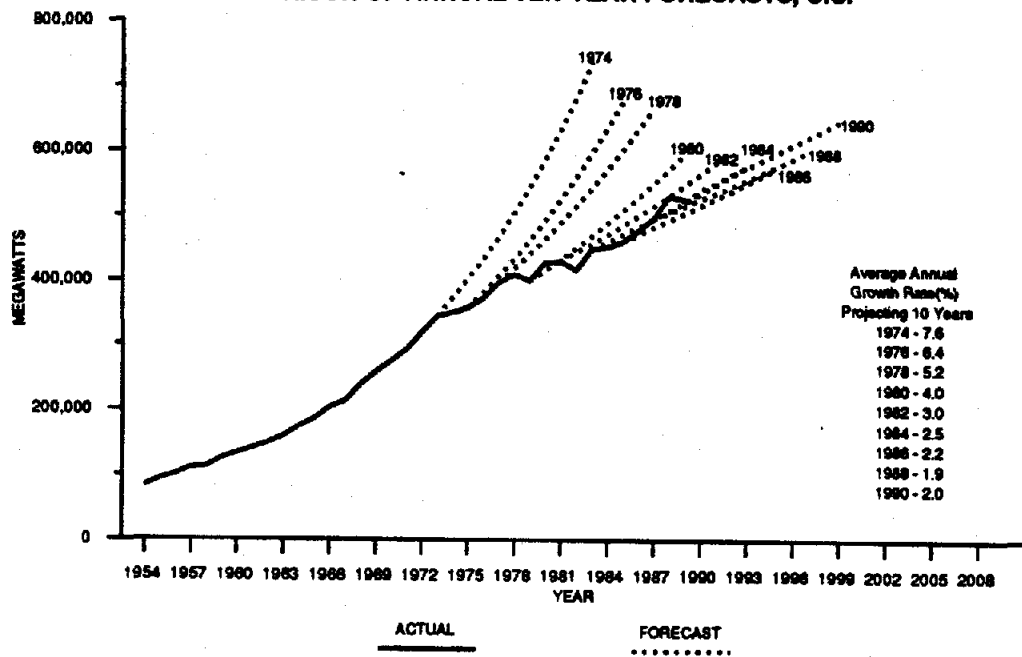
### TRADEMARK NOTICE

\*SAS and \*SAS/STAT are trademarks of SAS Institute Inc., Cary NC, USA.

#### Authors:

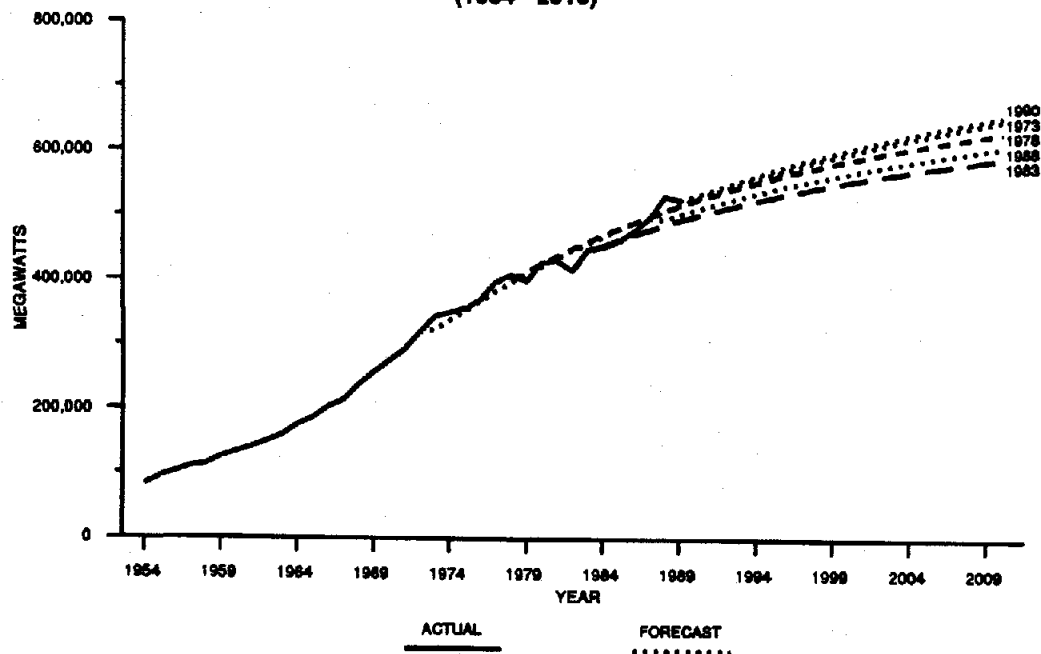
Galip Feyzioglu  
David Wang  
Public Utilities Commission of Ohio  
Division of Forecasting  
180 East Broad Street, Floor 3  
Columbus, Ohio 43266-0573

**EXHIBIT 1.  
NERC SUMMER PEAK LOAD PROJECTIONS  
COMPARISON OF ANNUAL TEN-YEAR FORECASTS, U.S.**



Source: 1974 - 1990 Electricity Supply & Demand, NERC.

**EXHIBIT 2.  
UNITED STATES SUMMER PEAK LOAD  
AS PROJECTED BY THE GOMPERTZ GROWTH MODEL  
(1954 - 2010)**



Source: EEI Statistical Yearbook and PUCO, Division of Forecasting.

### EXHIBIT 3. SAMPLE PROGRAM OF A THREE VARIABLE GOMPERTZ MODEL

```
DATA EXOG 89;
MERGE GF.F_USEXOG GF.RHS_1890;
BY YEAR;
IF YEAR LT 1890 OR YEAR GT 2010 THEN DELETE;
```

```
%MACRO EXOGVONB;
```

```
/* THREE VARIABLE GOMPERTZ MODEL */
```

```
%MACRO GOMP3T(LHS=,ALT=,XX=OH,Y0=1890,YM=2010,
YN=1989,YF=2010,YZ=YN,II=890,START=1,STOP=6,
INCLUDE=0,S1=.035,S2=.035,ALPHA=4,ADJUST=0,
CFP=0,CF0=0,L0 PI=A0 PI=U0 PI=P0 Y1=P0 1=,
WPO 1=P0 Y2=P0 2=WPO 2=LL1=,UNIT=,SCENARIO=0,
D_SCENAR=0,RHS=T&10 T&11 T&12,IO=37,I1=,I2=,
I3=88,W0=,W1=,W2=,W3=,WX=,R1=&L PI=&ALT PI<&U PI,
R2=&L I<&ALT I<&U I,R3=&L I1<&ALT I1<&U I1,
R4=&L I2<&ALT I2<&U I2,R5=&L I3<&ALT I3<&U I3);
```

```
DATA; SET; /* DATA SET */
%COVAR
%LIM G3
%STEPW_GV
```

```
PROC NLIN;
TITLE1 'THE TIME PATH TRAJECTORY OF';
TITLE2 '&LL1';
TITLE3 'FROM &Y0 TO &YF';
TITLE4 'AS DESCRIBED BY';
TITLE5 'THE GENERALIZED GOMPERTZ GROWTH MODEL.';
TITLE6 'A*EXP(-B*EXP(-(K*T&10+D*T&11+E*T&12)))';
TITLE7 'ERROR RATIO = &S1 IF YEAR LT &YM';
TITLE8 'ERROR RATIO = &S2 IF YEAR GE &YM';
TITLE9 'UNITS = &UNIT';
PARMS &LHS.PI=1 &LHS.I=1 &LHS.I1=1
&LHS.I2=1 &LHS.I3=1;
IF ITER =0 THEN IF _N_=1 THEN DO;
&LHS.PI=&LHS.PI;
&LHS.I=&LHS.I;
&LHS.I1=&LHS.I1;
&LHS.I2=&LHS.I2;
&LHS.I3=&LHS.I3;
END;
```

```
BOUNDS &R1,&R2,&R3,&R4,&R5;
```

```
%FUNN_G0
%FUNN_G1
%FUNN_G2
%FUNN_G3
%CF_G3
```

```
&LHS.XP=&LHS.K*T&10+&LHS.D*T&11+&LHS.E*T&12;
DXP_DPI=DK DPI*T&10+DD DPI*T&11+DE DPI*T&12;
DXP_DI=DK DI*T&10+DD DI*T&11+DE DI*T&12;
DXP_DI1=DK DI1*T&10+DD DI1*T&11+DE DI1*T&12;
DXP_DI2=0 +DD DI2*T&11+DE DI2*T&12;
DXP_DI3=0 +0 +DE DI3*T&12;
```

```
U=EXP(-&LHS.B*EXP(-&LHS.XP));
DU_DPI=(-DB DPI*EXP(-&LHS.XP)+(-&LHS.B*(-DXP_DPI)
*EXP(-&LHS.XP)))^U;
DU_DI=(-DB DI*EXP(-&LHS.XP)+(-&LHS.B*(-DXP_DI)
*EXP(-&LHS.XP)))^U;
DU_DI1=&LHS.B*(-DXP DI1*EXP(-&LHS.XP))^U;
DU_DI2=&LHS.B*(-DXP DI2*EXP(-&LHS.XP))^U;
DU_DI3=&LHS.B*(-DXP DI3*EXP(-&LHS.XP))^U;
```

```
MODEL &LHS&XX=&LHS.A^U;
DER.&LHS.PI=DA DPI^U+DU DPI*&LHS.A;
DER.&LHS.I=DU DI*&LHS.A;
DER.&LHS.I1=DU DI1*&LHS.A;
DER.&LHS.I2=DU DI2*&LHS.A;
DER.&LHS.I3=DU DI3*&LHS.A;
```

```
WEIGHT =W &LHS;
OUTPUT OUT=&XX&ALT P=P &ALT R=R &ALT
PARMS=PI&ALT &ALT I1&ALT I2&ALT I3&ALT;
```

```
DATA T &ALT; SET &XX&ALT; /* DATA SET */
%RECON_GV
```

```
Q1&ALT=LOG(I1&ALT/&LHS.A);
Q2&ALT=LOG(I2&ALT/&LHS.A);
Q3&ALT=LOG(I3&ALT/&LHS.A);
L1&ALT=LOG(Q1&ALT/Q0&ALT);
L2&ALT=LOG(Q2&ALT/Q0&ALT);
L3&ALT=LOG(Q3&ALT/Q0&ALT);
&ALT_K=-(1/((&I1-900*DUM_I1)-(&I0-900*DUM_I0)))
*LT&ALT;
&ALT_D=-(1/(&I2-(&I1-900*DUM_I1)))
*(L2&ALT+(&I2-(&I0-900*DUM_I0))*&ALT_K);
&ALT_E=-(1/(&I3-(&I2)))*(L3&ALT+(&I3-(&I0-900*DUM_I0))
*&ALT_K+(&I3-(&I1-900*DUM_I1))*&ALT_D);
XP&ALT=&ALT_K*T&10+&ALT_D*T&11+&ALT_E*T&12
+&SCENARIO;
XD&ALT=&ALT_K*DUM&10+&ALT_D*DUM&11+&ALT_E
*DUM&12+&D_SCENAR;
```

```
%D DT GO %PRESENT
%MEND GOMP3T;
```

```
/*
%IF %EVAL(&SYSUS=YES) %THEN %DO;
%GOMP3T(LHS=PP,ALT=PP,XX=US,Y0=1890,II=890,
S1=.0045,S2=.0045,ALPHA=4,STOP=1,L0 PI=122,
A0 PI=130,U0 PI=138,P0 Y1=&I1,P0 1=1,WPO 1=&W1,
P0 Y2=&I2,P0 2=2,WPO 2=&W2,LL1=U.S. POPULATION,
UNIT=10**6 PERSONS,I0=890,I1=28,I2=45,I3=89,
W0=63.056,W1=120.509,W2=139.928,W3=248.780,WX=&W3)
%END;
```

```
%IF %EVAL(&SYSOH=YES) %THEN %DO;
%GOMP2T(LHS=PP,ALT=PP,XX=OH,Y0=1930,YM=1954,II=30,
S1=.0130,S2=.0030,ALPHA=4,STOP=6,L0 PI=4.0,
A0 PI=5.0,U0 PI=6.5,P0 Y1=&I0,P0 1=0,WPO 1=&W0,
P0 Y2=&I1,P0 2=1,WPO 2=&W1,LL1=OHIO POPULATION,
UNIT=10**6 PERSONS,I0=44,I1=67,I2=89,
W0=6.839,W1=10.749,W2=10.907,WX=&W2)
%END;
```

```
%MEND EXOGVONB;
```

```
/*
%LET SYSOH=NO;
%LET SYSUS=YES;
/*
%EXOGVONB
```

# EXHIBIT 4.

THE TIME PATH TRAJECTORY OF  
U.S. POPULATION  
FROM 1890 TO 2010  
AS DESCRIBED BY  
THE GENERALIZED GOMPERTZ GROWTH MODEL:  
 $A = \exp(-B \cdot \exp(-(K \cdot T890 + D \cdot T28 + E \cdot T45)))$   
ERROR RATIO = .0048 IF YEAR LT 2010  
ERROR RATIO = .0045 IF YEAR GE 2010  
UNITS = 10\*\*6 PERSONS

## NON-LINEAR LEAST SQUARES SUMMARY STATISTICS DEPENDENT VARIABLE PPUS

| SOURCE            | DF | WEIGHTED SS  | WEIGHTED MS |
|-------------------|----|--------------|-------------|
| REGRESSION        | 5  | 3508088.8389 | 701217.7680 |
| RESIDUAL          | 66 | 83.9988      | 1.2727      |
| UNCORRECTED TOTAL | 71 | 3508172.8385 |             |
| (CORRECTED TOTAL) | 70 | 219764.2345  |             |

| PARAMETER | ESTIMATE    | ASYMPTOTIC<br>STD. ERROR | ASYMPTOTIC 95 %<br>CONFIDENCE INTERVAL |              |
|-----------|-------------|--------------------------|----------------------------------------|--------------|
|           |             |                          | LOWER                                  | UPPER        |
| PPPI      | 132.3589881 | 2.5888174892             | 127.22857052                           | 137.48540559 |
| PPI       | 82.4885000  | 1.8695888443             | 59.15505381                            | 85.82194839  |
| PPI1      | 120.1945583 | 0.2343592888             | 119.72684220                           | 120.66247035 |
| PPI2      | 138.8886000 | 0.2338510772             | 138.20170090                           | 139.13549940 |
| PPI3      | 247.3833298 | 0.4432718049             | 246.47830831                           | 248.24835144 |

# EXHIBIT 5.

TIME PATH TRAJECTORY OF U.S. POPULATION  
FROM 1890 TO 2010  
(GENERALIZED GOMPERTZ GROWTH MODEL)

