COMMONWEALTH OF KENTUCKY
BEFORE THE PUBLIC SERVICE COMMISSION

IN THE MATTER OF:	)	
	)	
NOTICE OF ADJUSTMENT OF THE RATES OF	)	CASE NO. 2004-00103
KENTUCKY AMERICAN WATER COMPANY	)	
EFFECTIVE ON AND AFTER MAY 30, 2004	)	

DIRECT TESTIMONY OF DR. EDWARD L. SPITZNAGEL, JR.

April 30, 2004

KENTUCKY-AMERICAN WATER COMPANY
PSC CASE NO. 2004-00103

DIRECT TESTIMONY
EDWARD L. SPITZNAGEL, JR.

1. Q. Please state your name, business address, and employer.

A. My name is Edward L. Spitznagel, Jr., and my business address is Campus Box 1146, One Brookings Drive, St Louis, Missouri 63130. I am employed by Washington University.

2. Q. What is your present position?

A. I am Professor of Mathematics in the College of Arts and Sciences at Washington University. I also hold a joint appointment in the Division of Biostatistics of the Washington University School of Medicine.

3. Q. Please review your educational background and work experience.

A. I hold a Bachelor of Science, summa cum laude, in mathematics, awarded in 1962 by Xavier University, Cincinnati, Ohio. I hold a Master of Science (1963) and Ph.D. (1965) in mathematics awarded by the University of Chicago. I have served on the Faculty of Arts and Sciences of Washington University since 1969. I have held a joint appointment in the Division of Biostatistics since 1978. From 1965 to 1969 I was on the faculty of Northwestern University.

Attached to my testimony is Appendix A, which provides a more detailed listing of my education and qualifications in the area of mathematics and statistics.

4. Q. What is the purpose of your testimony in this case?

A. I have been employed by Kentucky American Water Company to make weather-normalized predictions of water utilization by residential and commercial customers for the period December 2004 to November 2005.

5. Q. What is weather normalization?

A. From one year to the next, variations in temperature and precipitation lead to changes in water consumption. More water will generally be used during hotter, drier periods. The regulatory question is how to reflect those weather-related differences when setting rates.

For ratemaking purposes, revenues need to be set to as "normal" a level as possible, factoring out the potential or actual results of unusual weather conditions. This can be accomplished by building statistical models that predict water utilization from meteorological data and other possible predictors. An estimate of future utilization can then be made by using a long-term average of meteorological data (since there is no better way to forecast next year's weather than as an average) and known values of the other predictors.

6. Q. What are examples of these other, non-meteorological predictors?

A. One is the year itself. Due to gradual introduction of water-conserving plumbing fixtures and appliances, use of water appears to be gradually declining over time for both residential and commercial customer classes.

Another is the month of the year. While water utilization increases during the warmer, drier summer months, analysis of variance shows that month as a categorical variable is a powerful predictor even after temperature and moisture have been included in the model.

7. Q. What model for water utilization did you employ?

A. In a previous case before this Commission, I screened a large number of candidate predictors by examining data from sixteen different operating companies in five states, Kentucky, Missouri, Ohio, Tennessee, and Virginia.

I used as candidate predictors only those variables that correlated consistently with utilization for most or all of these operating companies.

I then fitted the surviving candidates in a multivariate model to predict utilization for Kentucky American Water Company. I found that calendar month was a strong predictor even in the presence of heat and moisture variables. Therefore I included month as a categorical variable. With month included, I added drought severity index, temperature, and calendar year as potential numeric predictors. I found that temperature was not a useful predictor in the presence of the other variables, so from that point onward, I did not use it.

I updated the model for the present case by fitting the same predictors to utilization data from December 1996 through December 2003.

8. Q. This is a period of seven years plus one month. Why did you not employ the customary period of ten years?

A. Until near the end of 1996, Kentucky American Water Company billed all residential and many commercial customers on a quarterly basis until near the end of 1996, when they were converted to monthly billing. To fit the statistical model, it is necessary to have either monthly or quarterly utilization across the entire period. It is mathematically impossible to convert the older quarterly information to monthly utilization. The reason is that the quarterly utilization values are a smoothing or "blurring" of the finer monthly detail, and it is not possible to reverse this smoothing.

In the case filed in 2000, I converted monthly information from the more recent years into quarterly information by forming running sums of three consecutive months. I did this because the three years of monthly data alone would not have been enough to fit the model. We now have seven years plus one month of quarterly billing data with which to fit the utilization model. This is sufficient information to allow us to avoid using the blurred quarterly utilization values.

9. Q. What variables were found to predict utilization?

A. The analysis of variance results are given in Appendix B. In addition to month as a categorical variable, both drought index and year were predictive of utilization.

Since billing during a given month reflects utilization partially from that month plus partially from the preceding month, I explored a variety of lagged versions of drought index, as well as the unlagged drought index, to obtain best-fitting models, as measured by the R-square values. For both residential and commercial customers, the best version of drought index I found was the average of the drought index for the billing month and the drought index for the preceding month, which I will refer to as PDSI12. Along with month of the year and calendar year, it gave R-square values of 0.8760 for residential customers and 0.8745 for commercial customers.

10. Q. Is there any time-series dependence issue with the models described above?

A. There is no bad effect on the estimates themselves. They will be unbiased. However, all standard errors that appear in the analysis of variance tables are computed under the assumption of independence. This assumption is ordinarily incorrect with any kind of time series data and usually results in p-values somewhat smaller than they really should be.

This was a much more serious issue in my previous weather normalizations for KAWC, where the quarterly billing led naturally to serial dependence due to the overlapping utilization periods. In the present case, I again tested for its presence using a specialized autoregressive modeling procedure. These results are given in Appendix C. The model for residential utilization shows evidence of first-order autoregression, while the model for commercial consumption shows no such evidence. In both cases, year since 1990 remains a statistically significant predictor with negative slope, meaning there is evidence of a decrease in utilization over time for both residential and commercial classes.

11. Q. Is there a simple solution to making estimates in the presence of serial dependence?

A. Yes. In the actual weather normalization calculations, I use twelve different models, one for each month. Since the data values within each of these models are twelve months apart, there is little if any serial correlation within these twelve models. Standard errors computed in those models are correct without adjustment.

12. Q. What variables did you use in those models?

A. From the overall analysis of variance, there is evidence that the predictive power of the drought index varies as a function of month, being strongest in the summer months and virtually non-predictive in winter and early spring.

To maintain consistency with previous cases' analyses, for May through December, I used the drought index while for January through April I did not use it. I included year since 1990 (time trend) as a predictor.

These results are in Appendix D.

13. Q. Once you had estimated the coefficients in these monthly models, how did you project utilization for December 2004 through November 2005?

A. I put the coefficients from the monthly regressions into Excel spreadsheets, one for each of the two customer classes. I then calculated the monthly mean PDSI12 over the 30-year period from January 1974 to December 2003. These spreadsheets are given in Appendix E.

14. Q. Having inserted the mean drought severity indices in the spreadsheets, how did you proceed?

A. I then projected an average daily utilization for each month. This was calculated from the fitted model multiplied by 1000, then divided by the number of days from the preceding month. Using the days from the preceding month allows for the fact that bills in, for example, March include utilization from the latter part of February.

15. Q. What are your projections of daily utilization under average weather for the three customer classes?

A. For residential customers: 165.42 gallons / customer / day
For commercial customers: 1385.52 gallons / customer / day

16. Q. We notice that a slightly different estimate for commercial customers, 1384.01, has been used in KAWC computations. What is the explanation for the difference?

A. In originally examining my model screening computations, I misread the R-square for predicting commercial utilization from one-month lagged PDSI as being the largest of all, so I based my calculations on that model. I forwarded the resulting estimate of 1384.01 to the company for their calculations. In reviewing my computations for this testimony, I discovered my error and have made the corrected estimate of 1385.52.

The actual R-square for the one-month lagged PDSI is 0.8735, which is less than the R-square of 0.8745 for PDSI12. However, when I was reading my results from my monitor, I probably misread the 3 as an 8 and thought the 0.8735 was 0.8785.

In the corrected result, it is satisfying to see that the average of current and previous months' PDSI is the strongest predictor for both residential and commercial utilization. This makes theoretical sense because customers billed throughout a given month will, on average, be billed about half for utilization in the billing month and half for utilization in the previous month.

It is also satisfying to see that the error resulted in an estimate that was only one tenth of one percent different from the correct value. That is, the methodology is robust over moderate time-shifts in PDSI.

17. Q. Does this conclude your testimony?

A. Yes, it does.