APPENDIX C: DETAILED RESULTS OF PHASE II

In this appendix, we present all the results of the runs carried out in Phase II. Table 1 presents the coefficients and related statistics. The variables are shown in the order in which they appear in the equations. Table 2 provides definitions and some comments on the variables.

Table 1

Coefficients and Their Associated F Statistics[1]

Independent Variables[3]	Runs[2] 1	2	3	4	5	6	7	8	9
Constant	2.836	1.187	4.394	3.011	0.106	0.147	2.852	2.789	3.002
LSZLN	0.093 (15.888)	0.119 (33.918)	- 0.074 (0.475)	0.053 (4.038)	0.052 (2.571)	- 0.012 (0.074)	0.111 (20.297)	0.090 (15.304)	0.098 (18.070)
AGE	- 0.002 (8.534)	- 0.002 (5.727)	- 0.005 (3.510)	- 0.002 (4.665)	- 0.000 (0.000)	0.001 (0.969)	- 0.002 (5.034)	- 0.002 (6.964)	- 0.002 (6.650)
COND	- 0.176 (22.482)	- 0.175 (25.865)	- 0.216 (2.648)	- 0.148 (13.172)	- 0.114 (4.835)	- 0.150 (2.323)	- 0.190 (23.737)	- 0.171 (21.645)	- 0.184 (24.540)
BMT	- 0.030 (3.774)	- 0.050 (11.781)	- 0.053 (1.231)	- 0.041 (5.166)	- 0.020 (0.595)	- 0.003 (0.007)	- 0.032 (4.317)	- 0.033 (4.428	- 0.033 (4.402)
AIR	- 0.046 (10.026)	- 0.048 (13.227)	- 0.019 (0.138)	- 0.052 (9.272)	- 0.057 (8.341)	- 0.071 (5.930)	- 0.046 (9.867)	- 0.045 (9.225)	- 0.045 (9.490)
GRGD	0.025 (0.089)	0.045 (0.447)	--	0.056 (0.328)	0.269 (6.153)	0.342 (7.614)	0.032 (0.135)	0.035 (0.170)	0.002 (0.001)

[1]Coefficients are not calculated for dummy variables for which the sample used has no observations. In addition, one dummy variable is omitted in each set, as shown by blanks in the Table.

[2]Dependent variable for each run was PVLN (defined on p. C-17).
Run 1 uses Model 1, enlarged sample (post-1974).
Run 2 uses Model 1, original sample (post-1974).
Run 3 uses Model 1, new sample (enlarged minus original sample, post-1974).
Run 4 uses Model 1, enlarged sample (post-1974).
Run 5 uses enlarged sample (pre-1974).
Run 6 uses original sample (pre-1974).
Run 7 uses Model 2, enlarged sample (post-1974).
Run 8 uses Model 3, enlarged sample (post-1974).
Run 9 uses Model 4, enlarged sample (post-1974).

[3]Independent variables are defined on pp. C-15 to C-18.

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
FPL	- 0.032 (3.825)	- 0.050 (10.165)	0.090 (3.512)	- 0.210 (1.26)	- 0.055 (5.139)	- 0.072 (2.806)	- 0.030 (5.003)	- 0.036 (5.106)	- 0.034 (4.41)
BTR	0.129 (67.021)	0.062 (14.185)	0.203 (19.716)	0.137 (57.534)	0.111 (17.938)	0.118 (7.179)	0.128 (66.654)	0.131 (69.175)	0.127 (64.87)
HRELN	0.101 (15.158)	0.278 (52.293)	0.007 (0.026)	0.116 (14.171)	0.281 (28.003)	0.255 (12.058)	0.107 (16.970)	0.106 (16.812)	0.104 (16.16)
GRGA	--	--	--	--	--	--	--	--	--
GRGB	0.061 (4.594)	0.072 (6.455)	0.009 (0.011)	0.092 (7.609)	0.081 (6.689)	0.151 (6.521)	0.058 (4.153)	0.063 (4.722)	0.056 (0.380)
GRGC	0.064 (4.514)	0.073 (6.135)	- 0.002 (0.001)	0.093 (6.919)	0.112 (9.708)	0.180 (7.455)	0.061 (4.015)	0.061 (4.014)	0.056 (3.49)
OTBN	0.044 (3.190)	0.051 (5.571)	0.076 (0.412)	0.053 (3.250)	0.058 (4.137)	0.048 (1.068)	0.041 (2.660)	0.043 (2.880)	0.043 2.951
DCBDLN	0.067 (1.586)	0.195 (5.184)	0.043 (.023)	0.066 (1.499)	0.006 (0.031)	0.004 (.039)	0.006 (0.014)	0.055 (2.087)	0.007 (0.022)

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
DS1	- 0.163 (2.247)	- 0.179 (3.762)	--	- 0.123 (0.913)	--	--	- 0.157 (2.106)	- 0.148 1.807	- 0.188 (2.89)
DS2	0.002 (0.001)	0.002 (0.002)	- 0.241 (1.173)	--	0.031 (0.167)	0.026 (0.086)	- 0.034 (0.308)	- 0.009 (0.023)	- 0.025 (0.167)
DS3	0.140 (4.940)	--	0.348 (7.245)	--	- 0.042 (0.138)	--	0.113 (3.354)	0.147 (6.260)	0.142 (5.94)
DS4	--	--	--	--	0.026 (0.017)	--	--	--	--
DSD	- 0.003 (0.001)	--	- 0.418 (3.293)	--	--	--	0.097 (0.911)	0.078 (0.468)	0.025 (0.034)
DS5	- 0.136 (1.601)	0.273 (7.47)	- 0.282 (0.597)	--	- 0.089 (1.203)	- 0.114 (0.621)	0.106 (1.021)	0.118 1.240	0.103 (.942)
ZNB	- 0.046 (0.446)	- 0.46 (0.716)	--	0.071 (0.799)	0.211 (3.652)	0.383 (8.219)	- 0.021 (0.096)	- 0.021 (0.062)	- 0.077 (1.121)

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
ZNF	- 0.137 (18.826)	- 0.062 (5.015)	--	- 0.141 (14.481)	- 0.037 (0.424)	- 0.049 (0.359)	- 0.073 (2.704)	- 0.132 (13.992)	- 0.121 (11.582)
ZNP	0.058 (1.250)	--	--	--	- 0.072 (1.359)	--	0.003 (0.398)	0.185 (0.169)	0.031 (0.463)
ZNQ	0.092 (1.075)	--	--	--	- 0.004 (0.001)	--	0.022 (0.089)	- 0.353 (0.002)	- 0.024 (0.097)
ZNA	- 0.005 (0.008)	0.003 (0.004)	0.782 (4.088)	0.029 (0.194)	--	--	0.000 (0.000)	0.008 (0.016)	- 0.026 (0.180)
ZNE	- 0.053 (6.665)	- 0.049 (5.607)	0.205 (0.909)	- 0.042 (3.256)	- 0.029 (0.827)	0.031 (0.318)	- 0.006 (0.049)	- 0.047 (3.699)	- 0.037 (2.054)
ZNJ	- 0.034 (0.413)	- 0.048 (0.798)	0.165 (0.195)	0.055 (0.561)	- 0.228 (5.657)	- 0.208 (3.800)	- 0.066 (1.284)	- 0.049 0.649	- 0.076 (1.645)
ZNO	--	--	--	--	- 0.097 (0.353)	- 0.230 (0.909)	--	--	--

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
PDEN	- 0.004 (0.006)	0.064 (.210)	0.239 (0.210)	- 0.014 (0.045)	- 0.073 (1.490)	0.039 (0.176)	- 0.042 (0.625)	- 0.069 (1.184)	0.053 (0.928)
CLE	- 0.490 (40.489)	- 0.386 (25.736)	- 0.885 (6.041)	- 0.479 (27.515)	- 0.310 (4.828)	- 0.765 (7.789)	- 0.492 (40.391)	- 0.482 (37.417)	- 0.505 (42.48)
CLF	- 0.370 (33.635)	- 0.281 (26.068)	- 0.563 (2.441)	- 0.363 (23.177)	0.066 (0.380)	- 0.472 (6.267)	- 0.379 (35.814)	- 0.370 (31.641)	- 0.378 (2.35)
CLG	- 0.098 (2.786)	- 0.055 (1.127)	- 0.149 (0.214)	- 0.169 (5.868)	0.037 (0.098)	- 0.203 (1.533)	- 0.105 (3.189)	- 0.087 (2.139)	- 0.091 (2.35)
CLH	--	--	--	--	--	--	--	--	--
PLG	0.148 (4.281)	0.202 (7.272)	0.087 (0.291)	0.162 (3.703)	- 0.021 (0.043)	- 0.072 (0.459)	0.158 (4.514)	0.176 (6.153)	1.70 (5.72)
PLV	0.140 (16.314)	0.135 (22.388)	0.426 (1.309)	0.147 (12.772)	- 0.020 (0.167)	- 0.025 (0.196)	0.145 (17.608)	0.146 (17.529)	0.153 (19.33)
DPWM	0.059 (3.101)	0.122 (10.334)	- 0.225 (2.609)	0.039 (1.122)	0.048 (0.816)	0.027 (.063)	0.027 (0.644)	0.049 (2.117)	0.044 (1.74)

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
DACM	- 0.011 (0.081)	- 0.061 (1.720)	- 0.003 (0.000)	- 0.021 (0.217)	- 0.060 (2.012)	- 0.108 (1.075)	- 0.004 (0.023)	- 0.006 (0.037)	0.037 (0.794)
DD1	--	--	--	--	--	--	--	--	--
DD2	- 0.025 (0.593)	0.004 (0.019)	- 0.596 (1.865)	- 0.013 (0.112)	--	--	--	--	--
DD3	0.056 (2.098)	- 0.005 (0.029)	- 0.364 (2.322)	- 0.082 (3.212)	0.061 (3.764)	0.058 (2.317)	--	--	--
DD4	0.009 (0.044)	0.086 (5.131)	- 0.767 (5.242)	0.035 (0.527)	0.045 (1.347)	0.102 (1.337)	--	--	--
DD5	- 0.013 (0.076)	0.818 (3.533)	- 0.051 (0.077)	0.037 (0.005)	0.055 (1.521)	0.144 (1.442)	--	--	--
DD6	- 0.053 (0.707)	0.220 (2.448)	- 0.015 (0.009)	- 0.044 (0.438)	0.091 (2.052)	--	--	--	--
DD7	- 0.071 (1.219)	--	0.096 (0.429)	0.006 (0.007)	0.010 (3.898)	--	--	--	--

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
DD8	- 0.174 (2.248)	--	--	- 0.122 (1.018)	--	--	--	--	--
QD4/QDD4[4]	--	--	--	--	--	--	--	--	- 0.133 (1.91)
QD1/QDD1	--	--	--	--	--	--	--	0.022 (0.096)	- 0.065 (2.17)
QD2/QDD2	--	--	--	--	--	--	--	0.024 (0.096)	- 0.044 (3.10)
QD3/QDD3	--	--	--	--	--	--	--	0.054 (0.400)	- 0.028 (1.41)
SDA	0.481 (57.442)	0.504 (41.599)	0.679 (26.041)	0.493 (42.520)	--	--	0.510 (64.918)	0.499 (61.440)	0.498 (61.735)
SDA1	0.464 (16.434)	--	0.542 (5.520)	0.520 (14.758)	--	--	0.434 (13.851)	0.451 (15.818)	0.452 (15.934)
SDA2	0.482 (31.670)	--	0.689 (25.540)	0.471 (21.770)	--	--	0.469 (30.581)	0.483 (32.066)	0.485 (30.985)

[4]QD1-4 are used in Run 8. QDD1-4 are used in Run 9.

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
SDB	0.498 (92.212)	0.499 (78.409)	0.692 (22.259)	0.514 (69.768)	--	--	0.522 (103.514)	0.511 (96.702)	0.518 (98.882)
SDC	0.512 (78.066)	0.490 (59.522)	0.756 (10.853)	0.524 (57.775)	--	--	0.525 (83.414)	0.524 (83.054)	0.528 (84.387)
SDD	0.416 (49.164)	0.442 (46.326)	0.552 (15.924)	0.431 (37.667)	--	--	0.417 (49.612)	0.414 (50.073)	0.424 (52.096)
SDE	0.383 (52.878)	0.376 (43.326)	0.507 (8.187)	0.396 (40.117)	--	--	0.405 (60.433)	0.394 (56.870)	0.409 (59.823)
SDF	0.505 (83.678)	0.534 (83.388)	0.544 (10.449)	0.506 (59.284)	--	--	0.520 (89.761)	0.514 (87.715)	0.524 (83.388)
SDG	0.495 (77.831)	0.481 (68.714)	0.774 (9.580)	0.493 (54.798)	--	--	0.509 (83.289)	0.501 (80.440)	0.513 (83.292)
SDH	0.456 (40.112)	0.476 (44.747)	0.331 (2.235)	0.455 (28.485)	--	--	0.478 (44.258)	0.470 (42.685)	0.475 (43.632)

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
SDI	0.469 (89.880)	0.435 (57.975)	0.658 (34.624)	0.472 (64.039)	--	--	0.490 (98.714)	0.478 (92.767)	0.486 (93.040)
SDJ	0.444 (98.654)	0.454 (69.177)	0.557 (38.768	0.443 (69.442)	--	--	0.462 (108.572)	0.450 (103.078	0.460 (105.938
SDK	0.315 (52.118)	0.329 (37.142)	0.359 (17.323)	0.276 (28.449)	--	--	0.328 (58.942)	0.325 (57.254)	0.331 (58.832)
SDL	0.259 (31.800)	0.290 (29.209)	0.301 (9.038)	0.267 (23.993)	--	--	0.267 (34.802)	0.268 (34.588)	0.274 (35.802)
SDM	0.255 (34.536)	0.271 (26.739)	0.274 (8.691)	0.245 (22.511)	--	--	0.260 (36.450)	0.263 (36.450)	0.269 (37.818)
SDN	0.212 (21.587)	0.227 (18.965)	0.376 (7.770)	0.214 (15.618)	--	--	0.223 (24.297)	0.223 (24.014)	0.227 (24.409)
SDO	0.201 (18.470)	0.197 (14.041)	0.355 (6.429)	0.206 (13.868)	--	--	0.207 (19.815)	0.210 (20.203)	0.217 (21.574)

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
SDP	0.189 (12.491)	0.165 (7.636)	0.274 (5.103)	0.192 (9.098)	--	--	0.211 (15.788)	0.195 (13.430)	0.202 (14.455)
SDQ	0.157 (10.464)	0.174 (10.125)	0.199 (2.311)	0.158 (7.403)	--	--	0.168 (12.182)	0.167 (11.869)	0.172 (12.500)
SDR	0.141 (10.274)	0.140 (7.593)	0.245 (2.717)	0.144 (7.499)	--	--	0.151 (11.973)	0.148 (11.216)	0.152 (11.894)
SDS	0.131 (6.196)	0.145 (6.602)	0.012 (0.003)	0.135 (4.621)	--	--	0.147 (7.977)	0.143 (7.501)	0.149 (8.025)
SDT	0.062 (1.357)	0.080 (2.045	0.127 (0.289)	0.063 (1.007)	--	--	0.094 (3.180)	0.080 (2.276)	0.085 (2.586
SDU	0.110 (5.502)	0.128 (5.796)	0.136 (1.368)	0.118 (4.561)	--	--	0.130 (7.985)	0.124 (7.100)	0.128 (7.614)
SDV	0.084 (3.142)	0.079 (2.229)	0.184 (1.050)	0.076 (1.841)	--	--	0.095 (4.135)	0.095 (4.093)	0.101 (4.599)

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
SDW	0.071 (1.957)	0.052 (0.850)	0.312 (5.735)	0.088 (2.088)	--	--	0.092 (3.308)	0.082 (2.601)	0.086 (2.920)
SDX	0.104 (3.428)	0.113 (3.368	0.193 (1.856)	0.128 (3.681)	--	--	0.115 (4.241)	0.119 (4.429)	0.125 (4.762)
SDY	0.050 (0.688)	- 0.018 (0.071)	0.152 (1.313)	0.047 (0.428)	--	--	0.068 (1.307)	0.053 (0.784)	0.063 (1.096)
SDZ	0.107 (2.951)	0.108 (2.455)	0.258 (3.297)	0.115 (2.434)	--	--	0.114 (3.406)	0.119 (3.685)	0.122 (3.875)
SDAA	0.089 (2.200)	0.164 (6.985)	- 0.022 (0.016)	0.098 (1.904)	--	--	0.103 (2.975)	0.098 (2.677)	0.106 (3.124)
SDBB	0.034 (0.222)	0.058 (0.595)	0.214 (1.537)	0.018 (0.044)	--	--	0.043 (0.358)	0.043 (0.361)	0.051 (0.519)
SDCC	0.082 (2.269)	- 0.068 (1.356)	0.174 (1.536)	- 0.001 (0.000)	--	--	0.095 (3.108)	0.091 (2.828)	0.097 (3.226)

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
SDDD	- 0.020 (0.135)	- 0.002 (0.001	0.130 (1.223)	- 0.040 (0.427)	--	--	- 0.012 (0.053)	- 0.011 (0.044)	- 0.004 (0.005)
SDEE	0.038 (0.595)	0.042 (0.591)	0.128 (1.228)	0.028 (0.229)	--	--	0.057 (1.388)	0.052 (1.166)	0.062 (1.625)
SDFF	--	--	--	--	--	--	--	--	--
SDTRND[5]	--	--	--	--	0.915 (265.864)	1.034 (200.752)	--	--	--
WKN	--	--	--	--	--	--	- 0.104 (5.030)	--	--
WSW74	--	--	--	--	--	--	- 0.031 (0.389)	--	--

[5]A sales trend was computed for sample 2 only in Phase II because of time and budgetary constraints. The sales trend is simply an aggregation of the coefficients of individual dummy variables computed in an earlier run, and, therefore, equivalent to the series of individual dummmy variables.

Table 1 (continued)

Coefficients and Their Associated F Statistics

Independent Variables	Runs 1	2	3	4	5	6	7	8	9
WDLB76	--	--	--	--	--	--	- 0.037 (0.087)	--	--
WDLA76	--	--	--	--	--	--	- 0.124 (3.281)	--	--
WM	--	--	--	--	--	--	- 0.020 (0.382)	--	--
M75Z1	--	--	--	--	--	--	0.074 (1.935)	--	--
WW	--	--	--	--	--	--	--	--	--

Summary Statistics

	1	2	3	4	5	6	7	8	9
$\bar{R}^2$	0.851	0.905	0.844	0.793	0.811	0.815	0.852	0.849	0.850
F	43.564	57.812	8.475	32.356	38.948	25.263	44.618	45.683	45.281
Sample Size	630	460	170	630	383	216	630	630	630

Table 2

Definitions of Variables and Comments

CODE	VARIABLE	DESCRIPTION/COMMENTS
AIR[1]	air conditioning	yes = 0, no = 1
AGE	age of house when sold	year built-sales date
BMT	basement	yes = 0, no = 1
BTR	bathroom	number of bathrooms
COND	condition	good (G) = 1, fair (F) = 2, poor (P) = 3
CLE[1]	classification for below average construction	yes = 1, no = 0
CLF[1]	classification for average construction	yes = 1, no = 0
CLG[1]	classification for above average construction	yes = 1, no = 0
CLH[1]	classification for good grade size and construction	yes = 1, no = 0
DACM	distance to highway access	distance measured in miles
DD1 to DD8[1]	distance from the waste dump in .5 mile dummies	DD1 = first 1/2 mile[2] DD2 = second 1/2 mile, etc.
DS1, 2, 3, 4, 5, D[3]	distance to schools	within 1/4 mile = .25 between 1/4 and 1/2 mile = .5 over 1/2 mile = 1
DPWM	distance to highway	distance measured in miles

Table 2 (continued)

Definitions of Variables and Comments

CODE	VARIABLE	DESCRIPTION/COMMENTS
DCBD	distance from central business district	
DCBDLN	natural log of DCBD ÷ distance to central business district	distance measured in miles
FPL[1]	fireplace	yes = 0, no = 1
GRGA[1]	no garage	yes = 1, no = 0
GRGB[1]	1 car garage	yes = 1, no = 0
GRGC[1]	2 car garage	yes = 1, no = 0
GRCD[1]	3 car garage	yes = 1, no = 0
HRELN[4]	natural log of HARE = total floor area	measured in square feet
LSZLN[4]	natural log of LSZ = lot size	frontage x dept, additional acres are added Total converted to square feet (1,000 ft^2)
OTBN	number of outbuildings	
PLG[1]	in-ground pool	yes = 1, no pool = 0
PLV[1]	vinyl pool	yes = 1, no pool = 0
PDEN	average number of rooms per person in enumeration district	
PVLN	natural log of PV = property values	sales price in $1,000

Table 2 (continued)

Definitions of Variables and Comments

CODE	VARIABLE	DESCRIPTION/COMMENTS
QD1-4	directional dummies	QD1 Northeast quadrant Yes = 1 No = 0 QD2 Southeast quadrant Yes = 1 No = 0 QD3 Southwest quadrant Yes = 1 No = 0 QD4 Northwest quadrant Yes = 1 No = 0
QDD1-4	directional distance variables (continuous)	QDD1 = distance from waste dump, miles if QD1 = 1 QDD2 = distance from waste dump, miles if QD2 = 1 QDD3 = distance from waste dump, miles if QD3 = 1 QDD4 = distance from waste dump, miles if QD4 = 1
SDFF to SDA2 (1974-1982)	sales dummies	time trended according to quarter and year beginning with second quarter 1982. SDA2; series runs SDA2 SDA, SDB...SDZ, SDAA, SDBB...SDFF, which is the first quarter of 1974
SDTRND	sales dummies trend for sample 2	$\sum_{i=A2}^{SDFF} = SD_i \times \text{coefficient } SD_i$
WKN	No information	Yes = 1 No = 0
WW	Wells determined not to be contaminated	Yes = 1 No = 0
WM	Municipal Water only	Yes = 1 No = 0

Table 2 (continued)

Definitions of Variables and Comments

CODE	VARIABLE	DESCRIPTION/COMMENTS		
WSW74	Wells contamined in 1974 which switched to municipal water	Yes = 1 No = 0		
M75Z1	Built in 1975 in contamination Zone 1	Yes = 1 No = 0		
WDLB76	Wells in Dugan's Lane area before 1976	Yes = 1 No = 0		
WDLA76	Wells in Dugan's Lane area after 1976 contamination	Yes = 1 No = 0		
ZNA[1] to ZNF and ZNJ to ZNQ	Zoning[4]	ZNA = Residential, Rural	Yes = 1	No = 0
		ZNB = Rural Highway Business (Commercial)	Yes = 1	No = 0
		ZNE = Residential R-150	Yes = 1	No = 0
		ZNF = Residential, Planned Retirement Community	Yes = 1	No = 0
		ZNJ = Residential R-400	Yes = 1	No = 0
		ZNO = Residential R-200	Yes = 1	No = 0
		ZNP = Residential R-90	Yes = 1	No = 0
		ZNQ = Residential R-50	Yes = 1	No = 0

[1]Dummy variables.

[2]DD1 and, in some cases, DD2 were the ommitted variables in the group.

[3]DSD is the same as 4 except that it is adjusted to account for residences sold before the school was built. This adjustment only had to be made in the pre-1974 sample.

[4]See "Further Comments," Appendix C, Volume I, for a full description.

APPENDIX D: COMMENTS OF REVIEWERS

In this appendix, some of the comments made on Phase I of the study by eight reviewers are presented and addressed.[1] Reviewers were asked both to criticize the existing text and to comment on whether the study warranted further analysis and, if so, to suggest new approaches. Some of the reviewers not only suggested new approaches, but also commented extensively on the approach adopted in Phase I. Ideally, some of these suggestions would have been addressed and incorporated into the main body of Volume I, however, time and budget constraints did not permit doing so. Also, by the time the comments were received, Phase II had been virtually completed so it was not possible to incorporate all the suggestions even into the latter Phase. However, most of the important criticisms and comments had already been taken into account in performing Phase II of the study. Discussions among the project staff, comments from EPA staff and informal conversations with outside students of environmental economics, including some of the reviewers, had brought forward many of the criticisms that appeared in the formal reviews.

In this appendix, we present and discuss the major comments and suggestions provided by the reviewers, particularly those that were not completely covered in the text of this report.

---

[1]The reviewers are A. Myrick Freeman, III, Department of Economics, Bowdoin College; Robert M. Schwab, Department of Economics, University of Maryland; Jon D. Harford, formerly on the Benefits Staff, Office of Policy Analysis; U.S. Environmental Protection Agency; Robert Mendelsohn, Department of Economics, University of Washington, Seattle, Washington; Allen V. Kneese, Resources for the Future, Washington, D.C.; V. Kerry Smith, Department of Economics, Vanderbilt University; Steve Beggs, Charles River Associates, Boston, Massachusetts; and Al McGartland, Benefits Staff, Office of Policy Analysis, U.S. Environmental Protection Agency.

The comments dealt with the theory and its application and both general and specific technical changes in the computational methodology. The reviewers made no suggestions for obtaining major new sets of data other than those that had already been incorporated in Phase II.

Before discussing the particular suggestions, it is appropriate to mention that several reviewers raised questions about homogeneity of tastes and income among the buyers and sellers in Pleasant Plains. We believe that problems of heterogeneity in these parameters are adequately controlled by virtue of the fact that the Pleasant Plains area is a relatively small part of a larger real estate market. Overall, the reviewers did not present arguments which would justify work beyond Phase II.

Suggestion 1: Very recent housing price data should be isolated to investigate the persistence over time of any impact of proximity to a hazardous waste site on house prices.

The purpose of such an examination would be to determine, first, whether and, if so, when the effect of information about a hazardous waste site became stale and prices settled at a new equilibrium, or, alternatively, whether awareness of the risks was long delayed or had a cumulative effect over time, contributing to a lag in realizing and stabilizing downward pressures on prices.[2] The consequences of a finding

[2] Hedonic equations can be the basis of benefit calculations only if it can be assumed that the housing market was in stock equilibrium in the time period for which the equation was estimated. However, with rational expectations, the market is always in equilibrium, given existing information. Hence, there is only a problem of information. If participants in the real estate market are not adequately informed, property prices will understate (if such risk is exaggerated, overstate) costs of the dump.

that prices do not stabilize until long after local residents had apparently become aware of the incident would suggest that the threshold date chosen to reflect the onset of public awareness was inappropriate and unlikely to generate reliable empirical results.

An investigation of this question was undertaken indirectly in Phase I. The sales-date dummy variables associated with a sample of transactions inside the contamination zones and a sample of the transactions outside the zones were compared to determine whether the rate of increase in house prices was the same in both samples. (The latter sample was to act as a control.) If the actual price impacts were greater inside than outside the contaminated zone, one might expect prices to rise more slowly, at some stage, inside the zone than outside it, before some long-run equilibrium were reached. The analysis shows no indication of such a differential. Indeed, prices generally seemed to rise faster inside the contaminated zones than outside. Unfortunately, a more direct investigation of the question, e.g., generating and testing subsamples, of observations by year or by groups of years, is not feasible, because the samples would be too small without a significant simplification of the model.

Another suggested approach was to introduce distance and time as an interactive variable. This would demonstrate whether prices as a function of distance were rising at an accelerating rate. Finally, it was suggested that an interrupted time series approach might be used on the whole sample to determine whether and, if so, when there was a distinguishable interruption in the response of property values to the incident. Neither of these approaches have been applied, largely because of lack of time and resources but also, in the case of the first, because with the failure to develop any trend associated with distance, there appeared little point in modifying the distance variable with a time subscript.

Suggestion 2: Not to use homogeneity of incomes as a criterion for site selection and consequently consider a larger number of sites as candidates for empirical research.

Heterogeneity of taste or incomes across households does not impair the validity of the results, if the model is used in one affected community which is a small part of a much larger urban housing market and if the sample is large enough to capture the behavior of the entire market. Consequently, a larger number of communities would be or could be considered as candidates.

In practice, no communities were rejected as subjects of empirical research because of not being sufficiently homogeneous.

Suggestion 3: Examine whether the adverse effects of the hazardous waste sites is more closely associated with other factors than with the proxy variables. If that were true, the proxy variables would not constitute reliable indicators of the effects of the site.

We judged that the sales-date dummy variables were most likely to capture some of the effects generated by the hazardous waste site as a consequence of the fact that there might be different impacts on house prices over time. On examination, this did not seem to have occurred. Another variable which might have been affected by the disamenity is lot size. This was not examined. Obviously there is an essentially unlimited list of variables that might have been important. In light of the fact that the models actually used proved to be good predictors of resources but also, in the case of the first, because with the failure to develop any trend associated with distance, there appeared little point in modifying the distance variable with a time subscript.

Suggestion 2: Not to use homogeneity of incomes as a criterion for site selection and consequently consider a larger number of sites as candidates for empirical research.

Heterogeneity of taste or incomes across households does not impair the validity of the results, if the model is used in one affected community which is a small part of a much larger urban housing market and if the sample is large enough to capture the behavior of the entire market. Consequently, a larger number of communities would be or could be considered as candidates.

In practice, no communities were rejected as subjects of empirical research because of not being sufficiently homogeneous.

Suggestion 3: Examine whether the adverse effects of the hazardous waste sites is more closely associated with other factors than with the proxy variables. If that were true, the proxy variables would not constitute reliable indicators of the effects of the site.

We judged that the sales-date dummy variables were most likely to capture some of the effects generated by the hazardous waste site as a consequence of the fact that there might be different impacts on house prices over time. On examination, this did not seem to have occurred. Another variable which might have been affected by the disamenity is lot size. This was not examined. Obviously there is an essentially unlimited list of variables that might have been important. In light of the fact that the models actually used proved to be good predictors of

With respect to the two comments concerning residents' awareness: Although it might have been possible to glean more information from news coverage at that time, without (and perhaps with) a survey it would be impossible to acertain the nature of individuals' perception of the risk. Documentation on the incident provides the only information for this study on the two questions of actual exposure and risk and perceived exposure and risk.

Suggestion 5: The method used to select independent variables, which involved maximizing the $\bar{R}^2$, is inappropriate.

Contrary to the interpretation of some reviewers, the purpose of maximizing the $\bar{R}^2$ was not primarily to maximize the goodness of fit. The technique was not used on the proxy variables; but it was used to deal with the problem of multicolinearity among the property characteristic variables. Selecting variables according to the criteria of maximizing the $\bar{R}^2$ is a widely accepted procedure (albeit not the only one) in such circumstances.

Suggestion 6: Amend the hedonic property model to deal with decision making in the presence of uncertainty.

The theoretical points associated with the issue of uncertainty have not been addressed in this or in other studies applying the hedonic models to evaluating disamenities. They constitute an important area for future work.

Uncertainty in decision making exists in situations where the magnitude of risks is not known or fully understood by the decision makers. While a determination was made of likely exposure of residents to hazardous substances in Pleasant Plains soon after the incident and the results of this determination were well published, there remained some uncertainty about the actual exposure of individual residents to these substances.

Further, knowledge of risk, by its nature, does not provide certainty. In addition, any real estate transaction is replete with uncertainty, since the present value of property is always a reflection of estimates of the effects of future developments in the market. The Pleasant Plains case, however, appears to be an example of a situation in which the area of uncertainty was reduced substantially soon after the incident and remained stable subsequently. Within the first few months of the incident, the contamination zones were defined and actions were taken to reduce the associated risks. While there was a new discovery of contamination in 1976, it was within the known area at risk, i.e., within the existing contamination zone. This suggests that uncertainty was not an important factor in decision making in the Pleasant Plains case.

Suggestion 7: Functional forms, in addition to the ones used in this study should be introduced, particularly for variables for which the semi-log specification may not be appropriate, e.g., some of the contamination, locational, and some property characteristic variables. Another suggestion was made concerning the specification of the dependant variable, i.e, that it should be expressed as price per square foot of house or lot. This is a technique used to deal with a potential problem of heteroskedasticity, i.e., where the variance of the error term is greater for larger lots or houses than for smaller ones.

While several functional forms for the contamination variables were tested in Phase I, only one, the semi-log transformation, was used for these variables in Phase II. While testing different functional forms for the relevant contamination variables may be a useful area for further research, it is unlikely to change the results substantially. The dummy variables representing distance (the only ones which permit investigation of alternative functional forms) have the "wrong" signs in this study, as well as being statistically insignificant.

Modifications of some of the more important variables other than the contamination variable were undertaken in Phase II, i.e., for lot size and house size; however, this was not done for other non-contamination variables. Of these other variables the specification of the locational variable may be of particular importance since there is some interaction between these and the relevant contamination variable. Distance from the highway is an example of a locational variable for which the semi-log specification may not be appropriate (prices are expected to rise at an increasing rate away from the highway, according to this specification). Further work is recommended in this area.

Another change in specification that it might be worthwhile to test is the conversion of property prices into price per square foot. Currently both lot size and house size are entered as independent variables; using price per square foot, instead, would result in dropping one of these variables from the right hand side of the equation and expressing it on the left. This procedure might be undertaken for both variables in separate runs, but not concurrently. Thus, it might be impossible to remove all potential heteroskedasticity. However, since the largest lots (lots more than two acres) were removed from the sample, any potential heteroskedasticity problem (with respect to lot size) is unlikely to be severe.

Suggestion 8: Either omit the date of sale or specify it as a continuous variable, instead of as a quarterly dummy variable. It could be omitted if sales price were expressed in constant dollars. Alternatively, if it were expressed as a continuous variable, it might be appropriate to express it as a high degree polynomial. In either case, the degrees of freedom available would be increased substantially.

While either of these two suggestions would increase the degrees of freedom available, neither would capture the nature of the non-linearity as well as do the quarterly dummies. In fact, a major reason for using the quarterly dummies was to examine the nature of any non-linearity associated with sales prices over time, since there is no way of predicting such non-linearity *a priori.* In Phase I, having established this pattern, we introduced into some of the regressions a single trend variable which incorporated the information observed through using the sales dummies. The purpose of doing so was to simplify the equation for the purpose of exposition. Although technically this also increases the degrees of freedom available, it does not provide any new information. (The coefficients on the other variables do not change, only the f and t ratios and the associated $\bar{R}^2$ improve).

Suggestion 9: Use dummy variables to represent individual neighborhoods (blocks, developments, etc.) instead of the locational variables used (distance from schools, highways, etc.).

It is argued that this approach would avoid the likelihood of substantial specification errors and that the locational variables we used serve as poor proxies for more important neighborhood characteristics.

While there is evidence to support the notion that some of the distance variables are misspecified (the school variables, for example, have different signs and some are statistically insignificant), creating plausible neighborhood dummies would involve extensive discussion with local realtors. This might be a worthwhile but costly endeavor.

In addition, further examination of the contamination zone dummies might reveal that some are acting as neighborhood dummies. This question would need to be investigated before separate (additional) neighborhood dummies were introduced. It is probably unnecessary to account for the smallest neighborhood units (blocks, developments, etc.) in a time-series analysis, since the characteristics differentiating them are unlikely to change substantially over a few years. However, some investigation of this issue would have to be made.

Suggestion 10: Select and test a sample of repeat sales (instances where particular properties are represented in the sample by more than one sale).

Where repeat sales data are identified, the estimation of the complete hedonic equation could be broken down into a two-stage procedure, in the first stage price changes would be explained by the relative changes in the environmental variables (as well as relative changes in depreciation and any other observed variables that are specified to change over time). The full hedonic equation could be estimated in stage 2, using the complete sample of single sales as well as repeat sales and imposing the coefficients on the environmental variables estimated in stage 1. The second stage is desirable, but not essential to such a modification of the study, as it is the coefficients of the environmental variable that are of major interest in this study (and they would be estimated in the first stage).

An advantage of identifying and testing such a sample is that it permits the use of a simpler model; only attributes which change over time

time need be included. House lot characteristics, property specific and neighborhood variables that remain constant over time for each property need not be included.

While this may be a reasonable assumption for some characteristics, it is not plausible for all of them. Many of the residences in the area, for example, were improved or expanded during the period of study. Accounting for all such changes would be feasible but difficult.

A principal advantage of this method is that it eliminates the need to include the locational variables. These, as indicated previously, may be misspecified. However, the method does not eliminate the need to include some sort of overall neighborhood variable since it is clear that the entire area has become more desirable. A suggestion was made to introduce a development variable, such as density, into a time series analysis of the kind described above. The "neighborhood" variable would reflect the fact that Pleasant Plains as a whole had undergone development during the period of study.

Obviously the great disadvantage of using repeat sales is that they constitute only a small sample of total transactions. The question of whether the consequent reduction in degrees of freedom would be more than or less than offset by the reduction in degrees of freedom required has not been examined in detail. Our casual impression is that the limitation on sample size would prove unduly restrictive.