

ECON 120C -Fall 2003
PROBLEM SET 2: Suggested Solutions

PART I

Session 1

By substituting $f=f_0+f_1*D82+\dots$ and $g=g_0+\dots$ into Model 0 we obtain Model 1:

$$\ln(Q_t) = \beta_0 + \beta_1 * D82 + \beta_2 * D86 + \beta_3 * ED1 + \beta_4 * ED2 + \beta_5 * \ln(Y_t) + \beta_6 * \ln(P_t) + \beta_7 * \ln(Y_t) * D82 + \\ + \beta_8 * \ln(Y_t) * D86 + \beta_9 * \ln(Y_t) * ED1 + \beta_{10} * \ln(Y_t) * ED2 + \beta_{11} * \ln(P_t) * D82 + \beta_{12} * \ln(P_t) * D86 + \\ + \beta_{13} * \ln(P_t) * ED1 + \beta_{14} * \ln(P_t) * ED2 + u_t$$

In Step 2) we create the log of Q, Y and P (l_Q, l_Y and l_P)

After this, we create the interaction terms in steps 3) and 4):
 $ED1 \ln Y = ED1 * l_Y$ and so on.

Session 2

6) We estimate model 1 by OLS:

Model 1

VARIABLE	COEFFICIENT	STD.DEV.	T STAT.	2Prob(t > T)
const	-1.01861	3.72883	-0.273	0.788711
D82	29.5156	8.08426	3.651	0.002620 ***
D86	-34.0056	14.2597	-2.385	0.031784 **
ED1	-43.1059	37.8678	-1.138	0.274090
ED2	103.479	178.617	0.579	0.571568
l_P	0.0724658	0.525032	0.138	0.892189
l_Y	0.219833	0.492373	0.446	0.662081
D82lnY	-3.55868	0.974109	-3.653	0.002608 ***
D86lnY	3.98500	1.68554	2.364	0.033054 **
ED1lnY	5.59127	4.74536	1.178	0.258333
ED2lnY	-13.7657	22.0094	-0.625	0.541740
D82lnP	0.743501	0.380092	1.956	0.070707 *
D86lnP	-0.184133	0.322694	-0.571	0.577308
ED1lnP	-5.11257	3.93147	-1.300	0.214457
ED2lnP	15.7754	14.2776	1.105	0.287828

R-squared = 0.960432

n=29

7) LM test for AR(2):

We first generate the residuals ($u_t = \hat{u}_t$) of last regression. Then we obtain the first and second lags of the residuals ($u_{t-1} = u_t(-1)$, $u_{t-2} = u_t(-2)$).

Then we estimate the following regression by OLS:

$$\hat{u}_t = \beta_0 + \delta_1 * D82 + \delta_2 * D86 + \delta_3 * ED1 + \delta_4 * ED2 + \delta_5 * \ln(Y_t) + \delta_6 * \ln(P_t) + \delta_7 * \ln(Y_t) * D82 + \\ + \delta_8 * \ln(Y_t) * D86 + \delta_9 * \ln(Y_t) * ED1 + \delta_{10} * \ln(Y_t) * ED2 + \delta_{11} * \ln(P_t) * D82 + \delta_{12} * \ln(P_t) * D86 + \\ + \delta_{13} * \ln(P_t) * ED1 + \delta_{14} * \ln(P_t) * ED2 + \rho_1 * \hat{u}_{t-1} + \rho_2 * \hat{u}_{t-2} + \varepsilon_t$$

OLS estimates

	VARIABLE	COEFFICIENT	STD.DEV.	T STAT.	2Prob(t > T)
0)	const	-0.280270	3.16016	-0.089	0.931081
5)	ED1	4.98615	34.5135	0.144	0.888000
6)	ED2	-3.05463	176.783	-0.017	0.986554
7)	D82	-1.48621	8.16050	-0.182	0.859126
8)	D86	0.162859	13.3823	0.012	0.990530
10)	l_P	-0.200631	0.506718	-0.396	0.700464
11)	l_Y	0.0404982	0.415708	0.097	0.924318
12)	ED1lnP	1.20698	3.29384	0.366	0.721675
13)	ED2lnP	-2.62339	12.3857	-0.212	0.836511
14)	D82lnP	-0.160775	0.365437	-0.440	0.669328
15)	D86lnP	0.0379079	0.300945	0.126	0.902258
16)	D82lnY	0.189665	0.985058	0.193	0.851171
17)	D86lnY	-0.0232808	1.58526	-0.015	0.988572
18)	ED1lny	-0.667022	4.31092	-0.155	0.880114
19)	ED2lny	0.604636	21.6654	0.028	0.978285
21)	ut1	-0.462566	0.330753	-1.399	0.192195
22)	ut2	-0.676561	0.255894	-2.644	0.024568 **

R-squared = 0.478564
n=29

The null hypothesis of this test is $H_0: \rho_1 = \rho_2 = 0$

We calculate the statistic $LM = (n-2) \cdot R^2$, which is distributed under H_0 Chi-squared with 2 degrees of freedom:

$$LM = 27 \cdot 0.478564 = 12.92$$

If $LM > \chi^2(2 \text{ degrees of freedom})$ we reject the null hypothesis of no serial correlation.

Since $LM = 12.92 > \chi^2(2) = 9.21$, we reject the null hypothesis of no serial correlation.

Therefore, given that there is evidence of serial correlation in the error term of Model 1, OLS estimates of Model 1 will be unbiased, consistent but inefficient, and the hypothesis tests based on them are invalid.

Session 3

9) We estimate Model 1 by the Generalized Cochran-Orcutt for AR(2):

VARIABLE	COEFFICIENT	STD.DEV.	T STAT.	2Prob(t > T)
const	-0.509543	2.43228	-0.209	0.837579
ED1	-24.9078	25.4327	-0.979	0.346734
ED2	21.8420	124.768	0.175	0.863951
D82	17.5659	8.52683	2.060	0.061763 *
D86	-34.2587	14.4215	-2.376	0.035047 **
l_P	0.0776736	0.342071	0.227	0.824193
l_Y	0.154719	0.323102	0.479	0.640650
ED1lnP	-2.71806	2.89225	-0.940	0.365867
ED2lnP	5.53570	9.91380	0.558	0.586850
D82lnP	0.233051	0.349186	0.667	0.517139
D86lnP	0.163408	0.272289	0.600	0.559583
D82lnY	-2.10478	1.02841	-2.047	0.063246 *
D86lnY	3.96298	1.69490	2.338	0.037509 **
ED1lny	3.24379	3.23478	1.003	0.335758
ED2lny	-3.21744	15.4121	-0.209	0.838137

AR(2) coefficients

u_1	-0.248938	0.112973	-2.204	0.036992
u_2	-0.832248	0.111087	-7.492	< 0.00001

10) We select our final model by omitting the least significant variable of the model until all the coefficient estimates are significant at the 10% level. Note that we omit one variable at a time:

First, we omit ED2. The regression results of the new model are:

VARIABLE	COEFFICIENT	STD.DEV.	T STAT.	2Prob(t > T)
const	-0.472671	2.28915	-0.206	0.839613
ED1	-20.8333	8.30857	-2.507	0.026218 **
D82	15.4898	7.43483	2.083	0.057517 *
D86	-31.8104	13.1551	-2.418	0.031015 **
l_P	0.0998264	0.300446	0.332	0.744989
l_Y	0.151127	0.304823	0.496	0.628323
ED1lnP	-2.34258	2.36231	-0.992	0.339469
ED2lnP	3.83406	6.25127	0.613	0.550242
D82lnP	0.178458	0.303427	0.588	0.566516
D86lnP	0.222693	0.261153	0.853	0.409245
D82lnY	-1.85545	0.894674	-2.074	0.058517 *
D86lnY	3.66900	1.54467	2.375	0.033604 **
ED1lnY	2.72231	1.17243	2.322	0.037115 **
ED2lnY	-0.492988	0.562645	-0.876	0.396825

Now, the least significant variable is l_P. We omit it and get:

VARIABLE	COEFFICIENT	STD.DEV.	T STAT.	2Prob(t > T)
const	-1.18314	0.783765	-1.510	0.153391
l_Y	0.245545	0.105815	2.321	0.035923 **
ED1lnP	-1.64573	1.05765	-1.556	0.142015
ED2lnP	2.42998	4.46506	0.544	0.594851
D82lnP	0.119029	0.240484	0.495	0.628308
D86lnP	0.198030	0.241353	0.820	0.425682
D82lnY	-1.75584	0.820300	-2.140	0.050402 *
D86lnY	3.67641	1.49619	2.457	0.027659 **
ED1lnY	2.34855	0.319967	7.340	< 0.00001 ***
ED2lnY	-0.362681	0.390850	-0.928	0.369167
ED1	-18.2470	2.80982	-6.494	0.000014 ***
D82	14.6913	6.84462	2.146	0.049850 **
D86	-31.8544	12.7426	-2.500	0.025475 **

We now omit D82lnP:

VARIABLE	COEFFICIENT	STD.DEV.	T STAT.	2Prob(t > T)
const	-1.16959	0.743238	-1.574	0.136422
ED1	-18.4854	2.57156	-7.188	< 0.00001 ***
D82	11.3377	4.85865	2.334	0.033949 **
D86	-31.2476	12.9401	-2.415	0.028975 **
l_Y	0.244450	0.100288	2.437	0.027714 **
ED1lnP	-1.46117	0.936530	-1.560	0.139559
ED2lnP	1.80519	4.07477	0.443	0.664076
D86lnP	0.331580	0.190189	1.743	0.101714
D82lnY	-1.35025	0.570555	-2.367	0.031836 **
D86lnY	3.58608	1.52165	2.357	0.032452 **
ED1lnY	2.36236	0.297183	7.949	< 0.00001 ***
ED2lnY	-0.320929	0.360046	-0.891	0.386813

Finally, we omit ED2lnP and obtain our final model, in which all the coefficients are significant at the 10% level:

VARIABLE	COEFFICIENT	STD.DEV.	T STAT.	2Prob(t > T)
const	-1.27720	0.695232	-1.837	0.084844 *
ED1	-18.4933	2.52234	-7.332	< 0.00001 ***
D82	10.0501	4.41535	2.276	0.036932 **
D86	-28.2211	12.1556	-2.322	0.033774 **
l_Y	0.258534	0.0942172	2.744	0.014410 **
ED1lnP	-1.05286	0.217701	-4.836	0.000182 ***
D86lnP	0.347920	0.186136	1.869	0.080016 *
D82lnY	-1.19917	0.518608	-2.312	0.034404 **
D86lnY	3.23087	1.43040	2.259	0.038217 **
ED1lnY	2.32428	0.273513	8.498	< 0.00001 ***
ED2lnY	-0.166325	0.0574408	-2.896	0.010537 **

AR coefficients:

u_1	-0.175295	0.0967261	-1.812	0.081972
u_2	-0.919021	0.0950578	-9.668	< 0.00001

ESS = 0.00586942
R-squared = 0.979191
Adjusted R-squared = 0.966185
Durbin-Watson = 3.10301

Elasticities using coefficients of the final model:

a) Regression models for the three periods:

1st period:

$$\ln(Q_t) = -1.227 - 18.493 \cdot ED1 + 0.258 \cdot \ln(Y_t) + 2.324 \cdot \ln(Y_t) \cdot ED1 - 0.166 \cdot \ln(Y_t) \cdot ED2 - 1.053 \cdot \ln(P_t) \cdot ED1 + \hat{u}_t$$

2nd period:

$$\ln(Q_t) = 8.823 - 18.493 \cdot ED1 - 0.941 \cdot \ln(Y_t) + 2.324 \cdot \ln(Y_t) \cdot ED1 - 0.166 \cdot \ln(Y_t) \cdot ED2 - 1.053 \cdot \ln(P_t) \cdot ED1 + \hat{u}_t$$

3rd period:

$$\ln(Q_t) = -19.398 - 18.493 \cdot ED1 + 2.290 \cdot \ln(Y_t) + 2.324 \cdot \ln(Y_t) \cdot ED1 - 0.166 \cdot \ln(Y_t) \cdot ED2 + 0.348 \cdot \ln(P_t) - 1.053 \cdot \ln(P_t) \cdot ED1 + \hat{u}_t$$

b) Effect of ED1 and ED2 on the elasticities:

1st period:

$$\partial \ln(Q_t) / \partial \ln(Y_t) = 0.258 + 2.324 \cdot ED1 - 0.166 \cdot ED2$$

$$\partial \ln(Q_t) / \partial \ln(P_t) = -1.053 \cdot ED1$$

2nd period:

$$\partial \ln(Q_t) / \partial \ln(Y_t) = -0.941 + 2.324 \cdot ED1 - 0.166 \cdot ED2$$

$$\partial \ln(Q_t) / \partial \ln(P_t) = -1.053 \cdot ED1$$

3rd period:

$$\partial \ln(Q_t) / \partial \ln(Y_t) = 2.290 + 2.324 \cdot ED1 - 0.166 \cdot ED2$$

$$\partial \ln(Q_t) / \partial \ln(P_t) = 0.348 - 1.053 \cdot ED1$$

The effect of ED1 and ED2 on income and price elasticities is the same for the three periods. What changes across periods is the constant term in the above expressions (λ_0).

As we can see, an increase in ED1 increases income elasticity and decreases price elasticity. Specifically, an increase of 1 percentage point in middle and high school enrollment increases income elasticity by 0.0232 and reduces price elasticity by 0.0105.

An increase in ED2 only affects income elasticity: An increase of 1 percentage point in ED2 reduces income elasticity by 0.0017.

c) Using $ED_1=0.27$ and $ED_2=0.06$ in the above expressions we obtain the elasticity estimates for each period:

Period	Income elasticity	Price Elasticity
1960-1981	0.876	-0.284
1982-1985	-0.323	-0.284
1986-1988	2.917	0.064

According to these estimates, demand is income inelastic for the first 2 periods and it is elastic for the last period, whereas it is always price inelastic.