Syracuse University

SURFACE

Economics - Faculty Scholarship

Maxwell School of Citizenship and Public Affairs

1998

State-Space versus Multiple Regression for Forecasting Urban Water Demand

R. Bruce Billings
University of Arizona

William C. Horrace
University of Arizona

Donald E. Agthe
Economic Consultant

Follow this and additional works at: https://surface.syr.edu/ecn

Part of the Economics Commons

Recommended Citation

State-Space versus Multiple Regression for Forecasting Urban Water Demand[1]

Discussion by:
James Zhou, San Diego Co. Water Authority
Kenneth A. Steele, San Diego Co. Water Authority
Richard C. Pyle, San Diego Co, Water Authority
Members, ASCE

The authors should be complimented for tackling a rather difficult problem of short-term urban water demand forecasting. The discussers would like to further explore some of their results and address the issues associated with forecasting methods for urban water use. The impact of weather variations in a semiarid area is one of key factors in forecasting short-term changes in urban water use.

The urban water demand is directly related to temperature and inversely related to precipitation. The authors' success in treatment of the weather factor, by independently employing temperature and precipitation and by taking their deviations from historic averages, confirms the same approach and findings in Zhou et al. (1996). In addition, the discussers have found in their study that water demand in a semiarid area is much more sensitive to precipitation than other weather factors such as air temperature, wind, and humidity.

Water is gradually becoming a commodity. But, unlike other consumer products, water is a basic necessity of life. Water also affects our quality of life. Therefore, the elasticity of a water price-use relationship must have a lower limit. In other words, although water use responds to price change, each household maintains a minimum consumption of water for their basic needs no matter what the price is.

Furthermore, the impact of price on water use is coupled with the income factor, which typically serves as a surrogate for household purchasing power. An increase in water price can be significant to a relatively low-income household, so that it responds with a reduction in water use. On the other hand, in a higher-income household the same price increase may result in no water saving actions but complaints via a public means. The impact—combining price, income, and relative purchasing capability—is essential to a forecasting model, but is not discussed in this technical note.

The discussers have found in their 1996 study that long-term water use is quite dependent on the strength of the local economy but not dependent on the price of water. Anecdotally, several observers here have noted that, while water usage may drop in the first few months of a price increase, it soon recovers. The authors' finding of a water price-use relationship in short-term water use forecasting tends to confirm this anecdotal evidence.

The discussers have also found in their study that water conservation efforts, along with an aggressive public education program on water saving techniques, would significantly affect overall water usage in the long run. The conservation efforts harden the urban water demand and further reduce the elasticity range of a water price-use relationship. There has been no quantitative analysis published on this subject. The discussers hope this subject will be included in future research by the scientific community. Lastly, it is noted that the state-space method was compared with an equally weighted, linear multiple regression method in the original paper. It would be interesting to know what the results would be for a comparison between the state-space method and an unequally weighted linear (or nonlinear) multiple regression model.

APPENDIX. REFERENCE

Zhou, J., Steele, K. A., and Pyle, R. C. (1996). ''Water use 'recession' in San Diego region.'' *Proc., ASCE North Am. Water and Envir. Congr.*,ASCE, Reston, Va., 2958–2963.

Discussion by:
Arun Kumar, Delhi College of Engineering
Vijay K. Minocha, Delhi College of Engineering

---

[1] March/April 1998, Vol. 124, No. 2, by R. Bruce Billings and Donald E. Agthe (Technical Note 13474).

The authors have presented a comparative study in forecasting monthly urban water supply using state-space and multiple regression models. The discussers feel that certain points raised in the study need further elaboration.

It is shown that the multiple regression model results in a poorer forecast when compared with a state-space model or even with a simple, seasonal, monthly mean model. The regression model in the forecasting model assigns zero value to $T^@$ and $R^@$, and the regression equation given in (4) reduces to

$$Q_i = a + b_3P^* + b_4D^* - b_5Y + d_iM_i \qquad (15)$$

The fact that (15) results in a poorer forecast than a monthly mean model $Q_i = Q_i$ simply suggests this: that perturbation of monthly demand from its respective monthly average is random, and incorporating economic parameters P^*, D^*, and Y is simply adding more noise to the model. This would have been obvious in the cross correlation structure between $(Qi - Qi)$ and the economic parameters.

The authors further state, "Independent variables whose estimated coefficients had an unexpected sign were dropped from individual runs while statistically insignificant coefficients with expected signs were retained." The procedure of selective elimination as followed by the authors is questionable. The unexpected sign will result because model parameters are being estimated by ordinary least square algorithm, wherein the sign constraint is not imposed. Possibly, a constrained least squares algorithm (CLS) might have helped in eliminating the unexpected signs, but at the cost of larger errors. The model parameter values of statistically estimated OLS cannot be attributed a physical interpretation.

The authors have obtained better results using the state-space model equations (5)–(8), simply highlighting the general state space model. It would be more desirable if the authors had given the exact structure of the model as used in the study.

The main advantage of state-space modeling over the regression model is that recursive algorithms are available for the sequential estimation of the state vector $\mathbf{X}_{t+1}$ (5) given the observation vector $\mathbf{Y}_t$. The Kalman gain factor $\mathbf{K}$ given in the (5) plays an important role in updating the state vector $\mathbf{X}_{t+1}$, and correct estimation of $\mathbf{K}$ is highly significant. The value of $\mathbf{K}$ is sensitive to the noise statistics in terms of modeling covariance Q and measurement error covariance R. It is not stated whether these parameters were kept constant or were updated with an additional data set.

The authors have observed that the state-space model has underestimated monthly forecast in fewer months than the regression and monthly average models. Based on the above observation, the authors have recommended that the forecaster who is more concerned about water shortages due to underestimating demand would probably prefer a state-space model.

The discussers believe it is only by chance that the state-space model has given an underestimated forecast less frequently. The parameter estimation procedure of a state-space model simply reduced the mean square error between the observed and the forecast values, irrespective of whether the forecast was higher or lower than the observed values. It is quite likely that in another data set the state model may yield more frequently underestimated forecasts than the other models. It is acceptable to say that the state-space model performs better, but to say that it underestimates on fewer occasions lacks statistical reasoning and scrutiny.

<div align="center">

Closure by:
R. Bruce Billings, University of Arizona
William C. Horrace, University of Arizona
Donald E. Agthe, Economic Consultant

</div>

We thank the editor for the opportunity to reply to the comments of Messrs. Zhou, Steele, and Pyle, and Kumar and Minocha. Our response plan is to address each of these comments in turn beginning with Zhou et al. (1999).

Zhou et al. (1999) suggest that water is becoming a commodity but is still unlike other consumer products, being a necessity of life. However, food, clothing, and housing are also necessities of life. Every commodity is whatever we make it to be through our institutions. The minimum amount of water for basic needs is very small—a few gallons per person per day for drinking, cooking, and washing. Other uses are dependent on social norms and economic conditions. A case study of Los Angeles revealed that only 11% of its water use was in meeting daily household necessity requirements (Johnson 1988).

Howe and Linaweaver (1967) carried out a landmark study showing that price is important to household water consumption decisions. Other notable studies in this early period that support the use of price as a rationing

device for water include Foster and Beattie (1979), and Gottleib (1963). Martin and Thomas (1986) compared water use among several relatively high-income desert climate cities with very different prices, and found an overall price elasticity of demand of 20.5. The prices and use levels ranged from $0.09 per kL for 595 L per capita per day in Phoenix, Arizona, to $14.04 per kL for 50 L per capita per day in Coober Pedy, South Australia. These results suggest that, in the long run, water use continues to respond to price changes, and that revenue will increase with price.

Both income and price are significant factors in the household water use decision. In Tucson, when prices were raised dramatically in the mid-1970s, water use fell in all areas of the city, regardless of income levels. Additionally, Agthe and Billings (1997) found that the price elasticity of demand did not vary much by the income level of individual households. Their estimated elasticities were 20.51 for low income households, 20.39 for middle income, and 20.54 for high income for a 10-year adjustment period. Zhou et al. suggest that price rises often set off shock waves that temporarily depress water use. While a widely publicized, large price increase may have this effect, the same outcome may occur because inflation erodes the real size of the increase over time. In the United States, this is especially relevant in the period inclusive of the decades of the 1970s and 1980s. Agthe et al. (1986) found that price elasticities of demand calculated for real prices in Tucson were considerably higher in the long run than in the short run, a result consistent with economic theory that hypothesizes long run substitutions of capital for water use.

Gottleib (1963), Morgan (1973), and Foster and Beattie (1979) found that rising incomes lead to increased water consumption. Large lawns, swimming pools, extra bathrooms in homes, and decorative uses of water tend to increase with income, consequently raising household water use. A study of Tucson revealed income elasticities of .31 and .36 depending on model chosen (Billings and Day 1989). The findings of Billings and Longstreth (unpublished, 1989) for Tucson contradict those of Zhou et al. for conservation campaigns. Two surveys—Billings and Longstreth (1989), and Dworkin (1981)—found that new residents of Tucson are much more concerned about water conservation than long-term residents. In Tucson, most conservation campaigns have a relatively short half-life, requiring constant reinforcement to remain effective.

Finally, Zhou et al. suggest that a comparison between a nonlinear multiple regression model and the state-space-model may be informative. Indeed, the spectrum of potential model specifications is unlimited and perhaps a more robust functional form could be found that nests our equally weighted scheme. However, given the excellent fit of the linear model (no R^2 less than .92), we felt that exploring myriad models was not justified. Furthermore, Billings (1982), using Tucson data, found no difference in R^2 between double-logarithmic and linear equations for a similar set of variables.

Kumar and Minocha were concerned with the statistical estimation of the state-space model. Billings and Agthe (1998) provided a background discussion of the state-space approach culminating with a general state-space specification in their equations (5)–(7). These are provided below. [We have substituted $\mathbf{Q}$ for $\mathbf{Y}$ to distinguish the endogenous variable (water demand) from the exogenous variable (income). Also, the exogenous effects matrix, $\mathbf{G}$, is s by j, not j by j as presented by Billings and Agthe (1998).

$$\mathbf{X}_{t+1} = \mathbf{FX}_t + \mathbf{GZ}_t + \mathbf{KE}_t$$

$$\mathbf{Q}_t = \mathbf{HX}_t + \mathbf{E}_t$$

$$\mathbf{Q}_t(m) = \mathbf{HX}_t$$

The software used is the ''Goodrich realization algorithm.'' The algorithm is too lengthy to describe in detail here; however, an excellent summary is provided in Goodrich (1992). Heuristically, the state vector $\mathbf{X}$ is an s-dimensional basis of the predictor space defined as the projection of forecasts of $\mathbf{Q}$ on past values of $\mathbf{Q}$ and $\mathbf{Z}$. $\mathbf{X}$ represents the first s significant canonical correlations associated with this projection, where significance is judged by the BIC principle. Therefore, $\mathbf{X}$ captures the s largest correlations between future realizations of $\mathbf{Q}$ and past values of $\mathbf{Q}$ and $\mathbf{Z}$.

One limitation of the general state-space model is its inability to accurately model data with a seasonal component. To handle this shortcoming, the Goodrich algorithm actually involves estimating two separate state-space models—a nonseasonal model and a seasonal model. The nonseasonal model is just the general model above. The seasonal model is the general model with $\mathbf{G} = 0$. In this case the state-space model can be recast as the Box-Jenkins model, which is amenable to seasonal estimation. The software used fits the raw data ($\mathbf{Q}$) in the seasonal model ($\mathbf{G} = 0$). It then deseasonalizes the $\mathbf{Q}$ and $\mathbf{Z}$ data using a seasonal state-space Kalman Filter (see Goodrich 1992, p. 182). The transformed $\mathbf{Q}$ and $\mathbf{Z}$ data are then used to fit the general model.

The software identified a single significant canonical correlation for both the seasonal and nonseasonal models, so the order of the system was $s = 1$ in both cases. Therefore, the general state-space system reduced to

$$X_{t+1} = A_1 X_t + G_1 P_t^* + G_2 D_t^* + G_3 Y_t + G_4 T_t + G_5 R_t + K_1 E_t,$$

$$Q_t = X_t + E_t,$$

$$Q_t(m) = X_t$$

where all elements are scalars. This resulted in estimation of only nine parameters (Complexity = 9) for the January 1981– December 1982 forecast period reported in Billings and Agthe (1998): A_1 and K_1 in the seasonal model, and A_1, K_1, G_1, . . ., $G5$ in the nonseasonal model. The results for the first forecast period are contained in Table 3 of Billings and Agthe (1998).

Kumar and Minocha observed that the monthly average forecast typically out-performs the multiple regression forecast. We believe that the within-sample fit of the multiple regressions was quite good (R2 of no less than 0.92 accompanied by reasonable Durbin-Watson tests) and they should not be dismissed as inaccurate based on their out-of-sample performance. ''A major disadvantage of these forecasting models is that the independent variables must themselves be forecast.'' Therefore, it is not unreasonable to speculate that much of the demand forecasting inaccuracy is due to poorly forecasted covariates, given the superior in-sample fit of the multiple regressions.

The practice of dropping covariates of incorrect sign is (admittedly) debatable. Our purpose was to highlight the relative merits of the state-space model vis-a`-vis a standard multiple regression model. To this end, we adopted a parsimonious modeling approach that would produce a good fit, given the usual economic sign restrictions, while avoiding a prolonged ''data-mining'' exercise. This approach was adopted for *both* the state space and the regression models, so comparisons are still valid. However, it is acknowledged that restricted least squares is a viable alternative in this instance.

It was certainly not our intent to leave the reader with a sense that state-space models *systematically* preclude underestimation, or that the regression models systematically cause it. We were merely implying that *for this particular data series* a forecaster might base his or her model selection decision on this particular criterion. Clearly, the propensity to over- or underestimate is an artifact of the data and not a salient feature of the theoretical models.

APPENDIX. REFERENCES

Agthe, D. E., and Billings, R. B. (1997). ''Equity and conservation pricing policy for a government-run water utility.'' *J. Water Supply Res. Technol.— Aqua*, 46(5), 252–260.

Agthe, D. E., Billings, R. B., Dobra, J., and Raffiee, K. (1986). ''A simultaneous demand model for block rates.'' *Water Resour. Res.*, 22(1), 1–4.

Billings, R. B. (1982). ''Specification of block rate price variables in demand models.'' *Land Economics*, 58(3), 386–394.

Billings, R. B., and Agthe, D. E. (1998). ''State-space versus multiple regression for forecasting urban water demand.'' *J. Water Resour. Plng. and Mgmt.*, 124(2), 113–117.

Billings, R. B., and Day, M. (1996). ''Demand management factors in residential water use: The southern Arizona experience.'' *J. Am. Water Works Assoc.*, 81(3), 58–64.

Dworkin, J. M. (1981). ''Identifying public preferences for water conservation measures.'' *Proc., Appl. Geography Conf.*, J. Frazier and B. Epstein, eds., 4, 95–106.

Foster, H., and Beattie, B. (1979). ''Urban residential demand for water in the United States.'' *Land Economics*, 55(1), 43–58.

Goodrich, R. L. (1992). *Applied statistical forecasting*. Business Forecasting Systems, Belmont, Mass.

Gottlieb, M. (1963). ''Urban domestic demand for water: A Kansas study.'' *Land Economics*, 39(2), 204–210.

Howe, C. W., and Linaweaver, F. P. Jr. (1967). ''The impact of price on residential water demand and its relation to system design and price structure.'' *Water Resour. Res.*, 3(1), 13–32.

Johnson, D. (1988). ''Down the Drain.'' *Los Angeles Times*, Feb. 2, Sect. 5, p. 1.

Martin, W. E., and Thomas, J. F. (1986). ''Policy relevance in studies of urban residential water demand.'' *Water Resour. Res.*, 22(13), 1735–1741.

Morgan, W. D. (1973). ''Residential water demand: The case from Microdata.''*Water Resour. Res.*, 9(4), 1065–1067.