

PROC AUTOREG TO FORECAST WITH A TIME- SERIES/CROSS-SECTION AUTOCORRELATION MODEL

Neil S. Fleming, Ph.D., C.Q.E.
Senior Consultant

Willy Waks
President

2W Systems Co., Inc.
5430 Alpha Road, Suite 200
Dallas, Texas 75240-4506

(214) 733-0588 (214) 661-8619

November 15, 1991

(2 W Systems Co., Inc.)

THE PROBLEM

- Gas utilities must read meters and bill customers on lagged basis by cycle each month.
- Housing units within billing cycles are related to:
 - 1) each other over time because same units are involved (time-series).
 - 2) housing units in different billing cycles because similar conditions (e.g., geographic proximity) exist for closely occurring cycles (cross-section).
- Structure suggests use of time-series/cross-section linear regression models.

AVERAGE USE STRUCTURAL MODEL

$$\begin{aligned} \text{AVG_USE}_{it} = & B_{0it} + B_1 \text{HDD}_{it} + B_2 \text{CDD}_{it} \\ & + B_{3m} \text{MOISm}_{it} + B_4 \text{YEAR_NUM}_{it} + e_{it} \end{aligned}$$

- Constant
- Heating degree days (HDD)
- Cooling degree days (CDD)
- Seasonal factors (MOISm)
- Linear annual trend (YEAR_NUM)
- Error e_{it} is associated with i th billing cycle and t th month

(2 W Systems Co., Inc.)

TOTAL USE IDENTITY

$$\text{TOT_USE}_{it} = \text{AVG_USE}_{it} * \text{COUNT}_{it}$$

- Average use in i th billing cycle and t th month (AVG_USE_{it})
- Number of customers in i th billing cycle and t th month (COUNT_{it})

(2 W Systems Co., Inc.)

OPERATIONAL ERROR STRUCTURE

$$e_j = \rho_{CS(1)} e_{j-1} + \rho_{CS(2)} e_{j-2} \\ \rho_{TS(1)} e_{j-20} + u_j$$

- Observations are sorted by month and billing cycle
- j = observation number:
 $j = 20 * (t-1) + i$, t =month, i =billing cycle

$$E(e_j^2) = \sigma_e^2$$

$$u_j \sim N(0, \sigma_u^2)$$

(2 W Systems Co., Inc.)

ESTIMATION OF AUTOCORRELATION FROM OLS

$$\hat{e}_j = \text{AVG_USE}_j - (\hat{B}_{0j} + \hat{B}_1 \text{HDD}_j + \hat{B}_2 \text{CDD}_j + \hat{B}_{3m} \text{MOISm}_j + \hat{B}_4 \text{YEAR_NUM}_j)$$

$$\hat{\gamma}_0 = \sum_{j=1}^n \hat{e}_j^2 \quad \hat{\gamma}_{20} = \sum_{j=21}^n \hat{e}_j \hat{e}_{j-20}$$

$$\hat{\rho}_{\text{TS}(1)} = \hat{\rho}_{20} = \sum_{j=21}^n \hat{e}_j \hat{e}_{j-20} / \sum_{j=1}^n \hat{e}_j^2$$

(2 W Systems Co., Inc.)

ESTIMATION OF AUTOCORRELATION

CONTINUED

$$\hat{\gamma}_1 = \sum_{j=2}^n \hat{e}_j e_{j-1} \quad \hat{\gamma}_2 = \sum_{j=3}^n \hat{e}_j \hat{e}_{j-2}$$

$$\hat{\rho}_{CS(1)} = \hat{\rho}_1 = \sum_{j=2}^n \hat{e}_j \hat{e}_{j-1} / \sum_{j=1}^n \hat{e}_j^2$$

$$\hat{\rho}_{CS(2)} = \hat{\rho}_2 = \sum_{j=3}^n \hat{e}_j \hat{e}_{j-2} / \sum_{j=1}^n \hat{e}_j^2$$

(2 W Systems Co., Inc.)

INDEPENDENT VARIABLES

WEATHER

- Average read date and average number of days since last meter read date are computed for all customers within cycle
- Heating degree days are from NOAA: max of 0 and $(65^{\circ}\text{F} - \text{daily temperature})$
- Cooling degree days are max of 0 and $(\text{daily temperature} - 65^{\circ}\text{F})$
- Both heating and cooling degree days for cycle are computed based on average read date and average days since last read date

INDEPENDENT VARIABLES

NON-WEATHER

- Average read date and average number of days since last meter read date are computed for all customers within cycle
- Cycles are assigned to specific calendar months according to majority of days between average read date and last read date
- Seasonality variables MOIS1-MOIS11 assigned based on month, with August omitted
- Annual linear trend variable is 0 in 1983, and incremented one through 1989

SAS CODE FOR PROC AUTOREG

```
DATA BYCYCLE1;  
SET RESID.BYCYCLE;  
IF _N_ GT 1200 THEN MYAVG=.;  
ELSE MYAVG=ACT_AVG;  
RUN;
```

```
PROC AUTOREG DATA=BYCYCLE1;  
MODEL MYAVG=HDD CDD MOIS1-MOIS11  
      YEAR_NUM/ NLAG=(1 2 20) BACKSTEP;  
OUTPUT OUT=AVG  
P=FOREAVGA;  
RUN;
```

(2 W Systems Co., Inc.)

DEPENDENT VARIABLE = AVG_USE

Ordinary Least Squares Estimates

SSE	35,736	DFE	1,185
MSE	30.157	Root MSE	5.49
Rsqr	0.9895	DW	1.48

<u>Variable*</u>	<u>DF</u>	<u>B value</u>	<u>SE</u>	<u>t Ratio</u>	<u>Prob</u>
Intercept	1	24.237	1.368	17.72	0.001
HDD	1	0.171	0.002	108.45	0.001
CDD	1	0.015	0.002	6.15	0.001
YEAR_NUM	1	-1.131	0.115	-9.83	0.001

*MOIS variables are not presented here.

(2 W Systems Co., Inc.)

ESTIMATES OF AUTOCORRELATIONS

Lag	Covar	Corr	-2	1	0	1	2	3	4	5	6	7	8	9	1
0	29.78	1.000													
1	7.730	0.260													
2	12.89	0.433													
3	0.901	0.030													
4	4.435	0.149													
5	-0.94	-0.031													
6	4.84	0.163													
8	5.451	0.183													
9	1.833	0.062													
10	8.168	0.274													
19	-1.30	-0.044													
20	13.62	0.457													

DEPENDENT VARIABLE = AVG_USE

Yule-Walker Estimates

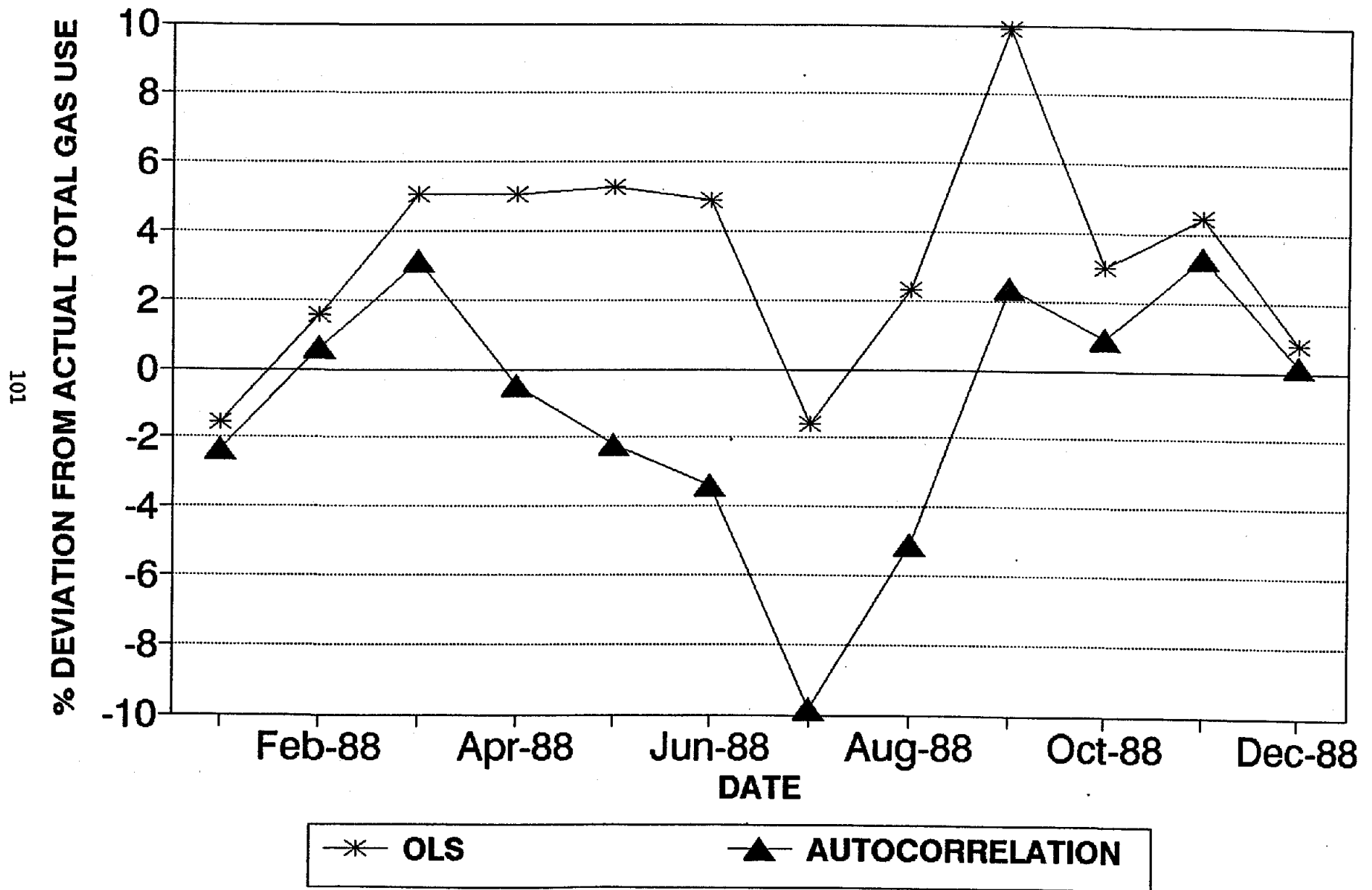
SSE	21,346	DFE	1,182
MSE	18.059	Root MSE	4.25
Rsqr	0.9909		

<u>Variable*</u>	<u>DF</u>	<u>B value</u>	<u>SE</u>	<u>t Ratio</u>	<u>Prob</u>
Intercept	1	23.194	2.690	8.62	0.001
HDD	1	0.164	0.002	96.72	0.001
CDD	1	0.016	0.003	5.68	0.001
YEAR_NUM	1	-0.677	0.120	-5.65	0.001
ρ_1	1	0.197	0.024	8.23	0.001
ρ_2	1	0.312	0.024	12.85	0.001
ρ_{20}	1	0.413	0.023	17.63	0.001

1988 FORECASTS OF RESIDENTIAL GAS USE COMPARING OLS/
AUTOCORRELATED MODELS BASED ON 1983-1987 DATA (CCF'S).

	MONTH	ACTUAL	OLS		AUTOCOR.	
			FORECAST	%DEV	FORECAST	%DEV
for	Jan-88	57,786,399	58,675,599	-1.54%	59,143,761	-2.35%
	Feb-88	45,915,904	45,205,443	1.55%	45,626,434	0.63%
	Mar-88	32,132,535	30,500,855	5.08%	31,118,088	3.16%
	Apr-88	16,198,409	15,376,409	5.07%	16,280,421	-0.51%
	May-88	9,769,418	9,253,889	5.28%	9,985,067	-2.21%
	Jun-88	9,486,663	9,021,326	4.91%	9,807,125	-3.38%
	Jul-88	8,848,898	8,990,375	-1.60%	9,719,153	-9.83%
	Aug-88	9,135,926	8,923,939	2.32%	9,601,997	-5.10%
	Sep-88	9,442,950	8,508,655	9.89%	9,218,229	2.38%
	Oct-88	16,452,915	15,958,226	3.01%	16,298,985	0.94%
	Nov-88	33,581,733	32,076,874	4.48%	32,474,523	3.30%
	Dec-88	47,387,784	47,014,713	0.79%	47,321,832	0.14%
	TOTAL	296,139,534	289,506,303	2.24%	296,595,615	-0.15%

COMPARING OLS & AUTOCORRELATION ESTIMATES IN % DEVIATION FROM ACTUAL BY MONTH.



CONCLUSION

Forecasting method presents:

- Better precision in forecasting than OLS
- Ease of implementation through both
 - 1) understandable statistical model
 - 2) available SAS statistical software

However:

- Cycles must be stable over time

(2 W Systems Co., Inc.)