A study of English house price data with spatial dependence

Illir Nase

Department of Real Estate & Housing

Faculty of Architecture and the Built Environment

Delft University of Technology

Outline

- ‘Housekeeping’
 - Theory
 - Variable design
 - Data mining
- ‘Entertaining’
 - Model(s): OLS → spatial panels
 - Specific to general
 - General to specific
 - Effect estimates
- ‘Seeing off’ ← • Follow-up

Theory

- Formation of regional house prices
 - Supply and demand + spatially dependent (spillovers)
- Supply dependent on existing stock
- Demand on income within commuting distance
- Displaced supply: high prices nearby cause supply to fall (suppliers looking for higher returns)
- Displaced demand: high prices nearby causes demand to increase (purchasers move to cheaper places)

Variable design

- House prices: Mean house prices for English local authorities (LA) (2004-2012)
- Supply: available housing stock
 - Also, additional stock per year
- Demand: based on income/purchasing power (within commuting distance of a locality)
- Other (socio-economic effects, school quality, crime rate effects, taxes, amenities, etc.)
- (available): crime data by LA

Variable design

- Supply (S_{it}): Yearly dwelling stock by LA (per 10 persons)
 - Stock: Calculated from 'Dwelling stock estimates' & 'Population estimates' (Source: NOMIS & ONS websites)
 - NAD1000: Net additional Dwellings by LA (per 1000 persons) (Source: DCLG Live tables & ONS)
- Demand: Yearly income by district (Y_{it}) = mean wage average (ω_{it}) * employment level (E_{it})
- Commuting flow matrix ($\mathbf{W}$): from 2011 population census (available from NOMIS & ONS; see also Baltagi et al. 2014 for more details)

Variable design

- Yearly income within commuting distance by LA weighted by each row of the (standardised) **W** holding the diagonal elements
- Unstandardised **W** extract:

	""Halton""	""Warrington""	""Blackburn with Darwen""	""Blackpool""	""Cheshire East""	""Cheshire West and Chester""
""Halton""	27270	5786	22	16	513	2680
""Warrington""	4674	50422	148	44	2005	2462
""Blackburn with Darwen""	27	208	31801	222	51	28
""Blackpool""	27	96	193	32638	25	19
""Cheshire East""	691	2073	69	26	94009	7996
""Cheshire West and Chester""	3663	3894	57	16	9041	80360

- Crime rates: Total offences recorded by the police and Community Safety Partnerships (per 100 persons) (source: ONS)

Data Mining (Issues)

- Changing LA boundaries (2009)
 - Pre-2009: N=365
 - 2009 onwards: N=326
- Offence records stored by Police Force Area (for certain LA & years)
- **W**: 2 aggregated records
 - N=324 → Final for this analysis

Data Mining (approach)

- Changing LA boundaries & 'house prices'

New ONS Code	Old ONS Code	LA Name	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
E06000048	00EM	Northumberland UA	115,526	141,409	153,796	165,140	178,371	181,190	170,912	180,428	169,904	167,730
E06000049	00EQ	Cheshire East UA	..	..	..	..	..	223,252	212,059	231,940	217,462	214,341
E06000050	00EW	Cheshire West and Chester UA	..	..	..	..	..	196,188	193,164	198,398	190,367	192,797
E06000051	00GG	Shropshire UA	152,718	177,534	188,424	201,367	213,622	211,639	198,086	208,309	204,035	199,890
E06000052	00HE	Cornwall UA	..	..	..	..	..	228,266	213,366	228,585	220,222	223,264
E06000053	00HF	Isles of Scilly UA	286,891	280,069	305,617	398,842	392,476	335,000	405,429	342,727	363,700	365,077
E06000054	00HY	Wiltshire UA	187,643	206,756	213,237	225,437	242,074	237,637	227,454	244,345	247,007	239,854
E06000055	00KB	Bedford UA	155,380	172,233	180,287	192,086	207,555	206,031	194,015	218,045	212,408	213,658
E06000056	00KC	Central Bedfordshire UA	..	..	..	..	..	218,843	206,136	223,526	219,350	224,178
E07000004	11UB	Aylesbury Vale	199,579	221,652	227,470	244,155	267,937	256,587	241,101	273,022	265,970	265,089

➤ (Centred) Moving Averages (table581 DCLG archives)

		2004				2005				2006				2007			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Cheshire East UA		..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	Congleton	156,071	163,660	169,012	181,323	181,435	171,919	179,735	170,759	179,249	185,534	193,344	194,208	196,377	201,383	204,363	209,144
	Crewe and Nantwich	134,712	148,193	157,203	154,007	159,923	154,919	161,398	148,227	146,875	166,688	172,505	170,219	164,346	180,203	175,142	190,133
	Macclesfield	218,269	217,538	249,162	236,188	223,886	243,294	235,692	243,752	236,701	255,997	271,738	260,431	269,344	271,809	307,648	306,967
Cheshire West and Chester UA		..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	Chester	170,092	186,307	198,588	195,782	183,740	187,356	192,167	199,479	190,509	202,893	202,715	217,959	213,081	220,637	218,330	225,930
	Ellesmere Port and Neston	135,277	148,083	154,424	166,795	155,785	159,816	158,706	165,365	160,696	164,725	157,770	165,582	158,952	164,714	169,386	171,535
	Vale Royal	142,241	161,078	179,227	168,330	165,378	167,240	176,311	180,802	175,320	186,676	189,990	185,542	186,501	192,622	199,439	190,962

Data Mining (approach)

- Changing LA boundaries & 'dwelling stock', 'population', 'income' & 'crime' data
- Changing **W** records & 'dwelling stock', 'population' & 'crime' data
- Adding & averaging
- Changing offence records & 'crime' data
- Forecasting models

[illegible]

Data Mining (theory & variables)

- Data ordered first by LA then by year
 - $i=1\dots324$: fast running index
 - $t=1\dots9$: slow running index
- **Y** Dependent variable vector (1×324 , $T=9$): (*logHP*)
- **X** matrix: (4×324 , $T=9$)
 - Income within commuting distance (*logincome*)
 - Available dwelling stock per 10 persons (*stock*)
 - Net additional Dwellings per 1000 persons (*NAD1000*)
 - Crime rate per 100 persons (*crime*)

Data Mining (theory & variables)

- '*crime*' & '*stock*' lagged 1 year to be treated as exogenous
 - ('Y' @ t depends on '*crime*' & '*stock*' @ t-1)
- '*NAD1000*' → includes demolitions & conversions
(Some cells have negative values)
following Glaeser et al. (2014): supply ≥ 0
These are restricted to = 0

(Pre-modelling) Caveats

- Population sampled exhaustively (sample=population)
- Travel flows change by year: assumed to be constant
 - **W**: 2011 census data (+ 2001 matrix large # of missing data)
- General model stationarity conditions :
Parameter space of, eg. (rho) $1/r_{\min} < \rho < 1/r_{\max}$ cannot be ensured beforehand ($r_{\min-\max}$: matrix real characteristic roots; for row standardisation $r_{\max} = 1$)

Models (specific to general)

- Pooled OLS estimates

Outcomes	OLS	OLS with S FE	OLS with T FE	OLS with ST FE
logincome	0.150360 (14.420395)	0.243531(23.951307)	0.138844 (12.577966)	0.020422 (2.482123)
stock	-0.024673 (-2.024467)	-0.003529 (-0.321161)	-0.003529 (-0.321161)	-0.093570 (-12.012658)
NAD1000	0.007645 (5.354274)	0.001330 (3.500329)	0.007739 (5.110095)	0.000815 (2.949740)
crime100	-0.014053 (-15.280970)	-0.007931 (-20.407119)	-0.013021 (-12.902746)	-0.001363 (-3.521221)
intercept	4.415257 (44.686239)			
R ²	0.101	0.3587	0.0849	0.0623
σ ²	0.0236	0.0007	0.0233	0.0003
R ² FE		0.9719	0.1122	0.9869
LogL		6379	1344.7	7491.6
Durbin-Watson	0.8174	0.8779	0.833	1.295
LM spatial lag	580.1539 (0.000)	5625.1397 (0.000)	7996.1955 (0.000)	5015.9805 (0.000)
LM spatial error	9237.6950 (0.000)	5598.4940 (0.000)	9198.7298 (0.000)	9198.7298 (0.000)
robustLM spatial lag	317.2271 (0.000)	954.4738 (0.000)	45.9182 (0.000)	0.7381 (0.390)
robustLM spatial error	8974.7681 (0.000)	927.8281 (0.000)	1248.4525 (0.000)	252.9612 (0.000)

(for coefficients 't' & for test results 'p' values in parentheses ; Nobs = 2916)

LR SFE, df, sig. : 12293.6546, 324, 0.0000

LR TFE, df, sig. : 2225.2040, 9, 0.0000

based on Elhorst (2014)

Models (weight matrices)

- Spatial Weight matrix **W**: The matrix of travel flows with diagonal elements set to zero
- Five different specifications → cut-off points
 - Number of commuters (30, 50, 100, 200, 500)
- Compare LM values (model results OLS with ST FE)

Models (cut-off & **W** sparsity)

- $C = 30$

Models (cut-off & **W** sparsity)

- $C = 50$

Models (cut-off & **W** sparsity)

- $C = 100$

Models (cut-off & **W** sparsity)

- $C = 200$

Models (cut-off & **w** sparsity)*

- $C = 500$

* No intentional nesting,
resulting configuration
from travel flow matrix
ordering $\rightarrow$ turns out to
be English regions

Models (general to specific)

- Panel SDM estimates

$$Y_t = \rho WY_t + \alpha \mathbf{1}_N + X_t \beta + WX_t \theta + u_t$$

min & max (rho) $\rho = -0.8273, 1.0000$

Hausman, df, p
445.6933, 9, 0.0000

Outcomes	SDM with ST FE (bias corr.)	SDM with S RE & T FE
logincome	0.010562 (1.573264)	0.065114** (9.687186)
stock	0.035244** (4.997596)	0.096810** (14.317834)
NAD1000	0.000550* (2.436883)	0.000960** (4.047367)
crime100	-0.002954** (-8.782338)	-0.003488** (-10.425119)
W*logincome	-0.047124** (-6.874298)	0.003114 (0.697777)
W*stock	0.042578** (7.021613)	0.037226** (6.117570)
W*NAD1000	0.000476* (2.279600)	0.001640** (8.737716)
W*crime100	0.000914** (5.502909)	0.001182** (8.581272)
W*dep.var	0.223656** (27.545977)	0.264641** (51.207286)
R ²	0.9923	0.99
σ^2	0.0002	0.0003
LogL	8124.5543	6079.346
Wald_spatial_lag	155.8823**	313.2177**
LR_spatial_lag	-32.4229	310.9135**
Wald_spatial_error	114.2005**	287.9046**
LR_spatial_error	-83.449	2367.6**

Effect estimates (theory)

- The matrix of partial derivatives of the expected value of $\mathbf{Y}$ with respect to the k^{th} explanatory variable of $\mathbf{X}$

$$\partial y / \partial x^{r'} = (I_n - \rho W)^{-1} (I_n \beta_r + W \theta)$$

- Direct effects $\rightarrow$ diagonal elements (summary stat.)
 - how changes in the r^{th} explanatory variable for the i^{th} LA impact the i^{th} LA's dependent variable
- Spillovers $\rightarrow$ off-diagonal elements (summary stat.)
 - the impact on the j^{th} LA outcomes y_j from a change in the r^{th} explanatory variable from the i^{th} LA

Effect estimates (results)

- Focus on '*crime*'

Coefficients	Model	
	1	2
logincome	0.010562	0.065114**
stock	0.035244**	0.096810**
NAD1000	0.000550*	0.000960**
crime100	-0.002954**	-0.003488**
W*logincome	-0.047124**	0.003114
W*stock	0.042578**	0.037226**
W*NAD1000	0.000476*	0.001640**
W*crime100	0.000914**	0.001182**
W*dep.var	0.223656**	0.264641**

Variable	Effects	Model 1	Model 2
		SDM with ST FE (bias corr.)	SDM with S RE & T FE
logincome			
	direct	0.0004 (0.0598)	0.0709 (10.4580)
	spillovers	-6.3502 (-0.2836)	0.1805 (5.4910)
	total	-0.2836 (-5.9169)	0.2514 (7.2972)
stock			
	direct	0.0467 (6.4137)	0.1159 (16.6386)
	spillovers	0.3242 (8.5395)	0.5695 (11.1103)
	total	0.3708 (9.1497)	0.6854 (12.8794)
NAD1000			
	direct	0.0007 (2.9279)	0.0015 (5.9485)
	spillovers	0.0038 (2.8773)	0.0170 (10.6416)
	total	0.0045 (3.1512)	0.0185 (10.9342)
Crime100			
	direct	-0.0029 (-8.5482)	-0.0034 (-9.7514)
	spillovers	0.0016 (1.5865)	0.0023 (2.4609)
	total	-0.0013 (-1.0902)	-0.0011 (-1.0025)

Follow-up

- ‘*crime*’ results interesting & following the ‘displaced’ logic
- ‘*stock*’ sign not in line with Baltagi et al. (2014) ‘displaced supply’
- To consider:
 - Multilevel Spatial Panels
 - Dynamic Spatial Panels
 - School quality data (tentative)
 - Different spec. of **W** (contiguity)

Thanks

+

Questions?