Agglomeration and Human Capital Spillovers

Verstraten, P.W.A.H.

2019

document version
Publisher's PDF, also known as Version of record

Appendix C

C.1 (Hedonic) control variables

Chapter 5 uses several hedonic house characteristics from the NVM dataset and other control variables, such as postal code fixed effects and postal code specific linear time trends. Table C.1 shows the mean values of these variables, before and after the construction of the tunnel, and their effect on house prices, following Equation (5.2).

Table C.1. Descriptive statistics of (hedonic) control variables and their effect on house prices

Variable	Mean value before tunnel	Mean value after tunnel	*p*-value difference ≠ 0	Effect on log house prices
Log lot size in m^2	5.522	5.569	0.000	0.188*** (0.006)
Log living area in m^2	4.768	4.754	0.000	0.568*** (0.017)
Log average floor height in meters (volume in m^3 divided by living area in m^2)	1.095	1.133	0.000	0.428*** (0.020)
Maintenance status				
Bad	0.004	0.004	0.803	(reference)
Bad – mediocre	0.002	0.001	0.002	0.228*** (0.065)
Mediocre	0.019	0.019	0.964	0.186*** (0.033)
Mediocre – reasonable	0.005	0.003	0.013	0.270*** (0.038)
Reasonable	0.115	0.110	0.211	0.331*** (0.032)
Reasonable – good	0.030	0.047	0.000	0.397*** (0.034)
Good	0.754	0.717	0.000	0.493*** (0.033)
Good – excellent	0.005	0.013	0.000	0.559*** (0.035)
Excellent	0.070	0.083	0.000	0.578*** (0.034)
Housing type				
Row house	0.379	0.348	0.000	–0.196*** (0.014)
Semi-detached house	0.019	0.025	0.001	–0.135*** (0.013)
Corner house	0.181	0.163	0.000	–0.185*** (0.012)
Duplex house	0.167	0.189	0.000	–0.097*** (0.010)
Detached house	0.253	0.275	0.000	(reference)
Parking lot				
No parking place	0.547	0.562	0.014	(reference)
Parking place on the street	0.011	0.018	0.000	0.021 (0.015)
Carport	0.019	0.020	0.555	0.052*** (0.010)
Single garage	0.385	0.354	0.000	0.080*** (0.006)

Carport and garage	0.007	0.009	0.104	0.084*** (0.012)
At least a double garage	0.030	0.036	0.007	0.097*** (0.010)
Period of construction				
Before 1906	0.073	0.003	0.000	–0.111*** (0.016)
Between 1906 en 1930	0.137	0.136	0.828	–0.168*** (0.015)
Between 1931 and 1944	0.074	0.060	0.000	–0.171*** (0.017)
Between 1945 and 1959	0.104	0.106	0.713	–0.169*** (0.012)
Between 1960 and 1970	0.164	0.152	0.005	–0.152*** (0.010)
Between 1971 and 1980	0.239	0.196	0.000	–0.115*** (0.013)
Between 1981 and 1990	0.113	0.094	0.000	–0.055*** (0.013)
Between 1991 and 2000	0.094	0.120	0.000	(reference)
2000 and later	0.001	0.042	0.000	0.014 (0.012)
Central heating system				
No	0.145	0.115	0.000	(reference)
Yes	0.855	0.885	0.000	0.099*** (0.007)
Month of transaction				
January	0.082	0.062	0.000	(reference)
February	0.089	0.072	0.000	0.006 (0.006)
March	0.087	0.085	0.664	0.013** (0.006)
April	0.082	0.086	0.279	0.022*** (0.006)
May	0.084	0.086	0.510	0.029*** (0.005)
June	0.082	0.089	0.060	0.022*** (0.005)
July	0.089	0.088	0.762	0.033*** (0.006)
August	0.084	0.083	0.864	0.030*** (0.005)
September	0.077	0.090	0.000	0.036*** (0.006)
October	0.092	0.094	0.636	0.038*** (0.005)
November	0.079	0.087	0.018	0.037*** (0.006)
December	0.074	0.078	0.204	0.040*** (0.005)

Variable	Mean value	Effect on log house prices
Year of transaction		
1995	0.026	(reference)
1996	0.034	0.080*** (0.010)
1997	0.039	0.123*** (0.011)
1998	0.041	0.181*** (0.012)
1999	0.049	0.306*** (0.014)
2000	0.054	0.406*** (0.017)
2001	0.065	0.497*** (0.017)
2002	0.064	0.573*** (0.016)
2003	0.062	0.625*** (0.016)

2004	0.063	0.667*** (0.015)
2005	0.072	0.716*** (0.014)
2006	0.072	0.742*** (0.012)
2007	0.070	0.762*** (0.011)
2008	0.058	0.775*** (0.011)
2009	0.044	0.728*** (0.010)
2010	0.044	0.728*** (0.011)
2011	0.043	0.709*** (0.012)
2012	0.050	0.656*** (0.014)
2013	0.050	0.624*** (0.016)
Postal code specific linear time trends (146)		–0.024*** to 0.031*** (a)
Constant		–2.775 (3.312)
Observations (postal codes)		27,835 (146)
Within R^2		0.869

Notes: Standard errors are clustered at the postal code level (in parentheses). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

(a) This effect varies over postal codes; the table shows the minimum and maximum value.

C.2 Housing data selection procedure

The initial dataset has been cleared to remove incomplete observations. One may think of houses with undisclosed selling price or with unknown lot size. In addition, we have deleted apartments as well as a few houses that have been sold more than five times during the period between 1985 and 2013. Table C.2 summarizes our data selection procedure in sequential order.

Table C.2. Housing data selection procedure and number of observations

Selection criteria	Number of observations
1. Initial dataset (1985–2013)	38,948
2. Remove cases with zero living area	37,233
3. Remove cases with zero lot size	33,434
4. Remove cases with lot size outside 0.1 and 99.9 percentiles	33,369
5. Remove cases with average floor height outside 0.1 and 99.9 percentiles	33,314
6. Remove cases with unknown transaction price	33,311
7. Remove cases with transaction price outside 0.5 and 99.5 percentiles	32,983
8. Remove data on apartments	32,428
9. Remove cases with unknown dwelling type	32,342
10. Remove cases with unknown year of construction	32,160
11. Remove dwellings that have been sold more than five times	31,625
12. Remove postal codes with less than 15 observations in 1985–2013	31,573
13. Remove years before 1995 (due to representativeness concerns)	27,835

C.3 Sensitivity to minimum number of house transactions per postal code

Table C.3 shows that adjusting the minimum number of house transactions yields similar results as the main specification. Hence, the current minimum of 13 transactions per postal code seems to be sufficient to run the analyses accurately. The estimations are even more robust if we add a regional interaction dummy.

Table C.3. Different minimum number of house transactions per postal code

Column:	(1)	(2)	(3)	(4)	(5)	(6)
Equation:	(5.2)	(5.2)	(5.2)	(5.2)	(5.2)	(5.2)
Transactions:	13 (baseline estimation)	50	100	13 (baseline estimation)	50	100
$\theta_{ant}\omega_{2000}$	0.033	−0.046	−0.077	0.219	0.178	0.043
	(0.161)	(0.169)	(0.142)	(0.230)	(0.244)	(0.211)
$\theta_{ant}\omega_{2001}$	−0.031	−0.029	−0.083	0.279	0.289	0.235
	(0.173)	(0.180)	(0.175)	(0.221)	(0.230)	(0.233)
$\theta_{ant}\omega_{2002}$	0.427*	0.447**	0.354	0.781***	0.815***	0.717***
	(0.218)	(0.225)	(0.240)	(0.236)	(0.242)	(0.265)
$\theta_{ant}\omega_{2003}$	0.599**	0.574**	0.392	1.077***	1.056***	0.901***
	(0.271)	(0.282)	(0.308)	(0.273)	(0.285)	(0.318)
θ	0.790***	0.788***	0.642**	1.497***	1.524***	1.451***
	(0.277)	(0.290)	(0.318)	(0.257)	(0.270)	(0.310)
$\theta_{ant}\omega_{2000}D_r^{South}$				−0.363**	−0.344*	−0.212
				(0.179)	(0.185)	(0.162)
$\theta_{ant}\omega_{2001}D_r^{South}$				−0.578***	−0.568***	−0.490***
				(0.163)	(0.166)	(0.164)
$\theta_{ant}\omega_{2002}D_r^{South}$				−0.661***	−0.658***	−0.560***
				(0.180)	(0.180)	(0.177)
$\theta_{ant}\omega_{2003}D_r^{South}$				−0.733***	−0.691***	−0.578**
				(0.240)	(0.241)	(0.244)
θD_r^{South}				−1.317***	−1.315***	−1.259***
				(0.202)	(0.204)	(0.212)
Observations	27,835	26,456	24,272	27,835	26,456	24,272
(postal codes)	(146)	(105)	(75)	(146)	(105)	(75)
Within R^2	0.869	0.869	0.871	0.870	0.869	0.872

Notes: All results are based on postal code fixed effects regressions. Standard errors are clustered at the postal code level (in parentheses). Year fixed effects, month fixed effects and hedonic controls for house characteristics are included as well as postal code specific linear time trends. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.4 Anticipation effects per quarter

Table C.4 reports the results for Equation (5.2) if we employ anticipation variables for the first quarter of 2000 until the first quarter of 2003 (instead of using yearly anticipation variables). Anticipation turns significant in the third quarter of 2002 and increases towards the last quarter of 2002. Compared to the previous quarter, there is no additional anticipation effect in the first quarter of 2003.

Table C.4. Effect of accessibility on house prices with anticipation per quarter

Column: Equation:	(1) (5.2) with finer-grained anticipation periods
$\theta_{ant}\omega_{2000,q1}$	–0.294* (0.176)
$\theta_{ant}\omega_{2000,q2}$	0.022 (0.156)
$\theta_{ant}\omega_{2000,q3}$	0.096 (0.169)
$\theta_{ant}\omega_{2000,q4}$	0.222 (0.186)
$\theta_{ant}\omega_{2001,q1}$	–0.176 (0.177)
$\theta_{ant}\omega_{2001,q2}$	–0.090 (0.176)
$\theta_{ant}\omega_{2001,q3}$	0.027 (0.202)
$\theta_{ant}\omega_{2001,q4}$	0.131 (0.183)
$\theta_{ant}\omega_{2002,q1}$	0.214 (0.227)
$\theta_{ant}\omega_{2002,q2}$	0.350 (0.220)
$\theta_{ant}\omega_{2002,q3}$	0.479** (0.230)
$\theta_{ant}\omega_{2002,q4}$	0.606** (0.243)
$\theta_{ant}\omega_{2003,q1}$	0.568** (0.270)
θ	0.793*** (0.277)
Observations (postal codes)	27,835 (146)
Within R^2	0.869

Notes: All results are based on postal code fixed effects regressions. Standard errors are clustered at the postal code level (in parentheses). Year fixed effects, month fixed effects and hedonic controls for house characteristics are included as well as postal code specific linear time trends. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.5 Results with an accessibility measure based on population

All results in Chapter 5 are based on the accessibility of jobs. Table C.5 shows that similar results for Table 5.3 are obtained once we calculate accessibility on the basis of population rather than jobs.

Table C.5. Accessibility measure based on population

Column: Equation:	(1) (5.1)	(2) (5.2)	(3) (5.3)	(4) (5.4)
$\theta_{ant}\omega_{2000}$		0.058 (0.167)		0.115 (0.158)
$\theta_{ant}\omega_{2001}$		0.014 (0.174)		0.092 (0.161)
$\theta_{ant}\omega_{2002}$		0.494* (0.206)		0.594*** (0.190)
$\theta_{ant}\omega_{2003}$		0.700*** (0.246)		0.834*** (0.228)
θ	0.539*** (0.0116)	0.907*** (0.247)	0.508*** (0.102)	1.050*** (0.0.224)
$\theta_{del}\omega_{2008-2013}$			–0.106 (0.134)	0.163 (0.107)
Observations (postal codes)	27,835 (146)	27,835 (146)	27,835 (146)	27,835 (146)
Within R^2	0.869	0.869	0.869	0.869

Notes: All results are based on postal code fixed effects regressions. Standard errors are clustered at the postal code level (in parentheses). Year fixed effects, month fixed effects and hedonic controls for house characteristics are included as well as postal code specific linear time trends. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.6 Results with a finer-grained postal code fixed-effects interaction

The last column of Table 5.6 shows the results of a regression with a postal code fixed-effects interaction term using a dummy variable that equals one if the postal code's fixed effect was larger than the median value, and zero otherwise. Table C.6 presents the results for a finer-grained division of postal code fixed effects into tertiles and quartiles.

Table C.6. A finer-grained postal code fixed-effects interaction

Column: Equation:	(1) (5.2) with postal code fixed-effects interaction (tertiles)	(2) (5.2) with postal code fixed-effects interaction (quartiles)
$\theta_{ant}\omega_{2000}$	–0.559*	–0.663
	(0.324)	(0.547)
$\theta_{ant}\omega_{2001}$	–0.797**	–0.845*
	(0.313)	(0.470)
$\theta_{ant}\omega_{2002}$	–0.355	–0.361
	(0.343)	(0.475)
$\theta_{ant}\omega_{2003}$	–0.239	0.145
	(0.523)	(0.569)
θ	–0.589	–0.111
	(0.500)	(0.521)
$\theta_{ant}\omega_{2000}Q_2$	0.625*	0.726
	(0.321)	(0.532)
$\theta_{ant}\omega_{2001}Q_2$	0.755***	0.796*
	(0.270)	(0.435)
$\theta_{ant}\omega_{2002}Q_2$	0.802***	0.726*
	0.289)	(0.430)
$\theta_{ant}\omega_{2003}Q_2$	0.893**	0.364
	(0.451)	(0.524)
θQ_2	1.247***	0.606
	(0.440)	(0.476)
$\theta_{ant}\omega_{2000}Q_3$	0.410	0.409
	(0.332)	(0.524)
$\theta_{ant}\omega_{2001}Q_3$	0.721**	0.609
	(0.293)	(0.424)
$\theta_{ant}\omega_{2002}Q_3$	0.639**	0.754*
	(0.280)	(0.410)
$\theta_{ant}\omega_{2003}Q_3$	0.655	0.685
	(0.448)	(0.524)
θQ_3	1.568***	1.186***
	(0.392)	(0.445)
$\theta_{ant}\omega_{2000}Q_4$		0.828
		(0.520)
$\theta_{ant}\omega_{2001}Q_4$		1.167***
		(0.430)
$\theta_{ant}\omega_{2002}Q_4$		0.945**
		(0.407)
$\theta_{ant}\omega_{2003}Q_4$		0.609
		(0.522)
θQ_4		1.417***
		(0.436)
N (postal codes)	27,835 (146)	27,835 (146)
Within R^2	0.869	0.870

Notes: All results are based on postal code fixed effects regressions. Standard errors are clustered at the postal code level (in parentheses). Year fixed effects, month fixed effects and hedonic controls for house characteristics are included as well as postal code specific linear time trends. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.