Residential water demand
in California – Preliminary findings

by Céline Nauges

Attached are some rough notes about the estimation of the residential water demand, both in the short-run and in the long-run. The analysis covers four cities over the 1999-2002 period, including a total of 24 time periods of two-months each. The two-month period is chosen because the frequency of billing was monthly in some places and bimonthly in others.

We report in a short-run analysis (i.e. the estimation of the static demand function), and a long-run analysis (i.e. the estimation of the dynamic demand function), giving the short and long run price elasticity of demand. When estimating the dynamic demand function, GMM would have provided more efficient estimates than IV but Stata could not handle GMM with such a big dataset.

The data

Table 1: Basic summary statistics

	Davis	EMWD	Santa Rosa	Vista
Period covered by the data	Jan-Feb 1999 – Nov-Dec 2002	Jan-Feb 2002 – Nov-Dec 2002	Jan-Feb 1999 – Nov-Dec 2002	Jan-Feb 1999 – Nov-Dec 2002
Total number of observations	136,272	22,014	75,864	172,224
Total number of households	5,678	3,669	3,161	7,176
Average monthly water use (1,000 gallons)	12.93	11.35	8.41	12.95
Average monthly water use from November to April (1,000 gallons)	8.07	8.82	5.87	10.10
Average monthly water use from May to October (1,000 gallons)	17.79	13.88	10.95	15.80
Average marginal price (US$ per 1,000 gallons)	0.58	1.46	2.19	1.85
Average minimum temperature (Celsius degrees)	9.06	15.50	7.77	11.63
Average maximum	23.93	30.36	21.84	25.40

temperature (Celsius degrees)				
Average monthly precipitation (inches)	3.72	0.179	6.95	0.74

Note on prices: Water price in Santa Rosa includes a fixed service charge and a usage charge. The price shown in the above table corresponds to the usage charge (per 1,000 gallons). In EMWD, water charge varies depending on the zone of residence, that we do not observe. Thus we use the average marginal price charged over the 5 zones. All prices are real prices. Correction has been made using the Consumer Price Index (CPI) for fuels and utilities (all urban consumers, not seasonally adjusted, base date: January 1997). The CPI from the Los Angeles area has been used to deflate EMWD and Vista water prices. The CPI from the San Francisco area has been used to deflate Davis and Santa Rosa water prices.

The data gather information on households' water use in four Californian water districts: Davis, East Municipal Water District (EMWD), Santa Rosa, and Vista. For three of them, observations cover the 1999-2002 period. For EMWD, data were gathered over the 2002 period only. Frequency of billing varies from one place to the other: monthly in EMWD and Santa Rosa, bimonthly in Davis and Vista. For consistency of the data across cities, we define a period of observations as a two-months period (Period 1 refers to January-February 1999 and period 24 refers to November-December 2002). In all four utilities, water is charged through a flat rate, i.e., the price per unit is constant whatever total consumption. In what follows, consumption will be measured in 1,000 gallons and price in US$ per 1,000 gallons. The dataset also contains detailed weather data for each month, in particular minimum temperature, maximum temperature, and total precipitations.
We keep those households who, in each city, have been billed regularly over the period.

Average monthly water use varies from 8.41 (1,000 gallons) in Santa Rosa to 12.95 units in Vista. Water use varies across the seasons. Between November and April, water use varies from 5.87 units to 10.10 units, while average consumption increases between May and October (from 10.95 units to 17.79 units per month) when outdoor water use increases. Figure 1 shows monthly water use for each of the four districts over the 24 periods covered by our data. The seasonal pattern is clear.

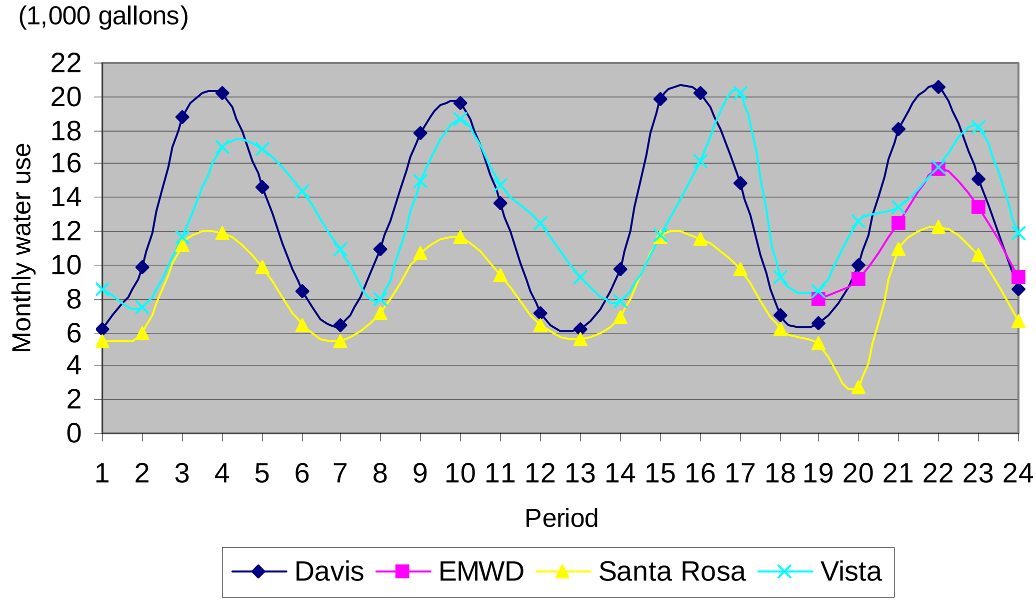

Figure 1: Average monthly water use over the period (Jan-Feb 1999 – Nov-Dec 2002)

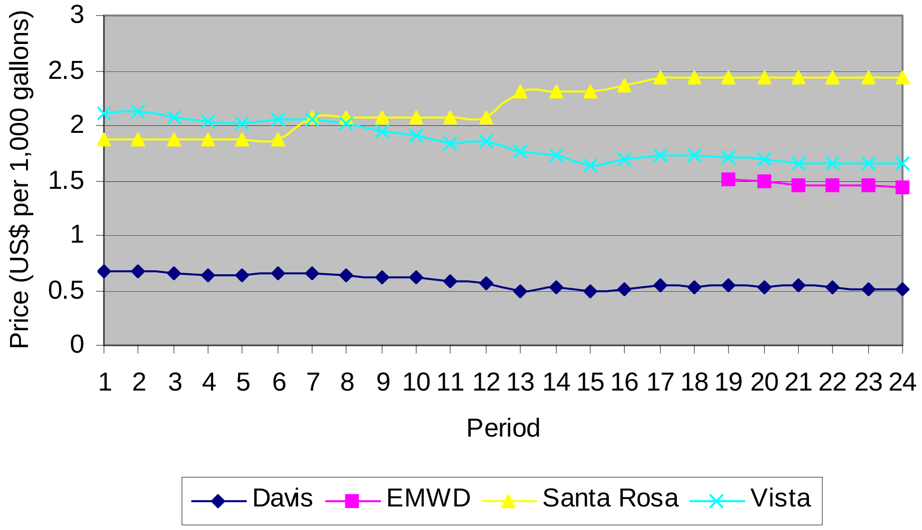

Figure 2: (Real) marginal price over the period (Jan-Feb 1999 – Nov-Dec 2002)

Real water prices have increased slightly in Santa Rosa while they decreased in Vista and Davis (nominal prices were constant in these two cities so real prices are declining).

Short run analysis of residential water demand

The "short-run" water demand function for the representative household is specified as follows:

$$\log(C_{it}) = a_0 + a_1 \log(P_{it}) + a_2 TMAX_{it} + a_3 TMIN_{it} + a_4 PREC_{it} + \mu_i + v_{it} \qquad (1)$$

where C_{it} is average monthly water use of household *i* in period *t* (a period covers two months), P_{it} is the price paid per unit (1,000 gallons) of water by household *i* in period *t*, $TMAX_{it}$ ($TMIN_{it}$) is the maximum (minimum) temperature at the closest weather station to household *i*'s residence in period *t*, and $PREC_{it}$ is the total amount of rainfall at the closest weather station to household *i*'s residence in period *t*. (a_0, a_1, a_2, a_3, a_4) is the vector of parameters to be estimated. The error term is decomposed into a household unobservable time-invariant effect (μ_i) and the usual idiosyncratic error term (v_{it}), the latter is assumed of mean 0 and constant variance. We also assume that the explanatory variables are not correlated with the idiosyncratic error term, i.e., $E(X_{it}v_{it}) = 0$ where $X_{it} = (P_{it}, TMAX_{it}, TMIN_{it}, PREC_{it})$, and that $E(X_{it}\mu_i) = 0$.

Since our data set does not include any household-specific information (household size, household income, house value, etc.), these characteristics are likely to be embedded in the unobservable term (μ_i). If μ_i is correlated with one or several explanatory variables, then usual estimation techniques (like Ordinary Least Squares) will give biased estimates. The most common estimation technique for panel data models such as equation (1) is the Within estimation technique, under which μ_i is assumed to be a fixed parameter (Baltagi, 2003). The Within estimation technique for panel data corresponds to the application of Ordinary Least Squares to equation (1) in which all variables have been deviated from their time means. This transformation eliminates all time-invariant components and in particular the household-specific parameter. The Within estimates are consistent whatever the correlation between the household-specific error term and the explanatory variables. Within estimates are shown in Table 2.

Table 2: Within estimates of the static water demand function

Dependent variable: log(consumption)	Estimated Coef.	Std. Err.	p-value
constant	1.680	0.0049	0.000
price (log)	-0.133	0.0087	0.000
maximum temperature	0.004	0.0002	0.000
minimum temperature	0.045	0.0002	0.000
total rainfall	-0.009	0.0002	0.000
Number of observations	406,374		
Number of households	19,684		

Since the dependent variable as well as the price are both in logarithms, the estimated coefficient of the price variable corresponds to the "short run" price elasticity of demand. Its value, -0.133, means that a 10% increase in price leads to a 1.33% decrease in water use. The weather variables have the expected signs.

The Fisher-test which tests that all household-specific effects are 0 is rejected at the 1% level. This result confirms that the Within estimation technique should be preferred to the simple OLS. Note that OLS overestimate the magnitude of the price elasticity of demand. In this case, price elasticity of demand as estimated by OLS is found equal to –0.25.

Long run analysis of residential water demand

The "long-run" analysis of water demand requires the specification of a dynamic demand equation, in which a lagged dependent variable $(C_{i,t-1})$ enters the vector of explanatory variables. The dynamic demand model is specified as follows:

$$\log(C_{it}) = a_0 + \delta \log(C_{i,t-1}) + a_1 \log(P_{it}) + a_2 TMAX_{it} + a_3 TMIN_{it} + a_4 PREC_{it} + \mu_i + v_{it} \qquad (2)$$

In this case, Within estimates have been proved to be inconsistent (Baltagi, 2003). The common approach is to use instruments, either an instrumental variables (IV) approach or a Generalized Methods of Moments (GMM) approach. The huge size of the sample makes it difficult to use the GMM approach so we consider here a more traditional IV approach. The procedure is the following. We first write equation (2) in first differences:

$$\log(C_{it})-\log(C_{i,t-1})=\delta\left[\log(C_{i,t-1})-\log(C_{i,t-2})\right]+a_1\left[\log(P_{it})-\log(P_{i,t-1})\right]+a_2\left[TMAX_{it}-TMAX_{i,t-1}\right]$$
$$+a_3\left[TMIN_{it}-TMIN_{i,t-1}\right]+a_4\left[PREC_{it}-PREC_{i,t-1}\right]+v_{it}-v_{i,t-1}.$$

This transformation eliminates the error term μ_i. We then estimate the model using IV techniques, where instruments are the dependent variable, lagged two to six periods. The model presented here yields the best fit to the data. Estimates are shown in Table 3.

Table 3: IV estimates of the dynamic water demand function

Dependent variable: log(consumption)	Estimated Coef.	Std. Err.	p-value
constant	-0.019	0.0014	0.000
lagged consumption (log)	0.653	0.0036	0.000
price (log)	-0.191	0.0366	0.000
maximum temperature	0.001	0.0002	0.000
total rainfall	-0.010	0.0002	0.000
Number of observations	316,120		
Number of households	16,015		

As expected the autoregressive parameter, δ, lies between 0 and 1. In model (2), the long run price elasticity is computed as follows:

$$\varepsilon=\frac{a_1}{1-\delta}$$

which, from Table (3), leads to a price elasticity equal to –0.55. This number means that a constant 10% increase of the water price (over several periods) would lead to a decrease in water use of 5.5%. As usually acknowledged in the literature, the long run price elasticity of demand is found larger in magnitude than the short run price elasticity.

References

Baltagi B.H., 2003. *Econometric analysis of panel data*. Second edition. John Wiley and Sons, LTD.