Admin-based floor area statistics for England and Wales using model-based imputation

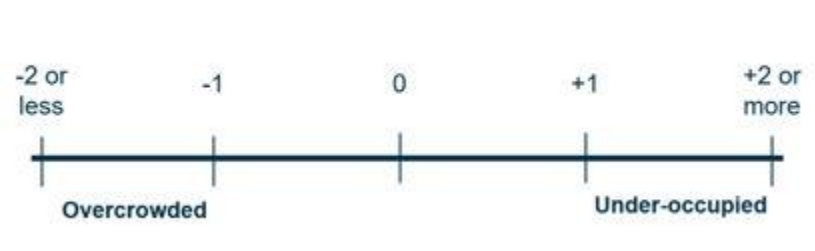
Stephan Tietz
Office for National Statistics, UK
Admin.Based.Characteristics@ons.gov.uk

5 October 2022

Overcrowding & floor area

- In the UK, overcrowding commonly measured using occupancy ratings (e.g. Bedroom Standard)

(Bed)room occupancy rating =
Actual number of (bed)rooms – Hypothetical number of bed(rooms) required



- Rooms can vary in size or
  actual use differs to the intended/recorded use
- “Available living space” (= floor area per person)
  can provide a more complete picture

Data source - Valuation Office Agency (VOA)

- VOA captures data about properties for Council Tax banding purposes
- Covers all residential properties in England and Wales
- But two different measures of floor area:
  - Reduced Cover Area (RCA) for houses and bungalows
  - Effective Floor Area (EFA) for flats and maisonettes

Producing floor area statistics

- Floor area statistics produced through surveys with small sample size (e.g. English Housing Survey) - limits sub-regional analysis
- VOA floor area generally follows expected distribution by property type with complete geographical coverage

Can we harmonise the two different VOA floor area measures across property types?

Data source - Energy Performance Certificate (EPC)

- EPC provides a measure of energy efficiency of a property (since 2007)
- Single measure of floor area: Total Floor Area (TFA)
- But covers only 50-60% of residential properties

Single measure of available living space

- Aim to produce a single measure of available living space for residential properties in England and Wales
- Build a statistical model that uses the geographical completeness of VOA floor area measures (RCA and EFA) to predict the EPC floor area measure (TFA)

Model exploration

- 57.2% % of residential properties have an EPC and could be linked to VOA data
- Removed outliers: <5m^2, 1st & 99th percentile of VOA - EPC
- Simple linear regression that uses VOA floor area to predict EPC floor area gives $R^2 = 0.84$
- Variables explored for model
  - Property characteristics: property type, floor area measure flag, number of rooms, number of bedrooms, number of bathrooms, property age
  - Geographical variables: country, region/GOR, local authority, urban/rural

Final model selected

- Best performing model is multiple linear regression that uses VOA floor area, VOA floor area measure flag (RCA vs EFA) and country (England & Wales) to predict EPC floor area

RMSE	R^2	ADJUSTED R^2	COEFFICENTS			INTERCEPT
14.61	0.86	0.86	VOA floor area	VOA floor area measure flag	Country	14.02
			0.93	-17.91	1.85	

Model performance (m^2 of actual property size)

- 41% of properties within +/-5m^2
- 69% of properties within +/-10m^2
- 88% of properties within +/-20m^2

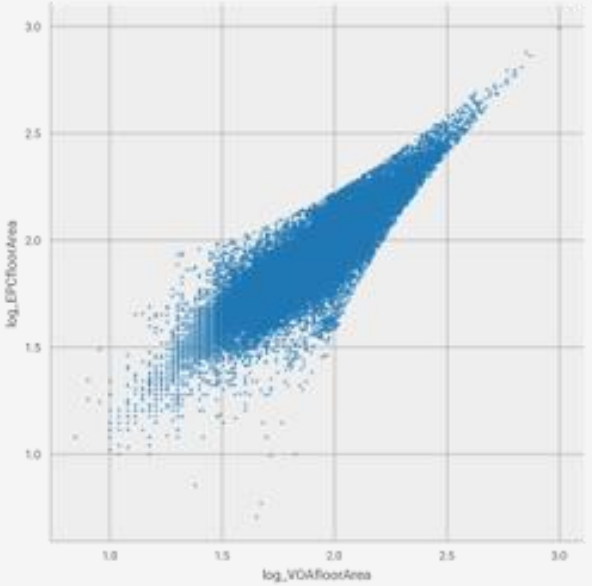
Model performance (% of actual prop. size)

- 31% of properties within +/-5%
- 60% of properties within +/-10%
- 84% of properties within +/-20%

Sources of variance

- Two primary factors causing the variance observed
  - Data quality
  - Differences in property structure that cannot be fully accounted for in the data available

Log transformed
VOA vs EPC floor area



Sources of variance: Data Quality

- Using structural equation modelling to estimate the measurement error of VOA and EPC data
- EPC data tends to have a larger measurement error for the floor area variable (6.5%) compared to VOA data (2.5%)
- Improvements in the quality of the floor area variable in the EPC data would improve the performance of the model

Sources of variance: Property Structure

- English Housing Survey (EHS) and Welsh Housing Conditions Survey (WHCS) use a statistical model to derive the useable floor area of an address from detailed measurements of the main rooms
- Linked to VOA & EPC data: highest correlation between any of the floor area measures was 0.91
- Differences in property structure that are too difficult to consistently account for in floor area measurements

Conclusions

- Model does not produce harmonised address-level floor area estimates of high enough statistical quality to provide an alternative measure of overcrowding that is comparable across all property types
- Observed variance caused by quality of the EPC floor area variable when used for statistical purposes; and differences in property structure that cannot be fully accounted for in the data available

Future developments

- ONS website release on 26th October
- Explore methods that use floor area to identify overcrowded or under-occupied addresses within the same property type
- Evaluate alternative ways of harmonising floor area across property types by using different statistical modelling methods such as machine learning, or through the inclusion of additional data sources

Contact: admin.based.characteristics@ons.gov.uk