Predicting Property Value within the Liverpool Authority District using Spatial Econometric Models

201374125[a]

[a]Department of Geography and Planning, University of Liverpool, Liverpool, L69 7ZX

This version was compiled on May 13, 2019

Word Count: 1643

1. Introduction

The importance of the spatial element in determining the factors driving property price is widely recognised (Bitter, Mulligan and Dall'erba, 2007), and controlling for this in the production of property pricing models is essential. The spatial variation in property price is driven by two key econometric issues;

1. Spatial Dependence

   The lack of independence often present in observations of cross-sectional data, i.e. the property price is influenced by surrounding property prices

2. Spatial Heterogeneity

   The underlying process cannot be assumed to be constant throughout the geography selected, i.e. the variation in property prices relies on an aggregate of geographical features that vary across space (Anselin, 2013; Anselin and Arribas-Bel, 2013)

Extending these concepts in relation to property prices will allow for an improved understanding of these key spatial processes that are not always considered in spatial models in property price research (Bowen, Mikelbank and Prestegaard, 2001). Typically empirical models of property price research have treated the chosen area (often metropolitan) as a single market (Bitter, Mulligan and Dall'erba, 2007), although recent developments in Multi Level models for use in spatial analysis have extended the analysis, allowing for the introduction of the concept of housing 'neighbourhoods' (e.g. Dong et al., 2019).

This report will first utilise a spatial fixed effect model, which will separate each individual area as a fixed effect (in this case two digit postcodes). The fixed effect model therefore will compare only properties that share the same postcode, isolating the effects of the chosen predictor variables. This model essentially allows the intercept to vary between the two digit postcodes within Liverpool (Anselin and Arribas-Bel, 2013). Secondly a Spatial Lag model will be used to assess the degree of spatial dependence that is observed in a selected variable.

2. Preliminary Analysis

Prior to running the models, a Kernel Density Map was produced which allowed for an exploration of the distribution of property sales (Figure 1). To select the bandwidth an adapted version of Scott's rule was implemented into *r* (Bowman and Azzalini, 1997):

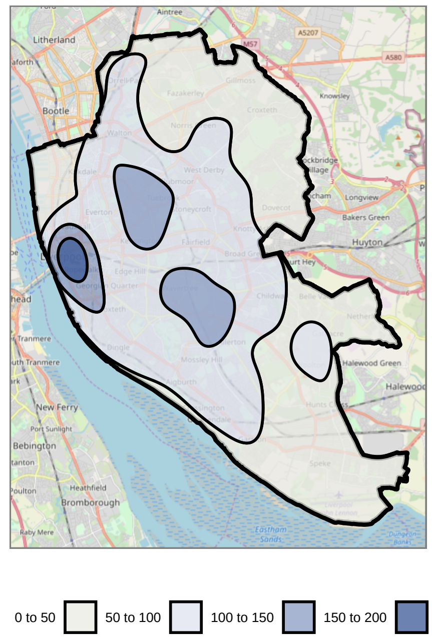

Fig. 1. Kernel Density Estimate showing property sales in the Liverpool region. Bandwidth selected using Scott's Rule

```r
choose_bw <- function(spdf) {
  X <- st_coordinates(spdf)
  sigma <- c(sd(X[, 1]), sd(X[, 2])) *
    (2 / (3 * nrow(X)))^(1 / 6)
  return(sigma / 1000)
}
```

Figure 1 shows that a high proportion of property transactions occur around the city centre, with fewer on the outskirts. Likely due to there being a higher density of properties in the city centre, and more flats as opposed to houses.

Figure 2 gives the results of a spatial interpolation technique showing the estimated distribution of property value throughout the LAD. North of the city centre there tends to be lower property values, whereas south of the city centre there is a band of high property values.

3. Regression Models

3.1. Baseline non-spatial regression. The results from the preliminary OLS regression analysis given on Table 1 show that each chosen predictor variable significantly influences the property price. Particularly, flats negatively influence the property price by the largest factor; flats are associated with a 0.84% decrease in the value of property price. Similarly Semi-detached and terraced properties are associated in a price reduction, suggesting that detached

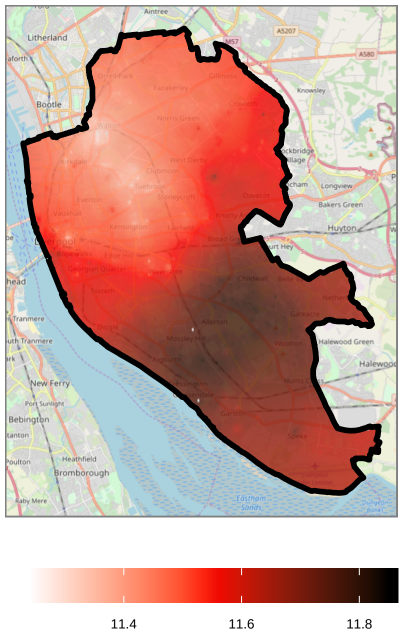

Fig. 2. Spatial Interpolation of Property Sales Price within Liverpool using Inverse Distance Weighting

Table 1. OLS Linear Regression Model

Variable	Model Estimate	Standard Errors
(Intercept)	12.69 ***	0.02
New Build	0.27 ***	0.02
IMD Score	-0.02 ***	0.00
Flat	-0.84 ***	0.02
Semi-Detached	-0.35 ***	0.02
Terraced	-0.66 ***	0.02

† R Squared = 51.01
*** Significant at the 0.001 level

properties are associated with a significant price increase.

Notably the Index of Multiple Deprivation (IMD) Score does negatively influence property price significantly, but only to a small degree, likely as IMD Score is an aggregate of many deprivation indicators, and the other chosen variables may be acting as a proxy towards these. Additionally it should be noted that IMD is an area level variable, and associated with the Lower Super Output Area (LSOA) where the property is located, therefore the assumption is made that every property is associated with the IMD Score of the containing LSOA. The equation for the OLS regression is given below, (Equation 1), house prices were log transformed as without log transformation the distribution was not normal, an assumption with OLS regression (Li et al., 2012).

$$\log P_i = \alpha + \beta_1 NEW_i + \beta_2 IMD_i + \beta_3 TYPE_i + \epsilon_i \qquad (1)$$

3.2. Spatial Fixed Effects Model based on 2 Digit Postcode. OLS regression assumes that all properties share the same constant term and alpha across all spatial areas within Liverpool, however, space can be included into the model, to act as a proxy for unobserved variables that act vary across space. The equation for this model is given (Equation 2). In this model, property 'type' is converted into the dummy variable 'detached' which identifies whether a

Table 2. Spatial Fixed Effects Model

Variable	Model Estimate	Standard Errors
L1	11.57 ***	0.03
L10	11.80 ***	0.06
L11	11.75 ***	0.04
L12	11.86 ***	0.03
L13	11.79 ***	0.03
L14	11.78 ***	0.04
L15	11.98 ***	0.02
L16	12.20 ***	0.03
L17	12.20 ***	0.02
L18	12.31 ***	0.02
L19	12.10 ***	0.02
L2	11.82 ***	0.09
L20	11.40 ***	0.21
L24	11.82 ***	0.04
L25	12.10 ***	0.02
L27	11.96 ***	0.08
L28	11.39 ***	0.17
L3	11.90 ***	0.02
L4	11.65 ***	0.03
L5	11.91 ***	0.05
L6	11.81 ***	0.04
L7	11.77 ***	0.04
L8	11.98 ***	0.03
L9	11.66 ***	0.03
New	0.28 ***	0.02
Detached	0.61 ***	0.02
IMD Score	-0.01 ***	0.00

† R Squared = 99.87
*** Significant at the 0.001 level

particular property is detached and given the value 1, otherwise given 0. Similarly the category 'new' has been changed to a value of either 1 or 0.

$$\log P_i = \alpha_r + \beta_1 NEW_i + \beta_2 IMD_i + \beta_3 DETACHED_i + \epsilon_i \qquad (2)$$

Here space is included into the model as it is known to have a significant influence on property price (Waddell, Berry and Hoch, 1993). Notably the accessibility of an area in terms of general connectivity with the city is known to drive an increase in property value (Narvaez, Penn and Griffiths, 2012). As well as the access to amenities in the immediate area, including public transport connections (Bartholomew and Ewing, 2011). In this model therefore, the inclusion of two digit postcodes will act as 'neighbourhoods' in which the baseline property value is expected to vary due to the variables noted above, in addition to spatial characters that are difficult to quantify. Essentially the spatial heterogeneity of Liverpool is now considered. Table 2 indicates that the postcode L18 is the most desirable postcode as it displays the highest log property price intercept (Model Estimate = 12.31) when detached homes, new builds, and IMD score is taken into account. Figure 3 displays these intercepts graphically, which gives an indication as to where the highest and lower intercepts are located. The high r^2 value shown on table 2 likely reflects only the high number of variables, and doesn't assist with model fit interpretation. It is clear from Figure 3 that property price is spatially correlated, and as such likely is effected by spatial dependence.

3.3. Including Spatial Lag of IMD Score. The inclusion of the spatial lag of IMD Score into the model builds on the assumption that

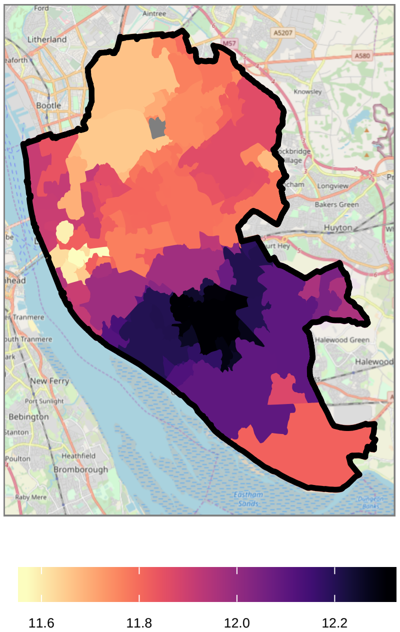

Fig. 3. Varying Spatial Estimates for Log property Price, produced by the Spatial Fixed Effects Model

Table 3. Spatial Lag Model

Variable	Model Estimate	Standard Errors
(Intercept)	12.22 ***	0.01
New Build	0.18 ***	0.02
Detached	0.63 ***	0.02
IMD Score	-0.01 ***	0.00
IMD Score (Lag)	-0.01 ***	0.00

† R Squared = 51.01
*** Significant at the 0.001 level

property value within a certain postcode is influenced by the IMD Score of surrounding properties (spatial dependence). It should be noted again that IMD Score is an LSOA level unit, so many properties share the exact same IMD Score, for this reason, the property sale prices were aggregated and averaged into each LSOA, queen weightings were used to demonstrate the lag in IMD Score between neighbouring LSOAs. Again, Table 3 reveals that detached properties have by far the greatest positive influence on property prices (Detached properties indicate a 0.64% rise in property price). Notably the spatial lag for IMD score is significant, but gives a relatively small negative value, potentially indicating either the lack of influence that IMD Spatial lag has, or that the new and detached properties dummy variables are acting as a proxies towards the majority of IMD Score indicators.

$$y_{lag-i} = \sum_j w_{ij} y_j \tag{3}$$

$$\log P_i = \alpha + \beta_1 NEW_i + \beta_2 DETACHED_i + \beta_3 IMD_i + \rho IMD_{lag-i} + \epsilon_i \tag{4}$$

4. Discussion

Spatial interpolation is able to alleviate the problem scale and resolution has in relation to the modifiable aerial unit problem (MAUP; Openshaw, 1984) if points were instead aggregated into spatial units such as LSOA (Fotheringham, 1989). LSOA units are broadly unrelated to the outcome variable property price, whereas interpolation uses the underlying data to estimate the distribution. While spatial interpolation tackles the issues with the MAUP, it introduces more assumptions, namely the assumption that property prices are all spatially correlated with each other, to the same degree, and that the distance decay selected with the calculation appropriately reflects that of the underlying, but missing, data (Kar and Hodgson, 2012).

The figure displaying the spatial interpolation of property prices (Figure 2) displays both spatial heterogeneity, where high and low values are clustered spatially, both towards the north and south of Liverpool, and spatial dependence as property price exhibits gradual shifts from high to lower values (See also 2; Anselin, 2013). The presence of these two econometric issues means that without considering space, predicting house price would be inaccurate purely by using variables in a non spatial model as they violate the underlying assumptions of regression analysis (Anselin, 2010).

Spatial fixed effects allow the properties within postcodes to be treated independently, attempting to remove the spatial correlations between postcodes to achieve a more accurate prediction of property price. However, it is limited in that once the number of spatial groups (e.g. Postcodes) become too large, the model itself becomes too large to appropriately interpret. Additionally, the r^2 value for the spatial fixed effect model chosen adds little information in terms of the model fit, essentially the model fit is often hard to interpret and quantify, so must be judged independently. While it is often suggested that spatial fixed effects models do remove spatial correlations, Anselin and Arribas-Bel (2013) note that with true spatial dependence this may not be true. As spatial fixed effects use space as a proxy for many unobserved variables, it is difficult to determine exactly what causes the between group differences, and due to this, any attempt to utilise the model outside of the specified area will not work.

Alternative methods for treating spatial dependence and heterogeneity exist, for example McMillen (2010) suggests that any observed spatial autocorrelation in models is due to omitting variables, and states that non spatial semi-parametric modelling should be used instead of using spatial fixed effects as a proxy for unobserved variables.

The primary issue with the inclusion of the IMD Score with this analysis is that IMD Score is automatically aggregated into LSOAs within Liverpool, and so is an area level indicator paired with individual level indicators such as property type, new builds, and price, and an assumption of spatial fixed effects is that all variables chosen are at an individual level (Anselin, 2013).

A multi level model approach to determining the area effect of house prices could consider the postcodes as a lower aggregation unit, and the entire LAD as the higher level unit. Results determined from a multi level analysis would indicate the level of variation between the two units, allowing for the consideration that housing spatial and temporal trends occur at more than just a macro level (Gibbons and Machin, 2008). Additionally, the majority of multi level models in property price research find that there is at least some unexplained variation between lower and higher level hierarchical units (Dong et al., 2019; Jones and Bullen, 1993), suggesting that there is an area effect that determines spatial differences in property price. The models utilised above however give insight into the spatial dependence and heterogeneity that is apparent with property prices within Liverpool.

References

Anselin, Luc. 2010. “Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity.” *Geographical Analysis* 20(1):1–17.

Anselin, Luc. 2013. *Spatial Econometrics: Methods and Models*. Vol. 4 Springer Science & Business Media.

Anselin, Luc and Daniel Arribas-Bel. 2013. “Spatial Fixed Effects and Spatial Dependence in a Single Cross-Section: Spatial Fixed Effects and Spatial Dependence.” *Papers in Regional Science* 92(1):3–17.

Bartholomew, Keith and Reid Ewing. 2011. “Hedonic Price Effects of Pedestrian- and Transit-Oriented Development.” *Journal of Planning Literature* 26(1):18–34.

Bitter, Christopher, Gordon F. Mulligan and Sandy Dall’erba. 2007. “Incorporating Spatial Variation in Housing Attribute Prices: A Comparison of Geographically Weighted Regression and the Spatial Expansion Method.” *Journal of Geographical Systems* 9(1):7–27.

Bowen, William M, Brian A Mikelbank and Dean M Prestegaard. 2001. “Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications.” *Growth and change* 32(4):466–490.

Bowman, Adrian W and Adelchi Azzalini. 1997. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Vol. 18 OUP Oxford.

Dong, Guanpeng, Levi Wolf, Alekos Alexiou and Dani Arribas-Bel. 2019. “Inferring Neighbourhood Quality with Property Transaction Records by Using a Locally Adaptive Spatial Multi-Level Model.” *Computers, Environment and Urban Systems* 73:118–125.

Fotheringham, A Stewart. 1989. “Scale-Independent Spatial Analysis.” *Accuracy of spatial databases* pp. 221–228.

Gibbons, S. and S. Machin. 2008. “Valuing School Quality, Better Transport, and Lower Crime: Evidence from House Prices.” *Oxford Review of Economic Policy* 24(1):99–119.

Jones, Kelvyn and Nina Bullen. 1993. “A Multi-Level Analysis of the Variations in Domestic Property Prices: Southern England, 1980-87.” *Urban Studies* 30(8):1409–1426.

Kar, Bandana and Michael E Hodgson. 2012. “A Process Oriented Areal Interpolation Technique: A Coastal County Example.” *Cartography and Geographic Information Science* 39(1):3–16.

Li, Xiang, Wanling Wong, Ecosse L. Lamoureux and Tien Y Wong. 2012. “Are Linear Regression Techniques Appropriate for Analysis When the Dependent (Outcome) Variable Is Not Normally Distributed?” *Investigative Opthalmology & Visual Science* 53(6):3082.

McMillen, Daniel P. 2010. “Issues in Spatial Data Analysis.” *Journal of Regional Science* 50(1):119–141.

Narvaez, Laura, Alan Penn and Sam Griffiths. 2012. Space Syntax Economics: Decoding Accessibility Using Property Value and Housing Price in Cardiff, Wales. In *Proceedings of the Eighth International Space Syntax Symposium, Santiago de Chile*. pp. 1–19.

Openshaw, S. 1984. “Ecological Fallacies and the Analysis of Areal Census Data.” *Environment and Planning A: Economy and Space* 16(1):17–31.

Waddell, Paul, Brian J. L. Berry and Irving Hoch. 1993. “Residential Property Values in a Multinodal Urban Area: New Evidence on the Implicit Price of Location.” *The Journal of Real Estate Finance and Economics* 7(2):117–141.

Code Appendix

201347125

This version was compiled on May 13, 2019

```
knitr::opts_chunk$set(
  eval = FALSE,
  message = FALSE,
  echo = TRUE,
  results = FALSE,
  warning = FALSE
)
```

1. Libraries

```
library(tidyverse)
library(ggplot2)
library(sf)
library(tmap)
library(tmaptools)
library(ggmap)
library(smoothr)
library(extrafont)
library(ggspatial)
library(gstat)
library(GISTools)
library(raster)
library(rasterVis)
library(broom)
library(kableExtra)
library(viridis) # colour schemes
library(spdep)
```

2. Data Preparation

2.1. Load in Data.

```
property <- read_sf("./data/house_transactions/liv_house_trans.shp")
property$price_log <- log(property$price)

# exclude 'other', removes very expensive non residential sales
property <- property[property$type != "O", ]

liv <- read_sf("./data/house_transactions/liv_outline.shp")
imd <- read.csv("./data/house_transactions/E08000012.csv")

lsoa <- read_sf("./data/Liverpool_lsoa11.shp")
```

2.2. Custom Map Theme.

```
theme_map <- function(...) {
  theme_minimal() +
    theme(
      axis.line = element_blank(),
```

```
      axis.text.x = element_blank(),
      axis.text.y = element_blank(),
      axis.ticks = element_blank(),
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      panel.border = element_rect(colour = "grey50", fill = NA, size = 0.5),
      ...
    )
}
```

3. Data Analysis

3.1. KDE.

```
## Function to choose bandwidth according to Bowman and Azalini / Scott's rule
## Adapted for sf
choose_bw <- function(spdf) {
  X <- st_coordinates(spdf)
  sigma <- c(sd(X[, 1]), sd(X[, 2])) * (2 / (3 * nrow(X)))^(1 / 6)
  return(sigma / 1000)
}

kde_property <- smooth_map(property, cover = liv, bandwidth = choose_bw(property))

area_threshold <- units::set_units(1000, km^2)
kde_property$polygons <- kde_property$polygons %>%
  st_transform(4326) %>%
  fill_holes(area_threshold)

bg <- liv %>%
  st_transform(4326) %>%
  st_union() %>%
  fill_holes(threshold = area_threshold)

# change level names to colours for slight ease
bins <- kde_property$polygons$level
bincol <- c("#ebebe5", "#dfe3ee", "#8b9dc3", "#3b5998")
kde_property$polygons$level <- bincol

kde_map <- ggplot() +
  annotation_map_tile(data = bg) +
  geom_sf(data = bg, size = 1, col = "black", fill = NA) +
  geom_sf(
    data = kde_property$polygons,
    aes(col = bins, fill = bins),
    color = "black", alpha = 0.75
  ) +
  scale_fill_manual(values = bincol) +
  theme_map() +
  theme(
    legend.title = element_blank(),
    legend.position = "bottom",
    legend.spacing.x = unit(.1, "cm"),
    legend.text = element_text(margin = margin(t = 0), size = 5)
  ) +
  guides(fill = guide_legend(
    title = "Cyl",
    label.position = "left"
  ))
```

```
kde_map
```

3.2. IDW Interpolation.

```
property_sp <- property %>%
  st_transform(4326) %>%
  as_Spatial()

bg_sp <- bg %>%
  as_Spatial()

set.seed(9)
# Create an empty grid where n is the total number of cells
grd <- as.data.frame(spsample(property_sp, "regular", n = 50000))
names(grd) <- c("X", "Y")
coordinates(grd) <- c("X", "Y")
gridded(grd) <- TRUE # Create SpatialPixel object
fullgrid(grd) <- TRUE # Create SpatialGrid object

# Add P's projection information to the empty grid
proj4string(grd) <- proj4string(property_sp)

# Interpolate the grid cells using a power value of 2 (idp=0.8)
P.idw <- idw(price_log ~ 1, property_sp, newdata = grd, idp = 0.8)

# Convert to raster object then clip to Texas
r <- raster(P.idw)
r.m <- mask(r, bg_sp)
```

```
# need to fortify to plot tile and border on same ggplot
bg_points <- fortify(bg_sp)
idw_map <- gplot(r.m) +
  annotation_map_tile(data = bg) +
  geom_raster(aes(fill = value), interpolate = TRUE, alpha = .8) +
  geom_path(data = bg_points, aes(x = long, y = lat), size = 1) +
  scale_fill_gradientn(
    colours = c("white", "red", "black"), limits = c(11.24, 11.85),
    name = " ", na.value = NA
  ) +
  theme_map() +
  theme(
    legend.key.width = unit(1, "cm"),
    legend.position = "bottom",
    legend.spacing.x = unit(.1, "cm"),
    legend.text = element_text(margin = margin(t = 0.5), size = 5),
  )
```

```
idw_map
```

4. Models

4.1. OLS Model.

```
property$pc <- as.character(lapply(
  strsplit(as.character(property$pcds), split = " "), "[", 1
))

property <- merge(property, imd)
```

```r
## convert p vals to asterisks
p_val <- function(x) {
  symnum(x, corr = FALSE, na = FALSE, cutpoints = c(
    0, 0.001, 0.01, 0.05, 1
  ), symbols = c("***", "**", "*", " "))
}
```

4.1.1. Table 1.

```r
# tell Kable to not show anything for NA values
options(knitr.kable.NA = "")
kable(ols_m1,
  digits = 2, caption = "OLS Linear Regression Model",
  linesep = "",
  longtable = FALSE, booktabs = TRUE,
  format = "latex",
  col.names = c(
    "Variable",
    "Model Estimate",
    "Standard Errors"
  )
) %>%
  # footnote for significance
  footnote(
    general_title = "",
    general = c(
      paste("† R Squared =", r),
      "*** Significant at the 0.001 level"
    )
  )
```

4.2. Spatial Fixed Effects.

```r
property$one <- 1
property$detached <- 0
property[property$type == "D", "detached"] <- 1
property$newB <- 1
property[property$new == "N", "newB"] <- 0

m2 <- lm("price_log ~ pc + newB + detached + imd_score - 1", property)

# tidy from broom package changes summary information into data.frames
t_m2 <- tidy(summary(m2))
# select only the important columns
tail(t_m2)

# Change p values into vectors for each model
p <- p_val(t_m2$p.value)

# Round statistic values to 2 digits and change to vector as well, keep zeros
t <- sprintf("%.2f", t_m2$estimate)

# Join statistic values and p asterisks
t_m2$estimate <- paste(t, p)

# Find R squared % for each model
r <- summary(m2)$r.squared * 100

# Round them to 2 digits
r <- round(r, 2)
# rename row names of table
```

```
t_m2$term <- as.character(lapply(
  strsplit(as.character(t_m2$term), split = "pc"), "[", 2
))
t_m2 <- t_m2[1:3]

val <- c("New", "Detached", "IMD Score")

t_m2$term[is.na(t_m2$term)] <- val
```

4.2.1. Table 2.

```
# tell Kable to not show anything for NA values
options(knitr.kable.NA = "")
kable(t_m2,
  digits = 2, caption = "Spatial Fixed Effects Model",
  linesep = "",
  longtable = FALSE, booktabs = TRUE,
  format = "latex",
  col.names = c(
    "Variable",
    "Model Estimate",
    "Standard Errors"
  )
) %>%
  # footnote for significance
  footnote(
    general_title = "",
    general = c(
      paste("† R Squared =", r),
      "*** Significant at the 0.001 level"
    )
  )
```

4.2.2. Mapping Intercepts.

```
int_map <- tidy(summary(m2))

int_map$pc <- as.character(lapply(
  strsplit(as.character(int_map$term), split = "pc"), "[", 2
))

int_map <- merge(property, int_map, by = "pc")
int_map <- lsoa %>%
  st_join(int_map) %>%
  group_by(LSOA11CD.x) %>%
  summarise(estimate = mean(estimate))
```

```
no_classes <- 6 # specify 6 bins
labels <- c() # empty vector

# find quantiles for each msoa
quantiles <- quantile(int_map$estimate,
  probs = seq(0, 1, length.out = no_classes + 1), na.rm = T
)

# define custom labels, remove decimal etc
labels <- c()
for (idx in 1:length(quantiles)) {
  labels <- c(labels, paste0(round(quantiles[idx + 1], 2)))
```

```
}

# remove final label
labels <- labels[1:length(labels) - 1]

# new variable in dataset with quantiles
int_map$estimate_qu <- cut(int_map$estimate,
  breaks = quantiles,
  labels = labels,
  include.lowest = T
)
```

```
int_map <- st_transform(int_map, 4326)
# remove slivers
int_map <- st_buffer(int_map, dist = 0.001)

map_int <- ggplot() +
  annotation_map_tile(data = bg) +
  geom_sf(data = int_map, aes(fill = estimate), colour = NA, size = 0) +
  geom_path(data = bg_points, aes(x = long, y = lat), size = 1) +
  theme_map() +
  theme(
    legend.key.width = unit(1, "cm"),
    legend.position = "bottom",
    legend.spacing.x = unit(.1, "cm"),
    legend.text = element_text(margin = margin(t = 0.5), size = 5)
  ) +
  scale_fill_viridis(
    option = "magma",
    direction = -1,
    name = " "
  )
```

```
map_int
```

4.3. Spatial Lag Model.

```
# aggregate by lsoa for imd score inclusion
prop_lsoa <- lsoa %>%
  st_join(property, by = LSOA11CD) %>%
  group_by(LSOA11CD.x) %>%
  summarise(
    price_log = mean(price_log),
    newB = sum(newB) / length(newB),
    imd_score = mean(imd_score),
    detached = sum(detached / length(detached))
  ) %>%
  na.omit()

# create queen weights for imd score using lsoa
lsoa_w <- poly2nb(prop_lsoa)
lsoa_lw <- nb2listw(lsoa_w)
prop_lsoa$imd_score_w <- lag.listw(lsoa_lw, prop_lsoa$imd_score)
```

```
m4 <- lm("price_log ~ newB + detached + imd_score + imd_score_w", prop_lsoa)
```

```
# spatial lag of imd_score
```

```r
property$imd_score_w <- lag.listw(hknn, property$imd_score)
```

```r
m4 <- lm("price_log ~ new + detached + imd_score + imd_score_w", property)
```

```r
# tidy from broom package changes summary information into data.frames
sp_m4 <- tidy(summary(m4))
# select only the important columns

# Change p values into vectors for each model
p <- p_val(sp_m4$p.value)

# Round statistic values to 2 digits and change to vector as well, keep zeros
t <- sprintf("%.2f", sp_m4$estimate)

# Join statistic values and p asterisks
sp_m4$estimate <- paste(t, p)

# Find R squared % for each model
r <- summary(m1)$r.squared * 100

# Round them to 2 digits
r <- round(r, 2)

# rename row names of table
sp_m4$term <- c("(Intercept)", "New Build", "Detached", "IMD Score", "IMD Score (Lag)")

sp_m4 <- sp_m4[1:3]
```

4.3.1. Table 3.

```r
# tell Kable to not show anything for NA values
options(knitr.kable.NA = "")

kable(sp_m4,
  digits = 2, caption = "Spatial Lag Model",
  linesep = "",
  longtable = FALSE, booktabs = TRUE,
  format = "latex",
  align = c("l", "c", "c", "c"),
  col.names = c(
    "Variable",
    "Model Estimate",
    "Standard Errors"
  )
) %>%
  # footnote for significance
  footnote(
    general_title = "",
    general = c(
      paste("† R Squared =", r),
      "*** Significant at the 0.001 level"
    )
  )
```