

## READ ME FILE

**Title:** The Apartment Shortage

**Authors:** Keaton Jenner and Peter Tulip

### Description

This 'read me' file contains details of the data and Stata code included in this archive that were used to generate the main results reported in RDP 2020-04.

### Data Files

There is a separate Excel workbook for each of these:

- rdp-2020-04-graph-data.xlsx
  - Plotting data for the figures appearing in the paper
- ABS Building Activity Survey Unpublished
- ABS Building Approvals
- Producer Price Indexes
- Australian Statistical Geography Standard (statistical areas 2 and 3)
- Not included due to licencing, size and/or commercial agreement:
  - Unit record property sales data can be purchased from CoreLogic
  - PSMA's Geocoded National Address File (G-NAF) can be downloaded from <https://psma.com.au/product/gnaf/>
  - Point of Interest data (for Appendix D) can be downloaded from [Spatial Services](#)

### Excel Work Files

- Cost Decomposition.xlsx
  - calculations underpinning Tables 3, 4, 5 and A3 with related data
- height.xlsx
  - Calculations for Figures 7 and 8 and Table 6
- Historical.xlsx
  - historical estimates of prices and marginal costs, for Figure 6
  - updates of 2016 estimates to 2018
  - The first tab, titled Summary, provides key details
- SA3.xlsx
  - Data disaggregated by SA3
  - The first tab, titled Summary, provides key details
- SA3 MC adj factors.xlsx
  - Calculates average-to-marginal cost adjustments by SA3 used in the construction of Figure 4
  - This file is in the Data/ABS folder as it is imported by STATA programs
- Disclaimer: we are aware that these files are not polished. There is some repetition, some entries are hard-coded and many series are not described well, if at all.

## Data Sources for Tables and Figures

The following tables indicate where estimates in tables and figures come from:

Data sources for tables		
Table no	File name/Source	Tab
1	Cost Decomposition	Tables 1, 3 and 5
2	STATA	
	Historical	Apartment prices
3	Cost Decomposition	Tables 1, 3 and 5
4	Cost Decomposition	Table 4
5	Cost Decomposition	Tables 1, 3 and 5
6	height	data
A1	STATA	
	Historical	Apartment prices
A2	Building Activity Survey/ STATA	ABS BAS_2013-2018_a
A3	Cost Decomposition	Table 9
B1	Cost Decomposition/ Ciesielski (2019)	Tables 1, 3 and 5
D1	STATA	
E1	STATA	

Data sources for figures		
Figure No	File name/Source	Tab
1	PowerPoint	
2	STATA	
3	rdp-2020-04-graph-data	Figure 3
4	rdp-2020-04-graph-data	Figure 4
5	SA3	Efficient Height calcs
6	Historical	Adjusted costs R7 Prices M12
7	height	Fig 7
8	height	Fig 8
9	SA3	Efficient Height calcs
A1	Historical	Building heights

## Stata Files

The results reported in this RDP were generated using Stata 16, including the use of several additional functions that need to be installed (instructions included in code files): `geoinpoly`, `geodist`, `spmap`, `nearstat`, `sepscatter`, `trimmean`, `shp2dta`.

The following programs are included in the code folder:

Apartment Shortage Master File.do – this master file installs functions (as required), initialises settings, and then runs each of the following code files needed to reproduce the detached housing results in order.

0. units - import & cleaning.do – imports and cleans the CoreLogic property sales and characteristics data.

1. units - create regional data files.do – processes the data and creates a separate data file for each city.

1.1 units - identify apartments.do – identifies street addresses with apartments (rather than other types of units).

X = 1 (Sydney), 2 (Melbourne), 3 (Brisbane)

2.X units - (X) – cleaning & stats.do – imports merged dataset. Trims out non-apartment units (e.g. townhouses) and other outliers. Calculates average sale prices at a city-wide and SA3 level.

3.(X)\_maps.do – imports and cleans ABS Building Approvals by SA3. Merges with cleaned CoreLogic dataset and ABS SA3 boundary data. Calculates difference between sale price and supply cost by SA3 and produces chloropleth map of results.

4.0 ABS cost regressions.do – imports and cleans unpublished data from the ABS Building Activity Survey which include average construction costs and building heights from 2013-2018 (aggregated). Estimates relationship between construction costs and building height.

5.0. K&T detached\_import & cleaning.do – imports and cleans CoreLogic property sales and characteristics data for detached sales.

5.1. K&T detached\_create regional data files.do – process detached property sales and creates a separate data file for each city.

5.2. K&T detached\_stats.do – trims outlier observations and calculates average detached sale prices at a city-wide and SA3 level

6.0 SA3 file conversion.do – imports CoreLogic detached house price data by SA3. Calculates cost per square metre and exports output for each city to a separate excel file.

7.0. units - Sydney - GNAF\_spatial.do – imports and cleans address file data for Sydney from G-NAF. Generates 'density' variable calculated as number of unique unit numbers within a building.

7.1. units - Sydney - GNAF\_merge\_spatial.do – merges G-NAF density variable with CoreLogic apartment sales for Sydney.

7.2. units - Sydney - spatial vars.do – imports and cleans Points of Interest data from Spatial Services NSW. Merges with combined CoreLogic/G-NAF data. Calculates distance (in kms) between Sydney apartments and Points of Interest (e.g. beaches, parks, train stations).

7.3 units - Sydney - hedonic regression.do – imports dataset construction in 7.2.do and estimates a hedonic regression of Sydney apartment prices including spatial controls. These estimates are used in Section 7 and reproduced in Appendix D.

#### **Additional notes on code files**

File paths for 'datadir' and 'workingdir' must be updated in the two master files. The datadir path should link to a folder containing the sales and property characteristics comma separate values (.csv) files available for purchase from CoreLogic. The workingdir path should be set to the file path of this zip folder, plus \Detached for the detached master file, and \Apartments for the units master file. Comment in/out the 'ssc install' lines of code as necessary to install/not install these functions. Loading in, processing, and storing the CoreLogic, G-NAF and Spatial Services NSW data files requires a lot of memory – this code has been tested only on computers with 16 gigabytes of memory.

5 August 2020