Australian Government
Department of Agriculture
ABARES

Measuring Australian broadacre farmland value:

Phase 1 - Statistical infrastructure

Will Chancellor, Shiji Zhao, Lucy Randall, Kenton Lawson and Khanh Hoang

Research by the Australian Bureau of Agricultural and Resource Economics and Sciences [ABARES]

Technical report 19.7
December 2019

Cataloguing data

This publication (and any material sourced from it) should be attributed as: Chancellor, W., Zhao, S., Randall, L., Lawson, K. and Hoang, K. 2019, *Measuring Australian broadacre farmland value: Phase 1 – Statistical infrastructure,* ABARES working paper, Canberra, December. CC BY 4.0. http://doi.org/10.25814/5de84a4ff6e09

ISBN 978-1-74323-458-7

ISSN 189-3128

Department of Agriculture
GPO Box 858 Canberra ACT 2601
Telephone 1800 900 090
Web agriculture.gov.au

Acknowledgements

The Corelogic dataset used for this research was partially funded [approx. 50%] by the Grains Research and Development Corporation (GRDC). In addition, the authors acknowledge the feedback provided from Terence Farrell of the GRDC.

The authors thank David Galeano and Peter Martin of ABARES for their contribution to this report. They also gratefully acknowledge advice and feedback provided by Professor Kevin Fox. Insights provided by members of the Organisation of Economic Co-operation and Development (OECD) Farm Level Analysis Network are importantly acknowledged by the authors.

Contents

Tables

Figures

Maps

Boxes

Abstract

This paper reports our preliminary findings on the construction of hedonic models to explain the determinants of the value of agricultural land (such as land size, soil condition and location) to estimate values of Australian farmland. The models are based on a large administrative dataset of land transactions (supplied by CoreLogic) and data from other sources. This report focuses on examining the efficacy of the CoreLogic data – its quality, usability and feasibility to be linked with other data sources for hedonic modelling -- rather than economic analysis of farmland value. As a 'technical' report, it aims to share our experience of using administrative land value data to estimate agricultural land values and seek feedback on our work.

We first transform the CoreLogic data (from tabular data to spatial polygons), so that various scientific datasets can be overlaid and linked to individual land parcels. Such spatial linkages allow for the derivation of farm specific characteristics (including land slope, soil characteristics, water cover etc.), overcoming a major challenge in the development of a national hedonic model for Australian broadacre farmland valuation. Our analysis uses a series of stratified hedonic models to identify drivers of broadacre farmland value and explore whether these drivers differ by region, farm size and price level. Results indicate that rainfall, temperature, distance from towns, characteristics of the farm house (if present), farm size, production type, and infrastructure such as buildings play an important role in driving the value of Australian broadacre farmland.

The results from hedonic modelling can be used to predict values of any farmland based on this 'characteristic' information. The coefficients of a sound hedonic model have wider applications in the analysis of agricultural economics and policy issues. For example, as they explicitly show the contributions of the 'characteristics' to the formation of farmland values, it is possible to use hedonic modelling to analyse the underlying drivers of price movements. This generates an important linkage between economic and environmental statistics, which may contribute to the development of environmental-economic accounting systems. Hedonic models may also be used to estimate real (historic) or hypothetical (future) farmland values under various simulated scenarios including economic incidents (e.g. a rise in interest rate), policy changes (e.g. structural reform) or external events (e.g. drought).

This report is a summary of the first stage of our research on farmland values using hedonic techniques. It focuses on the dataset that we have developed for hedonic modelling. While this study provides important insights into the drivers of Australian farmland value, it also serves as a foundation for future data and model refinement — progressing towards our long term vision to analysing economic and policy issues. This progress is greatly facilitated by our initial data transformation and linkage work, connecting economic data to environmental factors (among others) at the parcel level. The target audience of this report are those specialised in the analysis of agricultural farmland value and those interested in the integration of economic and environmental data. We expect that publication of these findings will facilitate communication with like-minded researchers to share experience and work towards developing improved methodologies for the analysis of agricultural land values.

Introduction

Farmland is an important asset for Australian farmers, agricultural industries and the nation as a whole. First, it is a productive asset which is essential for farms to run agricultural businesses, generate revenue and sustain their livelihood. For most farmers, land is their most important asset. Agricultural land is also an integral part of the Australian ecosystem and has a significant environmental value, of which farmers are responsible for managing approximately 385 million hectares (or 58 per cent of Australia).

Information about Australian farmland value is scarce. Values of farms are not observable unless farms are sold on the market and even then, farms are traded infrequently. ABARES produces price indexes of broadacre farmlands (Martin and Topp 2019) for the three ABARES agricultural zones (high rainfall, wheat-sheep and pastoral). However, while these statistics are informative, they have limitations. For example, as the statistics were derived from ABARES Australian Agricultural and Grazing Industry Survey (AAGIS), it is difficult to use the raw data to compile statistics at lower aggregation levels (such as for small farming areas or agricultural regions within a state) due to the limited sample size. Likewise, the index does not cover land used for producing dairy and other agricultural commodities. More importantly, it is not feasible to use the statistics to identify and quantify the determinants of land values (for example, between locations and farm types) and growth over time.

Hedonic models are a feasible and cost-effective statistical tool to fill this information gap. They fit the land value and other data into an equation (or model) using regression or machine learning techniques. In the case of regression, market prices are regressed on the farmland characteristics and the resulting coefficients can be interpreted as the implicit ‘prices’ of the characteristics (embedded in the market prices of farmland). Hence, based on some basic information about the properties (e.g. size and location etc.), a sound hedonic model would enable the estimation of the value of any farmland, including those which have not been sold recently.

While hedonic models have been used extensively to estimate market values of residential properties (Cho 1996; Conniffe and Duffy 1999)[1] and consumer goods (Triplett 2004), to our knowledge, comprehensive national hedonic farmland models for Australia do not yet exist. There are several possible explanations for this. The characteristics of farmland are more complex and difficult to identify, define and numerically measure compared to residential properties. There are also many different agricultural production systems, making the stratification of farmland more difficult. Finally, the volume of farmland transactions is smaller compared with the residential market, limiting the applicability of hedonic models to date.

The ‘characteristics’ (physical, natural, geographical and socioeconomic) expected to influence farmland values include terrain features, localised climate conditions (rainfall and temperature),

[1] Australian studies include Hansen (2006) and Soriano (2008)

transport access, soil conditions, buildings on the land, production type, among other factors. In addition to these farm specific factors, broader influences such as commodity prices, interest rates, productivity growth, economic growth, market sentiment, levels of rural debt, and the supply and demand for urban real estate assets may also affect the value of farmland.

Using CoreLogic land transaction data and other data ABARES has constructed hedonic models to explain the determinants of the value of agricultural land to estimate values of Australian farmland. The CoreLogic dataset has comprehensive coverage, containing farmland transactions in all States and Territories since 1975 and includes information on transaction values, land size, location, purposes of land use and residential properties on the land (number of bedrooms etc.).

Before constructing the hedonic models, the CoreLogic data required significant data transformation and filtering. The CoreLogic data, contained a total of 700,424 transaction records, including transactions which were obviously not farms. For example, some observations were labelled as 'construction sites', 'mines' and even 'church buildings' etc. Many of the records removed were for agricultural purposes (e.g. dairy, tomato farming, pineapples), however were excluded to fit our scope of dryland broadacre farms. This resulted in a final dataset of 166,994 broadacre farmland transactions covering the period from 1975 to 2019.

The CoreLogic data include information about the location of the land transaction and land parcel identifiers. Using this information, ABARES formed data linkages between these parcel identifiers and state government cadastral datasets, thereby enabling the derivation of parcel shape information for each CoreLogic land transaction record.

The CoreLogic data did not include data on a range of other possible explanatory variables such as climate data, land gradient, water cover and many others. Therefore, ABARES sourced and included data at the farm level on these other possible explanatory variables. The final data set included topography data from the Geoscience Australia Shuttle Radar Topographic Mission (SRTM) digital elevation model, Water Observations from Space (WOS), average rainfall and temperature from the Bureau of Meteorology (BoM) Australian Water Availability Project (AWAP) grids, among others. This means that physical property features (such as soil and climate attributes) can be derived at farm level and applied as explanatory variables to farmland value in the hedonic models.

After developing the dataset and a prototype model of co-variates to explain our dependent variable (land price per hectare), we apply a series of robust (Hamilton, 1991) and quantile regressions in order to observe the relationship between farmland value and potential explanatory variables by region, production type, size category and price segment. In doing so, we are able to identify some factors driving Australian broadacre farmland value as well as the magnitude and direction of these drivers. This study thereby makes considerable progress in use of hedonic method to measure Australian farmland values and quantify regional, price and size specific drivers.

The hedonic model and underlying data set have a number of potential applications. The underlying data, for example, can be used to generate price index(s) of Australian farmland at the national, state and regional levels, using conventional methods[2]. The model can be used to identify drivers of farmland price movements at the national, state or regional level as well as predict the impacts of changes in these variables on future farmland prices. For example, it may be possible to estimate the impact of commodity prices and interest rate movements on farmland values.

Future development of the hedonic models may allow assessments of other policy issues. For example, evaluating the impact of infrastructure investments such as construction of roads, railways and dams on the value of farm land. This could also include assessing the impact of drought and climate change on farmland values. Another possibility is to explore the relationship between natural capital and farm values.

Section 1 presents detailed information about the CoreLogic records, issues with the data and how it is linked to information from other sources. Section 2 explains the hedonic methods used in this preliminary study and Section 3 provides a short summary of the results of hedonic estimation. Section 4 concludes.

[2] For example, the stratification method has been used by the Australian Bureau of Statistics (ABS) and other official statistical agencies in the construction urban residential house price indexes.

1 Data

Raw CoreLogic data: an overview

The analysis in this paper is based on a dataset of CoreLogic property transactions, which are defined loosely as "all properties that are identified as for farming or agricultural use". CoreLogic obtains these transaction records from State Valuer General administrative datasets, and applies processes to integrate and maintain these transaction records. The raw 'agricultural' dataset provided by Corelogic contains 349,217 unique properties recorded between 1900 and 2019, made up of 700,424 property transactions.

Table 1 CoreLogic data variables

Number	Data field	Definition
1	Property ID	Unique record key within the core database for the property.
2	Real Property Description	The legal parcel(s) description of the property, depending on the scheme adopted for each state.
3	Lot Number	Lot Number component of the parcel's description. NSW, VIC, QLD, WA, SA, TAS, NT only
4	Full Property Address	Property Address
5	Property Type	CoreLogic identified category for the property such as House, Unit, Flats, Land, Business i.e. House, Unit, Flats, Business, Commercial, Community, Farm, Land, Storage Unit
6	Property Type Minor	Corelogic minor category for properties such as Multi Storey, Duplex, One story/Lowset, etc.
7	Primary Land Use	The Primary Land Use of the property such as single Unit Dwelling, House etc.
8	Latitude	The geographical latitude of the property.
9	Longitude	The geographical longitude of the property.
10	Bedrooms	The most recently recorded bedrooms count.
11	Bathrooms	The most recently recorded count of bathrooms for the property (inclusive of ensuites).
12	Land Area	Total size of the parcel/s in square metres.
13	Transfer ID	Unique record key within the core database for the transfer
14	Contract Date	Contract date of transfer which indicates the date on which the sale price was contractually committed between a vendor and a purchaser.
15	Transaction Date	Contract Date for states were VG Contract Date is provided include NSW, VIC, QLD, WA, TAS, ACT only
16	Contract Price	A proxy Contract Date with Settlement Date substituted for states where no VG Contract Date is provided. Allows for ordering transfers by the time that the transfer occurred.
17	Multi Sale	Sale price of transfer indicating the consideration for the property changing ownership (if available)

Source: CoreLogic

Definition of farmland prices

The variables provided (Table 1) in the base dataset are used in the regression analysis for the dependent variable (contract price per hectare [land area is converted from square meters to hectares]), some basic explanatory variables (bedrooms, bathrooms), years (which can be used as dummy variables), and linking variables (Lot number, geographical coordinates). The price per hectare calculation is complicated by 'multi sale' transactions, where several land parcels (usually different sizes) are grouped for sale. A single contract price is supplied for the 'multi sale', therefore simply calculating price per hectares as a division of contract price and land area (in hectares) is an incorrect representation of 'price per hectare'. In this study, we use multi sale transactions to aggregate land area by the sum of all land parcels within a multi sale, and then obtain 'price per hectare' as the total multi sale contract price divided by the total multi sale land area. A limitation of this approach is that we are treating all land transactions within a multi sale as having equal value (when in reality this may not be the case). Future iterations of this work will look to refine the treatment of multi sale transactions using quality adjustment weightings.

The main CoreLogic dataset is accompanied by a linking key dataset (see Appendix Tables A3 and A4), which contains variables needed to join to state level cadastral parcel boundaries. After deriving spatial parcel boundaries for each CoreLogic property, we can begin linking various spatial farm level variables to the CoreLogic dataset as outlined in the section to follow (Spatial transformation and variable construction).

Removing non-broadacre farm records

An initial review of the dataset indicated a large number of records unlikely to be broadacre farmland, highlighting typical 'administrative data issues' including duplicate records and missing values. Issues included a high number of hobby farms, mine sites, urban residences, and non-broadacre farms such as irrigated horticulture. Our approach strictly refines the data scope so that the transaction records closely resemble broadacre farmland. The raw dataset contained 700,424 property transactions (for 349,217 unique properties) between 1900 and 2019; however after completing our data cleaning, we are left with 166,994 transactions between 1975 and 2018. This cautious approach ensures that we are analysing true broadacre farmland and provides scope to increase the sample size with further data testing in future iterations. Figure 1 (on page 12) illustrates the average 'price per hectare' and sample size in the raw dataset. The volatility and magnitude demonstrates a need for careful refinement of the data scope.

Preliminary testing of the dataset revealed a high number of transactions where either contract price was $0 or land area was missing, resulting in a large number of '$0/[not available]' price per hectare. We expect that these low or zero values are either due to data/reporting errors (missing contract price or missing land area), or due to reasons in relation to legal definition (i.e. family transactions, transactions within trusts). Data testing also reveals unrealistically expensive properties on a 'per hectare basis'. Further investigation indicated that very small land parcels tend to experience erratic or unrealistic prices per hectare – many of which appear to be mine sites, commercial buildings or other non-broadacre farmland. For this reason, we remove 'quantile extremes', and drop transactions with small land areas.

The prevalence of hobby farms provides another challenge. These hobby farms tend to be operated for the purpose of lifestyle (rather than profit), tend to be situated on small and expensive land parcels, are often located in close proximity to urban areas, and usually include expensive buildings or homesteads. A simple check of hobby farms *(CL_Property_Type_Minor = "Hobby")* revealed that average price per hectare for these farms was $531,112 — far in excess of the most expensive broadacre farmland in Australia. We note that hobby farms may also be identified where CL_Primary_Land_Use is equal to “lifestyle”.

In addition to hobby farms, the prices of horticultural farms tends to be incomparable to broadacre farmland – due to the high capital costs associated with irrigation and perennial crops (i.e. fruit and nut trees). Similarly, dairy farms are also difficult to analyse alongside broadacre farmland due to the high proportion of fixed capital value in total sale price.

Therefore, we must address the following anomalies in setting our scope strictly to broadacre farms:

- Excessively high and excessive low price per hectare
- Excessively small farm size (in hectares)
- Non-broadacre farms according to CoreLogic identifier labels.

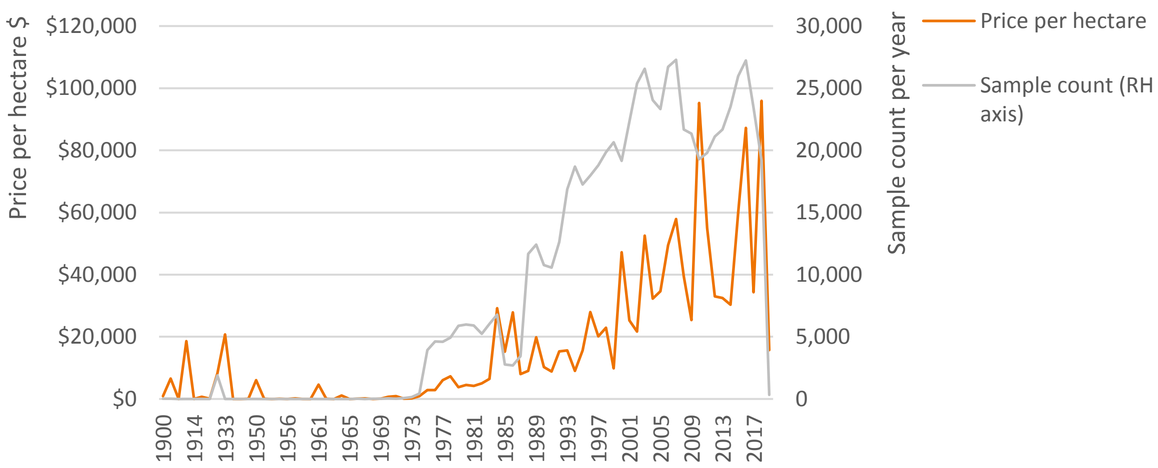
Further treatment of CoreLogic data

In order to address the above data issues, we apply a systematic data scope refinement process after linking the explanatory variables (discussed in the following section). This data linking process addressed some of the issues above (as well as limiting the year range to between 1975 and 2018) and reduces the number of transactions from 700,424 to 683,014. In order to ensure that the records in our analysis contains exclusive broadacre farms, further refinement is needed and the number of transactions in the final clean dataset is 166,994.

- Price per hectare:
    - (drop if price per hectare = 0) (111,419 observations deleted)
- Broadacre label trim [dummy variables] (see Appendix Table A5 and A6):
    - drop if cl_property_type ≠ "Farm" (180,406 observations deleted)
    - drop if cl_property_type_minor = "Hobby" (84,296 observations deleted)
    - drop if lifestyle = 1 (18,399 observations deleted)
    - drop if dairy = 1 (6,374 observations deleted)
    - drop if Irrigated = 1 (10,326 observations deleted)
    - drop if sugar = 1 (8,257 observations deleted)
    - drop if horticulture = 1 (8,936 observations deleted)
- Quantile extremes
    - Drop if hectares quantile is in the lower 20% (48,380 observations deleted)
    - Drop if price per hectare quantile by year is in the lower 10% or in the upper 90% (39,227 observations deleted)

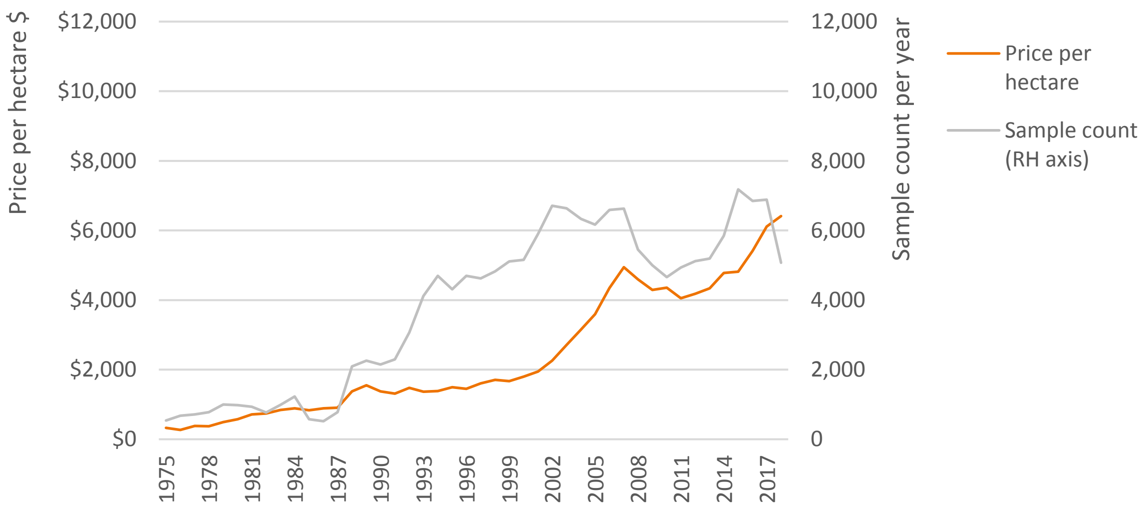
The quantiles for size (hectares) are calculated for the entire sample (using stata xtile [see Equation 8 in Methods section]), while the quantiles for price per hectare are calculated separately for each year. The impact of constraining the scope to broadacre farms is evident when comparing Figures 1 and 2 — which illustrate average annual price per hectare and sample size for the pre and post adjusted datasets. Note that the average price per hectare is unrealistically high and erratic in the 'pre' dataset (Figure 1), whereas this average becomes more normalised and stable in the 'post' dataset (Figure 2). As expected, the specification of a strict data scope and the removal of extreme (likely non-broadacre) transaction records has resulted in considerable loss of sample size. Currently this is preferable to risking inclusion of non broadacre farmland in the analysis to follow and enables us to generate meaningful results. Future work will look to review and loosen these data constraints — to improve sample size without jeopardising data quality.

Figure 1 Pre-data cleaning, average price per hectare and sample count



Source: authors estimates, custom CoreLogic dataset

Figure 2 Post-data cleaning, average price per hectare versus sample size

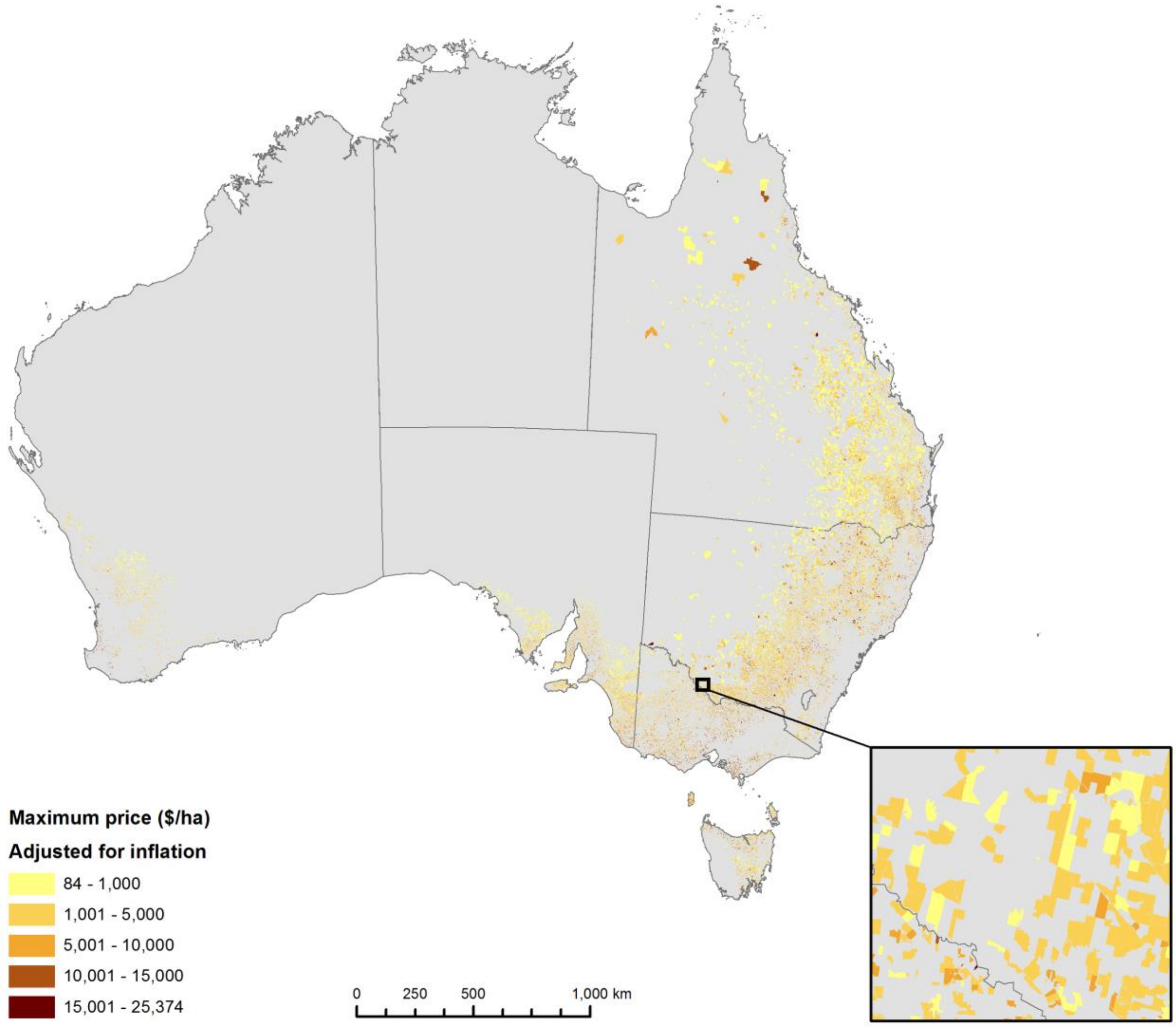


***Source:** authors estimates, custom CoreLogic dataset*

Spatial transformation and variable construction

An important feature of this study is the transformation of farmland transaction information into spatial parcel boundaries. These boundaries are presented in Map 1, with a colour scaling applied to represent the maximum real price per hectare for each parcel. The inset in Map 1 demonstrates the detailed land parcel level data and the differences in price per hectare that exist within a small area. The spatial parcel boundary linkage enables the derivation of farm specific attributes using various spatial datasets, so that a series of covariates can be developed for use in the hedonic models. This effectively enables us to determine the factors (e.g. soil, climate, and distance to towns) that are driving broadacre farmland value.

Map 1 Price per hectare (maximum) by parcel in clean dataset 1975-2018



Source: CoreLogic custom dataset, authors estimates.
Note: Maximum price by parcel adjusted to real 2018 prices for comparability.

We construct spatial variables to explain farmland price per hectare according to the following categories;

Spatial variables

Transport connectedness

Connectedness to the Australian transport network represents an ability to deliver farming inputs (e.g. fertiliser, seed, and fodder) and dispatch farming outputs (e.g. tonnes of wheat, bales of wool), which is important to the profitability of a farming businesses. Proximity to towns and cities may also have other benefits such as an increased access to amenities. We therefore are interested in the distance from the farm to towns and the respective transportation costs, which differ depending on road surface type.

To account for transport connectedness, we construct two variables — transport cost to town and transport cost to city, where a 'town' has a population of at least 1,000 residents and a 'city' has a population of at least 100,000 residents. Distance is calculated from the road network node (intersection) nearest to the property boundary polygon to a node assigned to the nearest respective town or city. The route chosen from farm to town or city minimises transport cost based on road class, formation and length (using the Djikstra algorithm). The cost rate is 1 unit per km for sealed principle roads and increases to 2 units per km for unsealed minor roads. Tracks are also included to ensure connectivity of the road network (5 units per kilometre), as are ferry and barge route lines to (e.g. Flinders Island, Daintree River) with a cost rate of 20 units per kilometre.

The road network is based on Geoscience Australia Topo250k series 3 data (Geodata, 2006), and town locations are based on the ABS Urban Centres and Localities (UCL) from the 2016 census (ABS, 2017).

Rainfall

Rainfall is an important factor of agricultural production, and is therefore expected to contribute to land value. Land situated in areas of favourable average rainfall can lead to more profitable and reliable crop or livestock production. High rainfall properties are therefore likely to attract higher prices per hectare than farmland in the arid areas of inland Australia. While average rainfall differs considerably between the ABARES High Rainfall Zone, the Wheat Sheep Zone, and the Pastoral Zone – there are also significant rainfall variations within these zones.

There are several possible rainfall variables that can be generated using the data available. We use actual quarterly rainfall by farm (available from 1960 to 2018) such that Q1=Jan-Mar, Q2=Apr-Jun; Q3=Jul-Sep; Q4=Oct-Dec. Rainfall is linked to farmland property using the BoM AWAP rainfall grids (which have a 0.05x0.05 degree grid cells [about 5x5km]) (BoM, 2019a). Where a property extends over more than one grid cell the area weighted average across the rectangular extent of the property has been calculated.

Temperature

As with rainfall, variations in temperature can impact agricultural production and possibly the value of farmland. Anecdotally, areas with hotter protracted periods may be less desirable for some farming production types (and may attract lower land prices) than areas which tend to experience more moderate average temperatures. The temperature variables are constructed in a similar way to the rainfall variables, with actual temperature variables representing quarterly annual temperature for each year by farm between 1960 and 2018 such that Q1=Jan-Mar, Q2=Apr-Jun; Q3=Jul-Sep; Q4=Oct-Dec. Temperature is linked to farmland property using the BoM AWAP grids (which have a 0.05x0.05 degree grid cells [about 5x5km] (BoM, 2019b). Where a property extends over more than one grid cell the area weighted average across the rectangular extent of the property has been calculated.

Water coverage

Water is considered to be a generally important resource for agricultural production, therefore the existence of water cover on farmland may impact farmland value. Yet, excessive water coverage might present flood risk, therefore it is difficult to predict the impact of water cover on farmland values. Using the Geoscience Australia Water Observations from Space data records, we apply a variable to capture the percentage frequency with which water has been detected over the Landsat archive for each 25x25m pixel (Geoscience Australia, 2019). High percentages indicate water storages, lakes and water courses; whereas lower percentages capture flooding and flood irrigation. Woody vegetation cover and clearing history

Vegetation

The presence of woody vegetation on broadacre farmland is expected to impact value for several reasons. First, vegetation on cropping land can reduced the productive capacity of that land due to physical obstacles and competition for water and nutrients. Similarly, grazing land with heavy tree cover may produce less pasture and increase costs associated with livestock herding. However, there is a case for retaining woody vegetation, such as for providing shelter from wind and extreme weather. Vegetation can also be beneficial to land quality, by reducing or preventing erosion and by returning nitrogen to the soil. Therefore the expected impact of land clearing and vegetation cover on farmland value is unclear and expected to vary by region and production type.

Vegetation at farm level for each of the available years between 1988 and 2018 is calculated such that 0 means all cleared and 100 means all woody vegetation. This covariate is based on the National Forest and Sparse Woody Vegetation Data (Version 3, 2018 Release) (Department of Environment and Energy, 2018). To practically test the relationship between land clearing and land value, a trend at farm level is derived over the available time period and the corresponding gradient provides an indication of either vegetation clearing or vegetation growth — and the extent of this clearing or growth.

Landsat satellite imagery is used to derive the variable to discriminate woody vegetation from forest, sparse woody and non-woody land cover across a time series from 1988 to 2018 (Department of Environment and Energy, 2019). A forest is defined as woody vegetation with a minimum 20% canopy cover, potentially reaching 2 metres high and a minimum area of 0.2 hectares. Sparse woody is defined as woody vegetation with a canopy cover between 5-19%. The National Carbon Accounting System (NCAS) woody vegetation cover grids include the following 23 years (omitting first two digits from year numbers): 88 89 91 92 95 98 00 02 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18. Each 0.00025 x 0.00025 degree pixel takes a value in {0,1,2,255} where 0 means non-woody cover; 1 means Sparse woody vegetation and 2 means Woody vegetation (forest). 255 means 'no data'. The variable calculated is the average value excluding 'no data' times 50. If all the woody cover grid cells in the property area for a given year are 0 then the average will also be 0. Similarly an output value of 100 from an average of 2 in a given year would mean all the grid cells for a given property have the value 2 in that year. Note that northern SA and southern NT are not included in the NCAS data.

Topography

The topography of broadacre farmland is likely to impact value, with land profiles potentially restricting or broadening the farm production opportunities available to the owner. Generally, steep hilly land may be inaccessible for cropping machinery and limit production opportunities — possibly impeding value. Conversely, flat or slightly undulating land may present few barriers to agricultural production and generally coincide with higher land values. Yet, it is anticipated that this impact of land slope on price per hectare is likely to be subtle. To account for topography, four classes ranging from 1 for flat to 4 for steep are generated to estimate the proportion of the property in each slope class. These variables use the Shuttle Radar Topographic Mission (SRTM) digital elevation model which has 0.0008333 degree pixels (about 90m) (Geoscience Australia, 2011).

Building cover

Physical buildings such as storage sheds or residential homesteads are generally expected to result in higher farmland values. The number of building points present on each property is derived using state data. Data are available for all states except Western Australia and Northern Territory via data.gov.au.

Land use type

Land use type is likely to impact farmland value. Land use variables are constructed by overlaying the Catchment Scale Land Use of Australia Update December 2018 (CLUM18) dataset (ABARES, 2019) on the CoreLogic parcel boundaries, and then assigning a primary, secondary, and tertiary land use category. CLUM18 was converted from a raster with 50m x 50m pixels to a polygon layer before the layer was overlaid. Overlaps between the layers that are less than half the area of one CLUM18 pixel have been ignored so the total of the hectares field will be a little less than the property area. Some very small properties could have been lost as a result. It is probable that these small properties are out of scope.

While this process generated a large number of potential variables, our focus is on the following land use (cropping, grazing, mining, wind power, and buildings/infrastructure) for the purpose of constructing dummy variables. For cropping and grazing, a flag is assigned where cropping or grazing land use respectively exceeds 50 hectares. For mine sites, wind power turbines, and building infrastructure, a flag is assigned where these respective land use features are present within the farmland polygon boundaries.

Soil condition

Soil may be important for pasture and crop production, and hence, broadacre farmland value. Degradation of land (such as erosion) may also be undesirable and impact value negatively, however, due to the complexity of these variables it is difficult to anticipate their impact on broadacre farmland value. To account for soil, we include four soil condition layers from Leys et al. (2017). The variables constructed from these layers are acidification risk, carbon risk, erosion by water risk, and erosion by wind risk. Interpretation of these variables with relation to land value is more difficult due to the multiple dimensions of soil characteristics. Acidification, for example, has increased in many areas due to the intensification of cropping and nitrogen fertiliser use (Leys et al. 2017).

Non-farm-specific variables

Broadacre productivity (by State)

Experimental (unpublished) state level estimates of the ABARES broadacre Total Factor Productivity (TFP) series are included in the hedonic estimation. These estimates are annual from 1978 to 2018. The purpose of including these estimates is to control for state specific technology progress and improvements in production (such as through management practices) that may have a corresponding impact on farmland values.

Rural debt

The level of rural debt provides an indication of the availability of finance and sentiment which may drive broadacre farmland values. We use the variable 'total rural debt' from the Reserve Bank of Australia Money and Credit and Statistics (RBA, 2019). Unfortunately this variable is general (i.e. not farm or location specific), providing only a basic indication of debt levels from 1975 to 2018.

Primary land use dummies

Dummies are generated to identify cropping, livestock, mixed, dairy, irrigated, vineyards, sugar, lifestyle, and horticulture; for the purposes of testing and data cleaning. These variables were constructed using key words in the CoreLogic primary land use variable (as in Appendix Table A5 and A6). These variables are useful as they provide an indication of production type at the time of sale, and according to CoreLogic are constantly maintained and generally considered to be high quality.

ABARES zone dummies

A flag of either (High Rainfall Zone [HFRZ], Wheat Sheep Zone [WSZ], Pastoral Zone [PZ]) is assigned to each CoreLogic transaction where geographic coordinates (latitude/longitude) fall within the respective ABARES zone. This flag is used as a control variable and as a hedonic model constraint.

GRDC zone dummies

A flag of either (GRDC Southern [GRDC_S], GRDC Northern [GRDC_N], GRDC Western [GRDC_W]) is assigned to each CoreLogic transaction where geographic coordinates (latitude/longitude) fall within the respective GRDC zone. This flag is used as a control variable and as a hedonic model constraint.

Time dummies

An annual flag generated for each year from 1975 to 2018 based on the transaction date of property.

2 Method

Many hedonic studies have attempted to estimate the values of residential real estate, yet few have attempted to generate estimates of Australian farmland – mainly reflecting data scarcity and limitations. Yet, the well-worn path of residential real estate hedonic estimation provides valuable lessons that can be applied in developing a hedonic model of Australian broadacre farmland values. For example, Waltl (2016) found that the number of bedrooms and the number of bathrooms affected house prices differently depending on price segment. Specifically, additional bathrooms led to higher values in the top price segment, whereas additional bedrooms are more important to the lower price segment. Such differences cannot be easily observed in typical linear hedonic models, since they assume the relationship between the dependent variable (price) and the co-variates (bedrooms and bathrooms in this example), are the same for all transactions. We therefore consider the importance of controlling for differences between price segments in our hedonic model — which we test using a simultaneous quantile regression approach.

An advantage of the quantile approach is that it becomes possible to differentiate between farmland sub markets (i.e. according to value or size), based on the assumption that the drivers of value for large farms may differ from small farms. Similarly, the value of inexpensive farmland may be driven by different factors than for expensive farmland. While the quantile approach provides important insights to the drivers of farmland value (by price segment, by size), a more traditional stratified robust ordinary least squares regression approach is used to generate hedonic farmland models (national level, zone).

The hedonic approach [which uses characteristics to determine price alongside historic and imputed sale price information in regression analysis to generate indexes], is one of several possible approaches to farmland valuation. Other methods include repeat sales, appraisal and stratification. The stratification method separates the total sample into strata according to property type, median price and other factors. The appraisal method is a process of matching an appraised value to a sale value for the same property in a previous period. The repeat sales method uses information from properties sold multiple times to create indexes — such that the sample only contains properties sold multiple times. Our approach uses both hedonics and stratification.

There are limitations and challenges with all methods. For example, obtaining sufficiently detailed data to explain value is a challenge (potentially leading to omitted variable bias). One possible solution to compensate for this data limitation is to use longitude and latitude as a way of controlling for spatial dependence (De Haan and Diewert, 2013). Some examples which use variations of hedonic price valuation with spatial dependence include Hill and Scholz (2014), Hill et al. (2010) and Hill (2011). Rather than control for spatial dependence (by using geographic coordinates as explanatory variables), we identify specific spatial farm level characteristics using spatial data linkages.

A broad range of intricate factors are thought to drive farmland value, and therefore specifying a logical theoretical model (as we have in Table 2) is an important step in obtaining meaningful results. Our theoretical model is loosely based on overseas examples of farmland hedonic estimation, such as in Pyykkönen (2006) who included a range factors to explain farmland value such as parcel size, land features, land quality, cropping yield, climate variables and infrastructure availability. Other studies such as Drescher et al. (2001) applied a very different range of explanatory variables including economic and government influences, expectations about the future, and market participant characteristics. Examples also exist where liveability and recreational amenities are controlled for (i.e., access to scenic landscapes, wetlands, lakes, forest (Ma and Swinton, 2011). Other factors used to explain farmland value include soil erosion (Palmquist and Danielson 1989), distance from urbanised areas (Huang et al. 2006), purpose of neighbouring land (Chicoine, 1981), rainfall and temperature (Mendelsohn et al. 1994) among many others. Practically, the development of a theoretical model is often limited by data availability. Obtaining, deriving and linking suitable data to develop such a model for Australian broadacre farmland was therefore a significant component of this study.

Only a handful of academic studies have attempted to estimate Australian farmland values using the hedonic approach, however these studies appear to be bound by data limitations, focusing on a specific regional area over a short period of time (Dent and Ward 2014, King and Sinden 1988, Eves 2016). We are unaware of any attempt to estimate Australian farmland values at the national level using a hedonic approach, particularly using such an extensive dataset which exceeds a 40 year period. However, there are examples of accounting type approaches which use historic trends to generate prices. Rural Bank (2016), for example, uses state government data to calculate farmland sale price averages over time by state, region and municipality. They provide commentary to suggest what the drivers of farmland value might include (location, climate, productivity, land quality, sentiment, interest rates, commodity prices, and general economic performance), however this appears to be anecdotal.

In this study, the method used to estimate the drivers of Australian broadacre farmland value is based on the conceptual framework in Table 2, where price per hectare is the dependent variable. The explanatory variables can be categorised as 'spatial farm specific', 'administrative' or 'macro-indicators'. The spatial variables are important in explaining the specific features of a given farmland property that contributes to value. Similarly, some of the administrative variables do provide farm specific explanatory power for land value (such as land area and house size where applicable). The macro indicators are more general, and relate to the state or national level, yet despite this 'broadness' they are still likely to impact on farmland value by influencing market sentiment and the availability of finance. These macro indicators are only used in the national level models.

Table 2 Theoretical model of explanatory variables

Data type	Explanatory variable	Description	Expected relationship to dependent variable
Administrative	Number of bedrooms	Number of bedrooms if house is present on transacted property	Positive – presence of house likely to increase land value. Larger house is expected to increase this further.
	Number of bathrooms	Number of bathrooms if house is present on transacted property	
	Hectares	Land size for transacted property	Negative – small farms generally expected to have higher value on 'per hectare' basis. As size increases, price per hectare is likely to decline.
	Multi-sale	Flag to indicate property parcels sold as grouping	Negative – sale of parcels as grouping may result in reduced price per hectare.

	Building points	Spatial building points on parcel	Positive, greater fixed capital investment likely to increase farmland value
	Transport cost to small town (population 1,000)	Distance from property to small town, accounting for quality/type of roads and infrastructure.	Negative – an increase in the distance / cost to a small or large town suggests the property is more remote and has less access to infrastructure and services. This is likely to reduce land value.
	Transport cost to large town (population 100,000)	Distance from property to large town, accounting for quality/type of roads and infrastructure.	
	Land clearing	Evidence of land clearing (negative) or revegetation (positive) between 1988 and 2018)	Unknown, potential to be positive or negative depending on region and farm type
	Land slope (flat)	Percentage of land parcel that is flat	Unknown – flatter land is expected to be a positive attribute for land value, whereas hilly or steep land is likely to be negative. However gradient may be irrelevant for some production types.
	Land slope (undulating)	Percentage of land parcel that is undulating	
	Land slope (Hilly)	Percentage of land parcel that is hilly	
	Land slope (Steep)	Percentage of land parcel that is Steep	
Spatial	Rainfall Jan-Mar	Average rainfall assigned to farm by quarter	Positive – Farmland with high average rainfall is likely to be positive for value due to the benefits for most agricultural production systems.
	Rainfall Apr-Jun		
	Rainfall Jul-Sep		
	Rainfall Oct-Dec		
	Maximum temperature Jan-Mar	Average maximum temperature assigned to farm	Negative – Areas with high average maximum temperatures are expected to demand lower prices, as this is unfavourable for most broadacre production activities.
	Maximum temperature Apr-Jun		
	Maximum temperature Jul-Sep		
	Maximum temperature Oct-Dec		
	Minimum temperature Jan-Mar	Average minimum temperature assigned to farm	Unknown – the effect of average minimum temperature on broadacre production is unclear, yet this may be positively related to value due to reduced frost risk which is unfavourable for some crop production systems.
	Minimum temperature Apr-Jun		
	Minimum temperature Jul-Sep		
	Minimum temperature Oct-Dec		
	Water cover 20%	Percentage of transacted farmland with water cover	Unknown – low levels of water cover are expected to be positive, however excessive water cover may be negative due to increased flood risk.
	Water cover 5%		

	Water cover 1.5%		
Spatial	Cropping	Dummy variable assigned where land used for cropping exceeds 50 hectares	Positive/unknown – may indicate more valuable farmland
	Grazing	Dummy variable assigned where land used for grazing exceeds 50 hectares	Negative/unknown – may indicate less valuable farmland
	Mines sites	Dummy variable assigned where mine site identified on farmland	Unknown. Some anecdotal information suggests a positive relationship
	Wind power	Dummy where wind turbines present on farmland	Positive – rental returns
	Buildings	Dummy where buildings/infrastructure are present	Positive – fixed investment may increase farmland value
	Soil Acid high	Acidification risk high	Negative
	Soil Carbon high	Soil carbon potential high	Negative
	Wind erosion high	Wind erosion risk high	Negative
	Water erosion high	Water erosion risk high	Negative
Macro	Productivity	State level experimental productivity estimates assigned as indicator of technology	Positive – productivity improvement may lead to improvements in land management and investment in fixed infrastructure through improved technology and farming practices.
	Rural debt	Total national rural debt (RBA)	Positive – Increases in debt may be an indication of increased demand for agricultural land and increase availability of lending.

The method used to generate the hedonic land value estimates varies slightly according to the scope of the analysis. For the analysis at the national level, by ABARES zones and GRDC regions, a robust regression method (Hamilton, 1991) is applied (rreg in stata). The dependent and explanatory variables remain the same throughout, however the strata and controls vary slightly. This method begins by fitting an ordinary least squares regression based on a typical OLS as in Equation (1), where the logged dependent variable (contract price per hectare) is represented by $\log(Y)$, the coefficients are represented by β, the logged explanatory variables are represented by $\log(X)$, the dummy explanatory variables are represented as XD, and the control variables for time (year) are CT (where the base year is automatically selected). The error term is represented by ϵ. Equation 1 is used for the analysis at the national level:

$$log(Y) = \beta_0 + \beta_1 \, log(X)_1 + \beta_2 XD_1 + \beta_3 CT_1 \ldots + \beta_x \, log(X)_x + \beta_x XD_x + \beta_x CT_x + \epsilon \qquad (1)$$

For ABARES zones, strata R is assigned to separate the analysis according to the High Rainfall Zone, the Wheat Sheep Zone, and the Pastoral zone as in Equation 2:

$$R(log(Y)) = R(\beta_0 + \beta_1 \, log(X)_1 + \beta_2 XD_1 + \beta_3 CT_1 \ldots + \beta_x \, log(X)_x + \beta_x XD_x + \beta_x CT_x) + \epsilon \qquad (2)$$

The GRDC regional analysis follows a similar logic, yet with the strata of G applied to control for the three GRDC regions (Southern, Northern, and Western). A second constraint is applied to limit the transactions to cropping farms, which is based on CLUM data, such that $cropping\ area\ \neq 0$ (represented as C) in Equation 3.

$$C.G(log(Y)) = C.G(\beta_0 + \beta_1\, log(X)_1 + \beta_2 XD_1 + \beta_3 CT_1 \ldots + \beta_x\, log(X)_x + \beta_x XD_x + \beta_x CT_x) + \epsilon \quad (3)$$

After performing the OLS regressions following Equations 1 to 3, the 'robust regression' approach (from Hamilton, 2012) next calculates weights, and then recalculates Equations 1 to 3 using these weights. Huber weighting is the first function used, where small residuals receive weights of 1, and variables with larger residuals are assigned gradually lower weights. As outlined in Hamilton (2012), Huber weights (Huber 1964) are used until convergence, and then, from that result, biweights (Beaton and Tukey 1974) are used until convergence. Both weighting functions are used because Huber weights have problems dealing with severe outliers, whereas biweights sometimes fail to converge or have multiple solutions. The initial Huber weighting should improve the behaviour of the biweight estimator.

This is demonstrated in Equation 4, where e_i represents the ith residual $Y_i - X_i B$ and the median absolute deviation MAD from the median residual med:

$$MAD = med(|e_i - med\{e_i\} \quad (4)$$

Noting that the ith scaled residual u_i is $u_i = e_i/s$ where s is the residual scale estimate. The robust OLS method uses $s = MAD/.6745$ and the Huber estimation finds case weights w_i such that:

$$w_i = \begin{cases} 1 & if\ |u_i| \leq c \\ c/|u_i| & otherwise \end{cases} \quad (5)$$

Where c is a tuning constant, such that $c = 1.345$, meaning that down weighting begins where absolute residuals exceed approximately $2 \cdot MAD$. The second weighting function used in robust OLS is referred to as 'biweight', where all non-zero residuals receive some down weighting according to the following function:

$$w_i = \begin{cases} [1 - (u_i/c)^2]^2 & if\ |u_i| \leq c \\ 0 & otherwise \end{cases} \quad (6)$$

Very large residuals ($|u_i| \geq c$) result in zero weights and these severe outliers drop out of the analysis. The biweight iterations employ a turning constant of $c = 4.685$, meaning that cases with absolute residuals of $7 \cdot MAD$ or more are assigned zero weights (and dropped). The turning constants $c = 1.345$ ($Huber$) and $c = 4.685$ ($biweight$) give these robust procedures about 95% of the efficiency of OLS when applied to data with normal distributed errors.

To analyse farmland value from another perspective, we are interested in the relationship between price per hectare and the explanatory variables according to different farm size categories. This can be achieved by following a variation of the robust regression approach using the constraint q, which is a constructed quantile that categorises farmland transactions into 1 of 3 groups according to their respective size in hectares. These categories are generated using the following stata function 'xtile' (Ryan, P. 2014).

$$(-\infty, x_{[p1]}], (x_{[p1]}, x_{[p2]}], \ldots, (x_{[pm-2]}, x_{[pm-1]}], (x_{[pm-1]}, +\infty) \quad (7)$$

Numbered, respectively,$1, 2, \dots, m$, based on the m quantiles given by the $p_k th$ percentiles, where $pk = 100\ k/m$ for $k = 1, 2, \dots, m-1$.

$$q.S(log(Y)) = q.S(\beta_0 + \beta_1\, log(X)_1 + \beta_2 XD_1 + \beta_3 CT_1 \dots + \beta_x\, log(X)_x + \beta_x XD_x + \beta_x CT_1) + \epsilon \qquad (8)$$

The final set of hedonic models use a simultaneous quantile regression approach (stata sqreg), to identify variations between the dependent and explanatory variables according to inexpensive, moderate and expensive farmland. This analysis is conducted from a national perspective only. The following functions are used for the national analysis, allowing quantiles to be specified and estimated simultaneously. An estimate of the entire variance-covariance matrix of the estimates is obtained by bootstrapping (Stata, 2014) with 20 iterations. We therefore estimate the following where quantiles are represented by Q_x in Equation 9:

$$[(Q_{.10}(log(Y)) = \beta_{0.Q_{.10}} + \beta_{1.Q_{.10}}\, log(X)_{1.Q_{.10}} + \beta_{2.Q_{.10}} XD_{1.Q_{.10}} + \beta_{3.Q_{.10}} CT_{1.Q_{.10}} \dots + \beta_{x.Q_{.10}}\, log(X)_{x.Q_{.10}} + \beta_{x.Q_{.10}} XD_{x.Q_{.10}} + \beta_{x.Q_{.10}} CT_{x.Q_{.10}} + \epsilon)\ \wedge$$

$$(Q_{.50}(log(Y)) = \beta_{0.Q_{.50}} + \beta_{1.Q_{.50}}\, log(X)_{1.Q_{.50}} + \beta_{2.Q_{.50}} XD_{1.Q_{.50}} + \beta_{3.Q_{.50}} CT_{1.Q_{.50}} \dots + \beta_{x.Q_{.50}}\, log(X)_{x.Q_{.50}} + \beta_{x.Q_{.50}} XD_{x.Q_{.50}} + \beta_{x.Q_{.50}} CT_{x.Q_{.50}} + \epsilon)\ \wedge$$

$$(Q_{.90}(log(Y)) = \beta_{0.Q_{.90}} + \beta_{1.Q_{.90}}\, log(X)_{1.Q_{.90}} + \beta_{2.Q_{.90}} XD_{1.Q_{.90}} + \beta_{3.Q_{.90}} CT_{1.Q_{.90}} \dots + \beta_{x.Q_{.90}}\, log(X)_{x.Q_{.90}} + \beta_{x.Q_{.90}} XD_{x.Q_{.90}} + \beta_{x.Q_{.90}} CT_{x.Q_{.90}} + \epsilon)] \qquad (9)$$

One feature of quantile regression is that it outputs a $pseudo\ R^2$ rather than R^2 as an indicator of model fit. $Pseudo\ R^2$ can be interpreted in the same way as R^2, however similarly high values will not be achieved. "*Those unfamiliar with ρ^2($pseudo\ R^2$) should be forewarned that its values tend to be considerably lower than those of the R^2 index and should not be judged by the standards for a "good fit" in ordinary regression analysis. For example, values of .2 to .4 for ρ^2 represent an excellent fit.*" (McFadden, 1977, p.34-35). $Pseudo\ R^2$ is calculated as follows:

$$Pseudo\ R^2\!:\ 1 - \frac{sum\ of\ weighted\ deviations\ about\ estimated\ quantile}{sum\ of\ weighted\ deviations\ about\ raw\ quantile} \qquad (10)$$

The simultaneous quantile regression allows for the use of bootstrapping (Stata, 2014), and effectively a resampled dataset. Resampling is repeated multiple times (in our case, 20 times), with a new random sample being used each time. This process builds a dataset of replicated statistics and calculates standard error using the standard formula for the sample standard deviation:

$$\widehat{se} = \left\{\frac{1}{k-1}\sum(\hat{\theta}_i - \bar{\theta})^2\right\}^{1/2} \qquad (11)$$

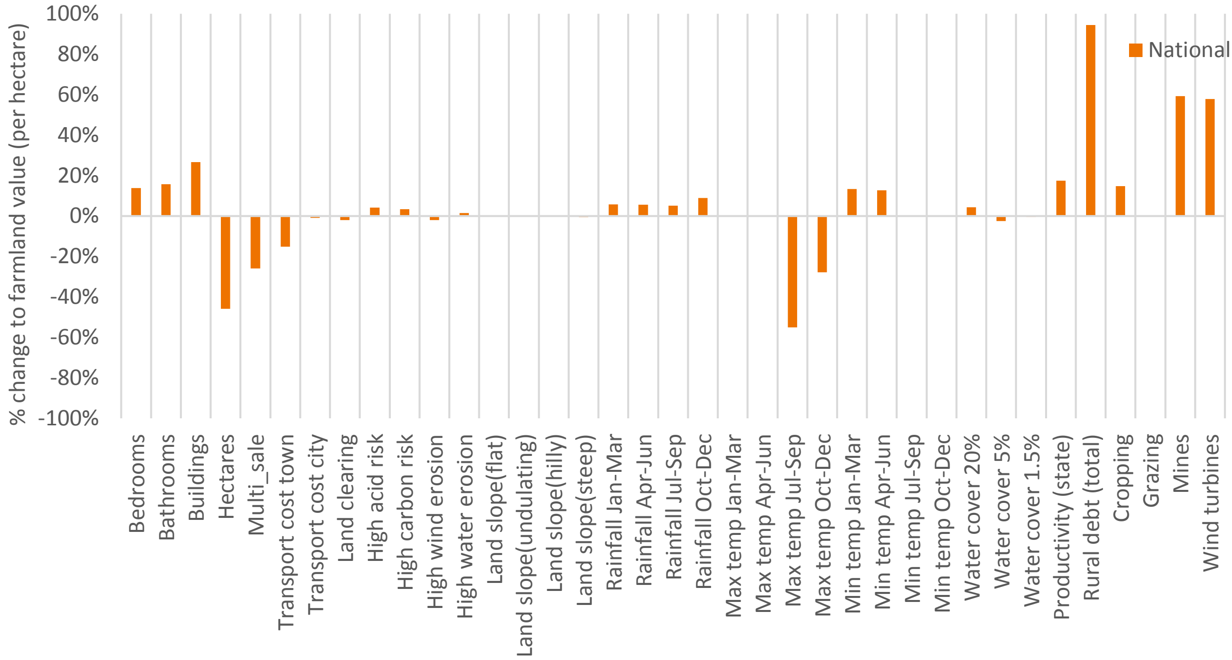
Where: $\hat{\theta}_i$ is the statistic calculated using the ith bootstrap sample, k is the number of replications, and $\bar{\theta}$ is the average of bootstrapped estimates.

3 Results

Results for each of the hedonic models are presented in the following order: National, ABARES zones, GRDC regions, size quantiles, and price per hectare quantiles. All results present the co-efficient of the explanatory variable in log form – such that the percentage change in the co-efficient results in a corresponding percentage change in the dependent variable (price her hectare). Any insignificant co-efficient results $where\ P > |t| => .05$ are omitted, meaning that these variables will appear as blank (no bar) in Figures 3 to 7.

Similarly, year control variables are omitted (for ease of presentation). Accompanying each set of results is a brief descriptive interpretation.

Figure 3 Land value drivers – National



Source: Authors estimates
Note: (R^2=.67), observations=166,994,

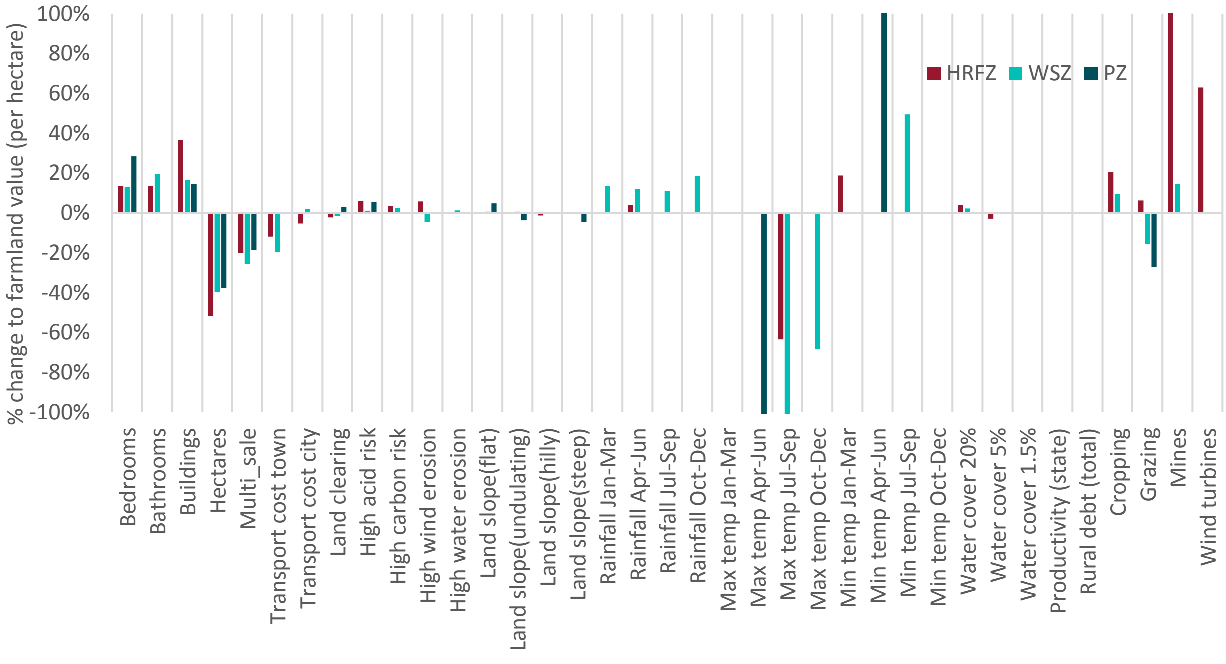
The national hedonic model provides a general overview of the factors which have driven farmland values from 1975 to 2018; and the proportional impact of these drivers. The presence of a house or buildings on the farmland appears to be important to value. Specifically, the size of the farm house (where present) and the number of buildings has a significant and positive relationship with price per hectare. Size (Hectares) is significant and negative, suggesting that price per hectare tends to be higher for smaller farms. Also aligning with expectations are transport costs, which are significant and negative for both towns and cities – suggesting that farmland value decreases with increased remoteness and reduced access to infrastructure. The multi-sale variable suggests that the grouping of land parcels into a single transaction drives reduced value per hectare.

The effect of land clearing is statistically significant and negative (-1.9%) at the national level, suggesting that a reduction in vegetation may result in a subtle land value increase. Steep land terrain also has a significant (yet subtle) effect on farmland values, suggesting that very hilly land may result in a slight reduction in land value.

Other general observations are that increased average rainfall, high average minimum temperature, productivity, increased average debt, production choice (cropping) and the presence of wind turbines or mine sites on farm appear to positively drive farmland value. Factors that appear to be negative drivers of broadacre farmland value at the national level include high average maximum temperatures in the Jul-Sep and Oct-Dec quarters, high wind erosion risk and water cover of 5% or 1.5% (possibly due to heightened flood risk).

While the coefficients align with expectations for most of the variables, results from some of the soil variables yielded unexpected signs, with high risk of acidification, carbon and water erosion all appearing to have a significant and positive (yet minor) relationship to farmland value. Although this appears to be counterintuitive, it is also possible that the soil coefficients are valid and that the initial interpretation (in Table 2) was incorrect, or that the interaction between soil risk and other variables (such as land clearing) and leading to spurious results. If we assume that the soil variable results are valid, the signs of these covariates could be related to farmland profitability. For example, farmers that are operating their land more intensively (higher stocking densities and more intensive fertiliser use), may achieve higher profitability, which may allow them to demand a higher price for their land (despite the long term degradation of the soil, which may be unnoticed by the land purchaser). This hypothetical scenario therefore may induce the positive relationship between soil risk and farmland value.

Figure 4 Land value drivers - ABARES zones



Source: Authors estimates

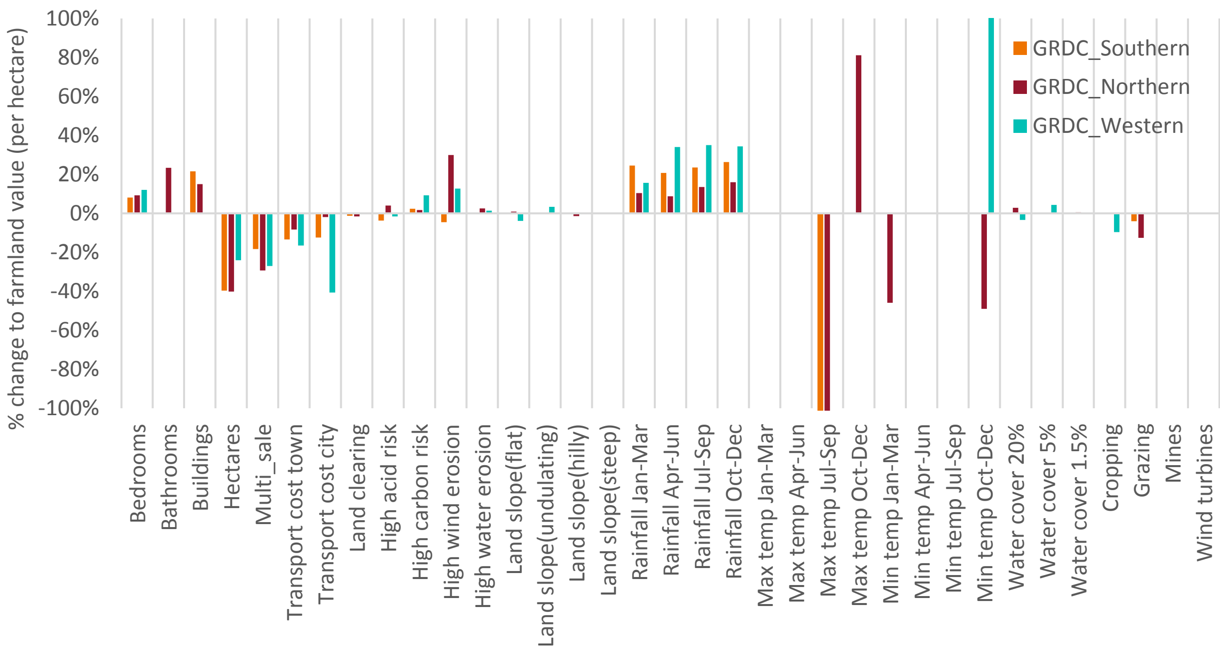
Note: High rainfall zone (R^2=.65), observations=105,061; Wheat Sheep zone (R^2=.64), observations= 85,423; Pastoral zone (R^2=.6581), observations=2,376

To explore these drivers of broadacre farmland further, the next set of results compares ABARES farming zones *(as in Appendix Box A1).* These zones are the High Rainfall Zone (HRFZ), the Wheat Sheep Zone (WSZ), and the Pastoral Zone (PZ). It is generally accepted farms in the HRFZ are typically smaller and tend to benefit from favourable climate conditions. Farms in the WSZ are generally larger than in the HRFZ and experience moderate climate conditions suitable for large scale cropping. Farms in the PZ are generally large scale grazing properties which anecdotally attract lower prices on a per hectare basis.

Many of the land value drivers identified at the national level hold true after disaggregation into the ABARES zones (Figure 4). The presence of a house, buildings or infrastructure appear to be significant and positive for all zones — however a large house (in terms of number of bedrooms) appears to be more important in the PZ and a high number of buildings (i.e. sheds and houses) appears to be more important to value in the HRFZ. The coefficient sign (negative) and significance of farm size (hectares) and multi-sale holds for all ABARES zones. Multi-sale transactions attract lower land values in the WSZ, whereas farmland in the HRFZ is price sensitive to size (such that increasingly large farms may attract lower values on a price per hectare basis). Increased distance from towns negatively impacts land values in the HRFZ and WSZ, however is insignificant for the PZ.

Land value in the WSZ appears to be largely driven by favourable climatic conditions. Increased average rainfall in the WSZ is positive and significant in all quarters, and high maximum average temperature in the Jul-Sep / Oct-Dec quarters is negative and significant. High risk of wind erosion is also a negative driver of broadacre farmland value in the WSZ. Note that cropping production type is positive for WSZ, whereas grazing production is negative. This combination of features may reflect that higher values are assigned to favourable cropping areas within the WSZ (compared to less favourable grazing areas of the WSZ with hotter and drier climate conditions).

Figure 5 Land value drivers – GRDC regions (farms with cropping)



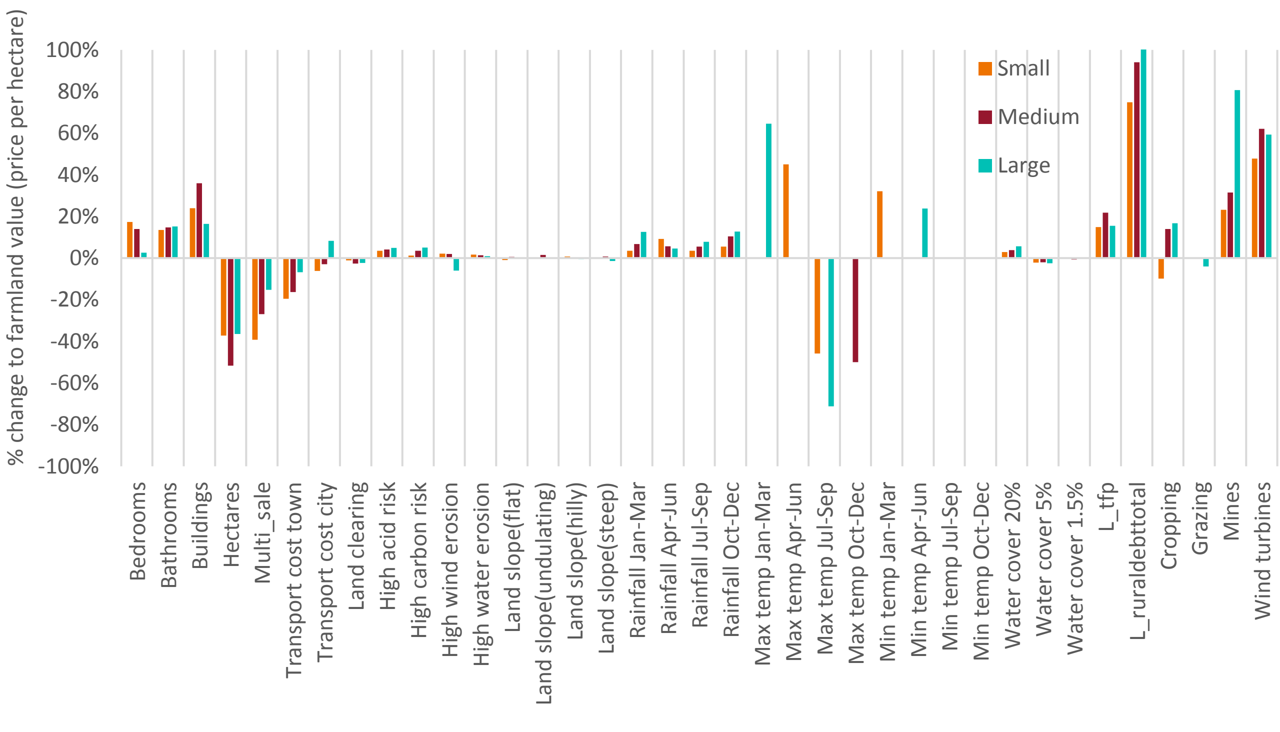
Source: Authors estimates
Note: Southern GRDC region crop farms (R^2=.64), observations=10,609; Northern GRDC region crop farms (R^2=.64), observations=38,865; Western GRDC region crop farms (R^2=.73), observations=6,990;

Next, the model in Figure 5 illustrates cropping farms by GRDC region. Climate appears to be the most important factor, with higher average rainfall in all seasons positively related to land value in all GRDC regions. High average maximum temperatures in the Jul-Sep quarter appears to drive values downward considerably in the Southern and Northern GRDC regions. Increased transportation cost is significant and negative for all GRDC regions, suggesting that increased remoteness is a negative land value driver. Larger farm size and multi-sale transactions also appear to correspond with lower land values on a price per hectare basis.

When the hedonic models are generated (as presented in the 'Results' section), the dataset is refined further through stratification. The strata are ABARES agricultural production zones (High Rainfall, Wheat Sheep, and Pastoral) and GRDC zones (crop farms in the GRDC Southern Region, GRDC Northern Region, and GRDC Western Region). Size and price quantiles are also used as a form of stratification using quantile regression approaches.

Three separate models according to size quantiles are used to generate the results for Figure 6, where hectares are categorised into three groups (using stata xtile). House characteristics appear to contribute positively for all size categories, yet are the most influential to land values on small farms (based on number of bedrooms) and then progressively less influential as farm size increases. Production type also appears to be important to value by size, with the crop farm flag positively affecting medium and large broadacre farmland value – compared to the grazing flag which is negative for large farms. An unusual coefficient is the 'transport cost to city' for large farms, which implies that the value of large farms benefits from increased distance to cities. This is difficult to interpret, however may be due to some expensive inland pastoral holdings.

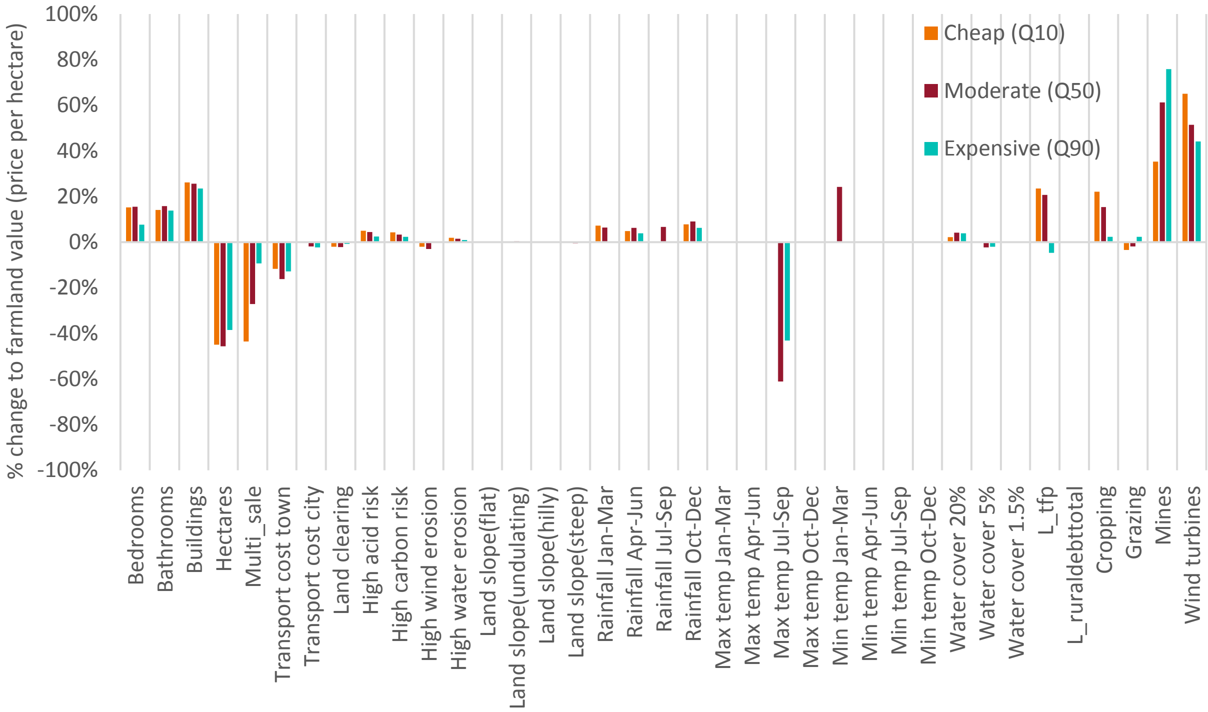
Figure 6 Land value drivers – National – by size quantiles



Source: Authors estimates
Note: National small (R^2=.58), observations=55,665; National medium (R^2=.51), observations=55,664;

The following estimates in Figure 7 were generated using a simultaneous quantile regression approach, in order to observe whether the drivers of farmland value differ according to price segment (price per hectare). For this analysis, we assign quantiles [using stata sqreg], to differentiate between cheap, moderate and expensive farmland. As this is a national level analysis we re-introduce the macro variables for productivity (L_TFP) and (L_ruraldebttotal). TFP appears to be a positive driver for cheap and moderately valued farmland, however is negative for expensive farmland. This implies that the most expensive farms (on a per hectare basis) may be less concerned with profitability and productivity. One possibility is that this quantile is capturing some farms which are operated for leisure. Cropping production type is also positively correlated for cheap and moderate farms. The importance of climate factors varies slightly by value. Interestingly, high average maximum temperatures are negative for Jul-Sep for moderate and expensive broadacre farmland (and insignificant for cheap farms).

Figure 7 Land value drivers – National – by price per hectare quantile



Source: Authors estimates
Note: Cheap (.10) (PR^2=.42); Moderate (PR^2=.43), Expensive (PR^2=.39), Number of observations=166,994

4 Conclusions and future research

This paper reports on the preliminary findings from our effort to develop statistical infrastructure around a large administrative farmland transaction dataset. The data contains rich information, covering transaction records over a long time period and across all the States and Territories. We have supplemented the CoreLogic transaction records with data from other sources to increase its power in the estimation of farmland values. After an intensive analysis, we have come to the conclusion that the custom CoreLogic dataset is of sufficient quality and is 'fit for purpose'. However, in order to generate significant economic insights, it is necessary to clean and carefully define the scope of the CoreLogic data, as well as complement it with data from other sources.

Preliminary results from applying the first iteration of hedonic models were presented in this report. Factors that typically appear to have a positive relationship with farmland value are the presence of a farm house (and the size of that house), the number of buildings on the property and higher average rainfall. The existence of wind turbines and mine sites also appears to be related with higher farmland values. The broad economic variables (rural debt and productivity) appear to similarly be positively correlated with farmland values. Factors that generally have a negative effect on farmland value include farm size (larger farms tend to be sold at a lower price per hectare), multi-sale (price per hectare is lower when farms are sold as a group), remoteness (farms further from towns and cities tend to be cheaper), land clearing (implying that cleared properties achieve a slightly higher price), wind erosion risk (suggests that increased erosion is slightly negative for farmland price), and high average maximum temperature from July to December.

Opportunities for further investigation and refinement of explanatory variables should be explored. For example, the soil acidification variable had a statistically significant and positive (yet small) correlation with land values in all hedonic models except for two of the regions in the crop specific GRDC model. These findings were somewhat contrary to initial expectations, such that higher soil acidification was expected to indicate depletion of natural capital (e.g. through excessive fertiliser use), and therefore be correlated with lower land values. Such interactions and nuances demonstrate the complexity in analysing and 'fitting' farmland value statistical models; but also the potential they have to help answer research and policy questions.

While our study covers considerable ground in constructing national level hedonic farmland models, the findings are 'experimental' and opportunities remain for further refinement. First, we take a conservative approach to data cleaning – removing about 67% of transaction records to meet our strict scope of non-irrigated broadacre farmland. This cautious approach was taken to increase our confidence in the transaction records we analysed. It is highly possible that the sample size could be improved without adversely impacting the dataset quality by carefully reviewing the excluded transactions. For administrative data, 'cleaning' and 'scoping' is resource intensive, however it is necessary — particularly in regions with low sample size or high frequency of missing data (as appears to be the case in Western Australia).

With this research report, we look to prompt feedback which can assist to enhance our datasets and make improvements on the prototype hedonic models. After peer review and further refinement, we will work towards a set of fully-fledged hedonic models - by exploring estimation techniques and calibrating model specifications (such as determining functional forms, selecting of variables and testing explanatory power etc.). After developing a good understanding of the data at hand and resolving major methodological issues with the application of hedonic approach to farmland, we will expand our testing of estimation techniques — namely, the use of machine learning. After that, we will start using the hedonic model for estimating values of specific farms and analysing some of the empirical issues that have important economic and policy implications.

Appendix A: Statistical classifications and variables

Box A1 Farming zones and regions

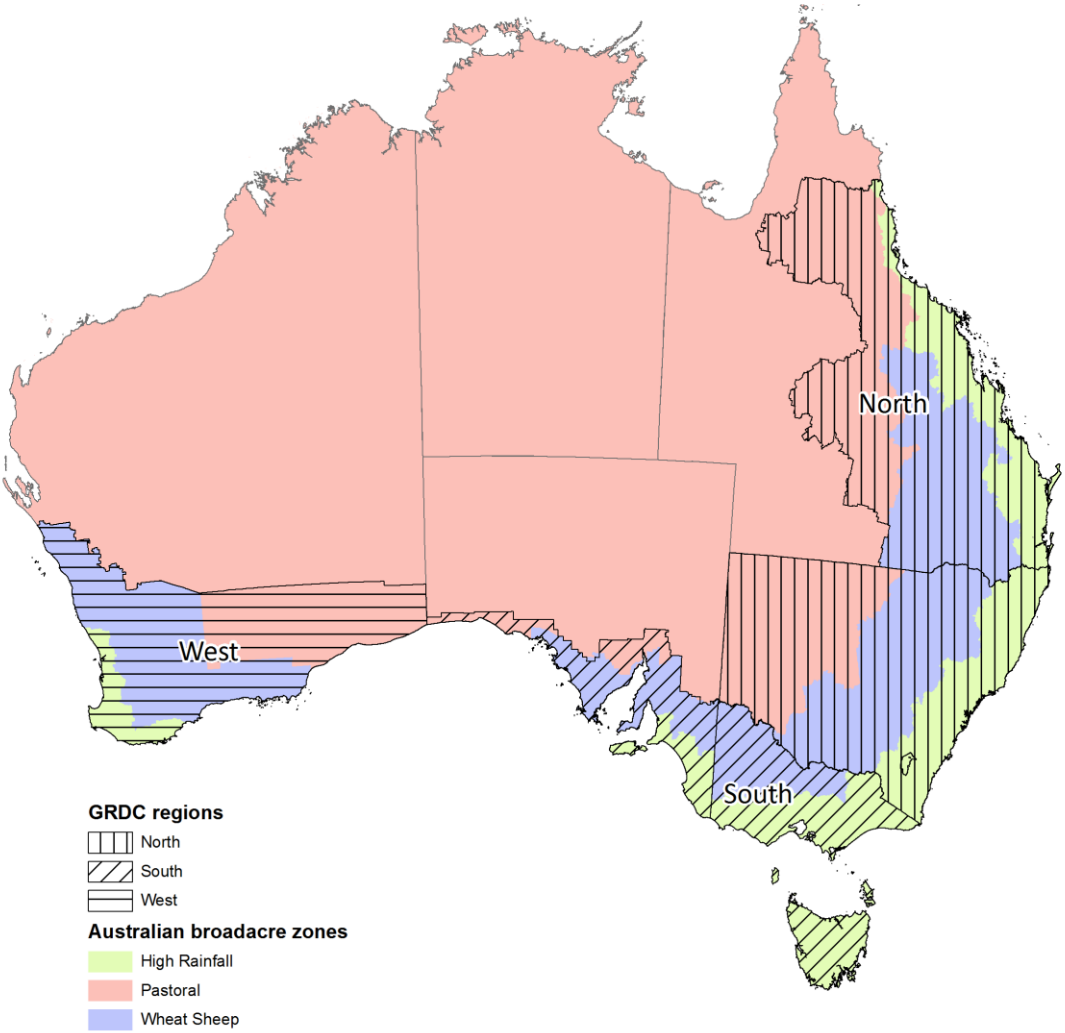


Table A3 CoreLogic main data variables

No.	Data Field	Definition
1	Property ID	Unique record key within the core database for the property.
2	Real Property Description	The legal parcel(s) description of the property, depending on the scheme adopted for each state.
3	Lot Number	Lot Number component of the parcel's description. NSW, VIC, QLD, WA, SA, TAS, NT only
4	Full Property Address	Property Address
5	Property Type	CoreLogic identified category for the property such as House, Unit, Flats, Land, Business i.e. House, Unit, Flats, Business, Commercial, Community, Farm, Land, Storage Unit
6	Property Type Minor	Corelogic minor category for properties such as Multi Storey, Duplex, One story/Lowset, etc.
7	Primary Land Use	The Primary Land Use of the property such as single Unit Dwelling, House etc.
8	Latitude	The geographical latitude of the property.
9	Longitude	The geographical longitude of the property.
10	Bedrooms	The most recently recorded bedrooms count.
11	Bathrooms	The most recently recorded count of bathrooms for the property (inclusive of ensuites).
12	Land Area	Total size of the parcel/s in square metres.
13	Transfer ID	Unique record key within the core database for the transfer
14	Contract Date	Contract date of transfer which indicates the date on which the sale price was contractually committed between a vendor and a purchaser. Contract Date for states were VG Contract Date is provided include NSW, VIC, QLD, WA, TAS, ACT only
15	Transaction Date	A proxy Contract Date with Settlement Date substituted for states where no VG Contract Date is provided. Allows for ordering transfers by the time that the transfer occurred.
16	Contract Price	Sale price of transfer indicating the consideration for the property changing ownership (if available)
17	Multi Sale	A ‘stapled’ transfer with multiple properties in the transfer. The contract price is the total for all properties in the transfer.

Table A4 CoreLogic additional linking variables

No.	Data Field	Definition
1	Lot Type	Lot Type component of the parcel's description. SA only.
2	Lot Part	Lot Part component of the parcel's description. QLD only
3	Plan Prefix	Plan Prefix component of the parcel's description. WA, VIC, SA, QLD, NSW only
4	Sub Division	The sub division of the parcel, QLD only.
5	Survey Plan	The Survey Plan. This is not the same as Plan - Plan, lot and location are per parcel while survey plan is per property. NT only.
6	Unit	Unit component of the parcel's description. ACT only
7	Hundred	Hundred component of the parcel's description. SA only
8	District	District component of the parcel's description. ACT only
9	Division	Division component of the parcel's description. ACT only
10	Location Code	The code of the location as supplied in the source data. NT only
11	Location	Location component of the parcel's description. Number describing the planning area - NT only
12	Parish Code	The code of the parish as supplied in the source data. VIC only.
13	Parish	Parish component of the parcel's description. VIC and QLD only
14	Parcel Display Name	Displays the name of the parcel type. This will vary per state as follows: ACT: Section/Block NSW, NT, QLD, SA, TAS, VIC, WA: Lot/Plan
15	Parcel Display Value	Displays the parcel value as a single line string and is used within the RPP platform. Rules for Parcel display per State can be found here
16	Parcel Status	An alphanumeric value to describe the status of the parcel, e.g. 'Approved', 'Proposed', 'Cancelled'
17	Primary Plan	Primary Plan component of the parcel's description. ACT only. Contains Section and Block.
18	Plan Number	Plan Number component of the parcel's description. NSW, VIC, QLD, WA, SA, TAS, NT only. Eg. SP1231
19	Parcel ID	Unique record key within the core database for the parcel of the property
20	Standard Parcel Identifier	The standard parcel description as supplied by the source
21	Jurisdiction ID	Link to external CadLite data source
22	Block	Block component of the parcel's description. VIC, ACT, SA only
23	Section	Section component of the parcel's description. VIC, SA, NSW, ACT only
24	Section Reference	Section Reference component of the parcel's description. SA only
25	Portion	Portion component of the parcel's description. VIC only
26	Accessory Lot Flag	Has its own title – Garage spot belongs to unit. Can't be dwelling place. Must be sold together as one.
27	Crown Allotment Number	Crown Lot component of the parcel's description. NSW, VIC, WA only. A blank value indicates a CROWN lot.

28	Crown Status	A code that identifies a characteristic of the crown description. Valid codes are CROWN LAND, PRE-EMPTIVE RIGHT, CROWN SPECIAL SURVEY, M-Parcels identified by further DESCRIPTION ONLY and VESTED LAND. Further detail may be available if the FURTHER DESCRIPTION field is populated on the Source data. Only VIC

Table A5 Categorical property type variables

CL_Property_Type	CL_Property_Type_Minor	CL_Primary_Land_Use
Business	Accommodation/Restaurant	1-Unknown ABATTOIR Abattoirs Agriculture Agriculture (Extensive) Agriculture N.E.C. Agriculture N.E.C. - Irrigated Agriculture-Intensive Almonds Almonds - Irrigated Animals-Special Aquaculture Aquaculture-Fish Farm Aquaculture-Licenced Beds Aquaculture-Research Facility ART GALLERY BAKERY, RESIDENCE Bank Banks Base Metals - Mines Beekeeping Berry Fruits Berry Fruits - Irrigated Building Units (Primary Use Only) Business and Residence Business Services N.E.C. BUSINESS, RESIDENCE Cafe CAMEL FARM Caretkr Qtrs Cattle Breeding & Fattening Cattle Fattening Cattle Feed Lot Cattle Grazing & Breeding Cattle-Beef Cattle-Beef - Irrigated Pasture Cattle-Beef - Stock Paddocks Cattle-Beef - Stock Watering Cattle-Beef - Stud Cattle-Dairy Cattle-Dairy - Irrigated Pasture Cattle-Dairy - Stock Paddocks Cattle-Dairy - Stock Watering Cattle-Dairy - Stud CEMENT WORKS Cemeteries Cereals Cereals - Irrigated Cereals and Cattle Cereals and Cattle - Irrigated Cereals and Cattle - Stock Watering Cereals and Fodder Cereals and Fodder - Irrigated Cereals and Pigs Cereals and Pigs - Irrigated Cereals and Sheep Cereals and Sheep - Irrigated Cereals and Sheep - Stock Watering Cereals, Stock, Horticulture Cereals, Stock, Horticulture - Irrigated Citrus Citrus - Irrigated Citrus and Others Citrus and Others - Irrigated Closed Roads Com Dev Site COMMERCIAL Commercial Flower and Plant Growing – (outdoor) COMMON - (RESIDENTIAL) Composting Coolstore/Coldstore Cotton Cream Crop Production Fodder Crops Crop Production Mixed/Other Crop Production Other Grains/Oil Seeds Crop Production Wheat Cultural Activities and Nature Exhibitions N.E.C. DAIRY Dairying and Pigs - Irrigated Dairying and Potatoes Dairying and Potatoes - Irrigated DEPOT Detached Dwelling Detached Dwelling (existing) Detached Dwelling (new) Domestic Livestock Grazing Dwelling - Large Housesite Engineering Extractive Fabricated Metal Products, Except Machinery and Equipment N.E.C. Factory FARM Farm - Residence Farm Products, Warehousing Storage and Silos (Excl. Stockyards) FARM, RESIDENCE FARMING Farming Speciality Animals Farming-Cropping Farming-Cropping-All Irrigate Farming-Cropping-Irrig.Scheme Farming-Cropping-Not Irrigated Farming-Cropping-Part Irrigate Farming-Dairy-All Irrigated Farming-Dairying Farming-Dairy-Irrigat.Scheme Farming-Dairy-Not Irrigated Farming-Dairy-Part Irrigated Farming-Grazing/Pastoral Farming-Horses Farming-Horses Open,Run,Bush Farming-Horses-Not Irrigated Farming-Horses-Part Irrigated Farming-Mixed Farming-Mixed-All Irrigated Farming-Mixed-Irrigat.Scheme Farming-Mixed-Not Irrigated Farming-Mixed-Part Irrigated Farming-Mutton Bird-Crown Farming-Mutton Bird-Private Farming-Pigs Farming-Poultry Farming-Speciality FIRE SERVICE Fishing Flats Flowers Flowers - Irrigated Fodder Crops Fodder Crops - Irrigated Forestry Forestry & Logs Forestry N.E.C. - Private Forestry-Artificial Plantation Forestry-Artificial-Authority Forestry-Artificial-Private Forestry-Natural Bush Forestry-Natural Bush-Authorit Forestry-Natural Bush-Private Forestry-Nursery Forestry-Nursery-Private Fuel Outlet/Garage/Service Station Garage/Outbuilding Gas Production General Cropping General Industry General Purpose Factory General Purpose Warehouse G'House/Nurse/Flower-All Irrig G'House/Nurse/Flower-Irr.Scheme G'House/Nurse/Flower-No Retail G'House/Nurse/Flower-Not Irrig G'House/Nurse/Flower-Pt. Irrig Glasshouse Glasshouse - Irrigated Glasshouse Plant/Vegetable Prod Goats Grains Gravel/Stone Grazing/Pastoral-All Irrigated Grazing/Pastoral-Irrig.Scheme Grazing/Pastoral-Not Irrigated Grazing/Pastoral-Open,Run,Bush Grazing/Pastoral-Part Irrigate Grocer Group Title (Primary Use Only) Guest House/Private Hotel Guest Lodge Guest Lodge/Back Packers/Bunkhouse/Hostel Halls and Service Clubrooms Harbour Industries Hardwood Plantation Health Clinic HOL/UNITS Holiday Home / Shack Holiday Home / Shack Private Land Hops Hops-All Irrigated Hops-Not Irrigated Hops-Part Irrigated Horse Stud Horse Stud/Training Facilities/Stables Horses Horses - Irrigated Pasture Horses - Stock Paddocks Horses - Stud Horses and Riding School Horticulture N.E.C. Horticulture N.E.C. - Irrigated Horticulture N.E.C. - Nursery Horticulture/Market Gardening Hotel Hotel/Tavern House House - Business House - Farm House - Kennel House - Land House - Orchard House - Shed House - Stable House - Vineyard House & Flat/S House and Agriculture (Non-Viable) House and Forestry (Non-Viable) House and Horticulture (Non-Viable) House and Livestock (Non-Viable) House and Market Garden (Non-Viable) House and Mixed Farming (Non-Viable) House and Plant Nursery (Non-Viable) House and Poultry (Non-Viable) House Or Cottage House With Unestablished Grounds/Gardens HOUSE, COTTAGE HOUSE, FARM HOUSE, FLAT HOUSE, NURSERY HOUSE, ORCHARD HOUSE, VINEYARD Houses Individual Car Park Site Indoor Sports Centre INDUSTRIAL Industrial Dev Site Kennel/Cattery LA STANDPIPE Light Industry Livestock Livestock N.E.C. Livestock N.E.C. - Irrigated Pasture Livestock N.E.C. - Stock Paddocks Livestock N.E.C. - Stock Watering Livestock Production Beef Cattle Livestock Production Dairy Cattle Livestock Production Sheep Maisonette Major Water Conduits Manufacturing Factory Market Garden Market Garden – Vegetables MARKET GARDEN, RESIDENCE Market Garden-All Irrigate Market Gardening Market Gardening and Orchard Market Gardening and Orchard - Irrigated Market Gardening N.E.C. Market Gardening N.E.C. - Irrigated Market Garden-Irrigat. Scheme Market Garden-Not Irrigated Market Garden-Part Irrigated Median Strips, Plantations, Road Reserves, Standpipes and Undefined Land Wh Medical and Health Services Inc. Veterinary N.E.C. Medical Centre/Surgery Milk-No Quota Milk-
	Agricultural Services	
	Automotive	
	Business/Office Services	
	Fishing & Forestry	
	Health & Education	
	Manufacturing	
	Mining	
	Personal/Other Services	
	Retail Food	
	Retail Trade	
	Tourist & Leisure Services	
	Transport & Storage	
	Wholesale Trade	
Commercial	Industrial Building	
	Office Building	
	Retail Building	
Community	Centres	
	Education	
	Government	
	Religious	
Farm	Cattle Beef	
	Cattle Dairy	
	General	
	Grain & Other Crops	
	Hobby	
	Horticulture/Fruit Growing	
	Other Livestock	
	Poultry	

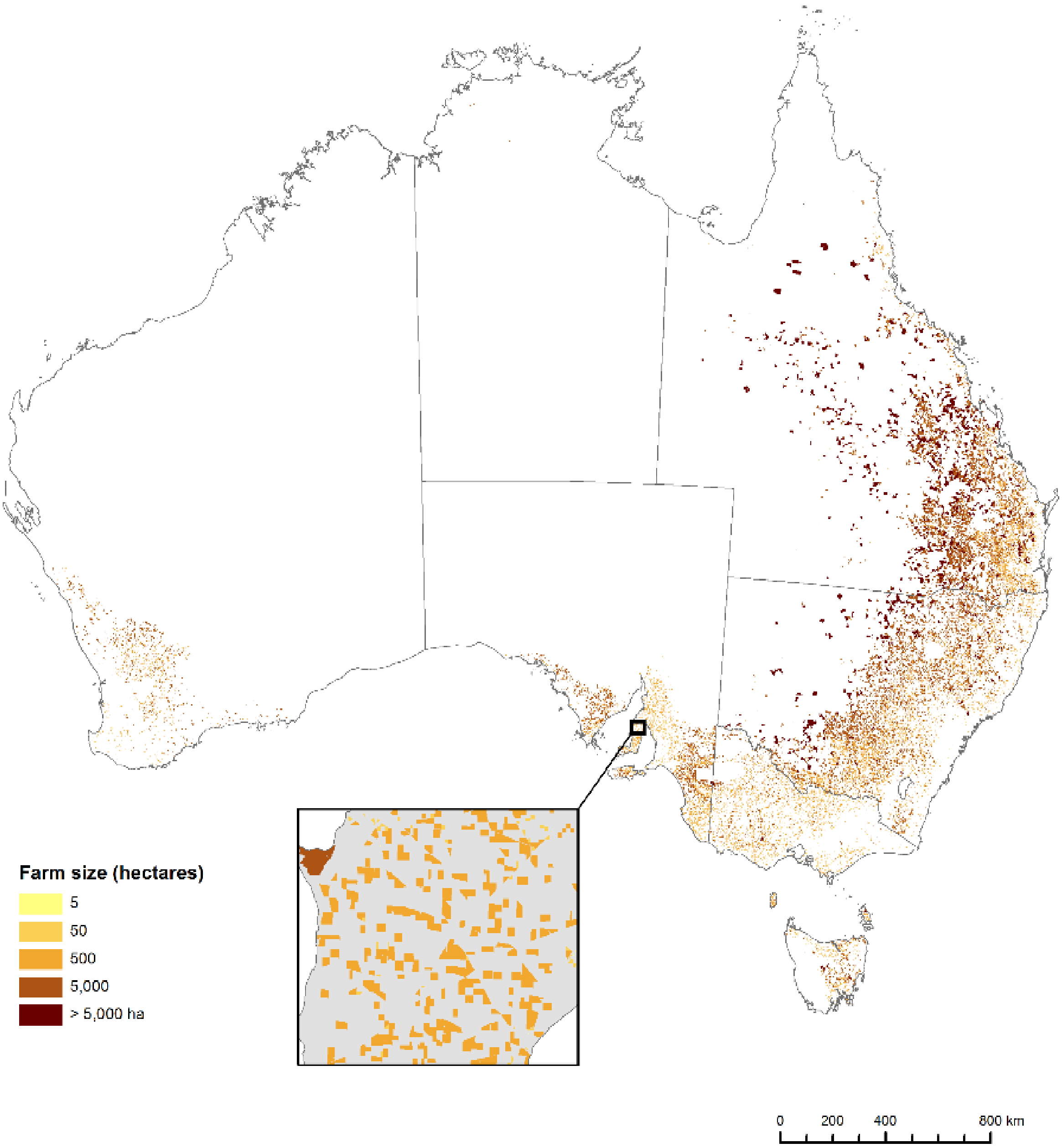
	Sheep	Quota MILL MISCELLANEOUS Miscellaneous Improvements on Residential Land Miscellaneous Improvements on Residential Rural Land Miscellaneous Primary Production Mixed Farming Mixed farming and grazing Mixed farming and grazing with infrastructure Mixed farming and grazing without infrastructure Mixed Farming N.E.C. Mixed Farming N.E.C. - Irrigated Mixed Farming N.E.C. - Stock Watering Mixed Use Occupation Motel Motor Vehicle Sales Multi Unit Dwelling (Flats) Multi-Level Office Building Native Animals Native Hardwood (standing timber) Non-Native Animals Noxious/Offensive Industry (Including Abbatoir) Nurseries (Plants) Nursery (Plants) OBSERVATORY Office OFFICE, WORKSHOP Oil Depot & Refinery Oil Seeds Oilseed Old Folks' Homes Onions ORCHARD Orchard - Residence ORCHARD, RESIDENCE Orchard-All Irrigated Orchard-Irrigation Scheme Orchard-Not Irrigated Orchard-Part Irrigated Orchards Orchards, Groves and Plantations Outbuildings OYO Subdivided Dwelling PADDOCK PADDOCKS Parks and Gardens Including Picnicking Peanuts Piggery Pigs Pigs - Irrigated Pasture Pineapples Place of Worship Plant/Tree Nursery Playhouse/Traditional Theatre Pome Fruits Pome Fruits - Irrigated Pome Fruits - Stock Watering Post Office Potatoes Potatoes - Irrigated Poultry Poultry - Broiler Poultry - Eggs Poultry - Hatchery Poultry - Open Range Poultry (broiler production) Poultry (egg production) POULTRY FARM Poultry N.E.C. Primary Production Primary School Professional Offices Pub/Tavern/Restaurant/Nightclub Public Conveniences R10-Unknown Railway Line Railways (Incl. Rapid Rail Transit and Street Car Transport) Res Dev Site Res Investment Flats Res Rural / Rural Lifestyle Reserve for Drainage or Sewerage Purposes Reservoir, Dams, Bores Residential Retail Plant Nursery Retail Premises (multiple occupancies) Retail Premises (single occupancy) Retail Store/Showroom Retirement Village Unit RIFLE RANGE Rural Living Rural Residential Rural Residential House (House Without Primary Production) Sand and Gravel - Mines SCHOOL FARM Semi-Detached/Terrace Home/Row House Separate Dwelling and Curtilage Service Station Serviced Apart/Holiday Units Shack Shack Site (Not In Conformity With Requirements Under the Planning Act) Shed - Land SHED LAND Sheep and Cattle Sheep and Cattle - Irrigated Pasture Sheep and Cattle - Stock Paddocks Sheep and Cattle - Stock Watering Sheep and Cattle - Stud Sheep Breeding Sheep Grazing-Dry Sheep-Mutton Sheep-Mutton - Stock Watering Sheep-Wool Sheep-Wool - Irrigated Pasture Sheep-Wool - Stock Paddocks Sheep-Wool - Stud Shop Shop & Dwelling (single occupancy) Shopping Group (2 to 6 Shops) SHOPS Shop-Single SHOWGROUND, RACECOURSE, AIRFIELD Showroom/Store Silo - Concrete Cells Silo - Steel Cells Single Res Dwelling Single Strata Unit/Villa Unit/Townhouse Single Unit Dwelling Slaughtering, Preparation, Preserving of Meat, Abattoirs Small Crops & Fodder - Irrigated Small Crops & Fodder - Non Irrigated Small Seeds Small Seeds - Irrigated Social Service and Welfare Provision Soft Fruit & Nut Soft Fruit & Nut-All Irrigated Soft Fruit & Nut-Not Irrigated Soft Fruit & Nut-Part Irrigate Softwood Plantation Solid Waste Disposal Special Tourist Attraction Specialised Cropping Sportsclubs/Facilities STANDPIPE State(Secondary Land Use Only) Steep Or Rocky Land Stock and Poultry Stockyard Stockyard Services - Stables Stone and Pome Fruits Stone and Pome Fruits - Irrigated Stone and Pome Fruits - Stock Watering Stone Fruits Stone Fruits - Irrigated Stone Fruits and Others Stone Fruits and Others - Irrigated Strata Unit or Flat (Unspecified) Strata/Subdivided Office Sub-div Land (Multi Lot) SUBDIVIDED LAND - DISCOUNTED BY LG Subdivisional Land (In globo/Potential) Sugar Cane SURGERY Swamp Or Land Subject to Flooding Telecommunications N.E.C. Terrace Tomatoes - Irrigated Tourist Park/Caravan Park/Camping Ground Townhouse Townhouse - Defined As Home Unit With Both Ground and First Floor Areas Training Facilities TRIPLEX UNIT Tropical Fruits Truck Freight Garaging and Equipment Maintenance Truck Freight Terminal Turf Farms Undetermined Land Use Undeveloped Reserve Utility Services-Sewer/Water Vacant - Large Housesite Vacant Allotment Conservation Or Recreation Vacant Gov Admin Dev Site Vacant In globo Residential Subdivisional Land Vacant Land Vacant Land - Native Veg/Bushland with Covenant Vacant Land - Native Vegetation/Bushland VACANT LAND - NON-RESIDENTIAL VACANT LAND - RESIDENTIAL Vacant Land - Rural Residential (No Primary Production) Vacant Land With Minor Improvements (Rural Living) Vacant Land With Minor Improvements (Urban) Vacant Land-Urban Vacant Res Rural / Rural Lifestyle Vacant Residential Dwelling Site/Surveyed Lot Vacant Rural Land (Excl 01 & 04) Vacant Urban Land Vacant-Englobo/Broad Hectares Vacant-Residential Vacant-Rur Resid.With Rural Cl Vacant-Rural Residential Vegetables Vegetables - Irrigated Vegetables - Stock Watering Vines Vines - Irrigated Vines - Stock Watering Vines (Non-Viable) Vines and Others Vines and Others - Irrigated Vines and Others - Nursery Vines and Others - Stock Watering Vines and Stock Vines and Stock - Irrigated Vineyard VINEYARD, RESIDENCE Vineyard-All Irrigated Vineyard-Irrigation Scheme Vineyard-Not Irrigated Vineyard-Part Irrigated Vineyards Warehouse Warehouse & Bulk Stores Warehouse/Showroom Water Catchment Area Water Storage Water Store Dam (Non-Catchment) Water, Sewage Disposal Wind Farm Electricity Generation Wooded Area WORKSHOP YARD (blank)
Flats	Boarding House	
	Self Contained	
House	Acreage	
	Dual Occupancy	
	Duplex	
	Multi Storey	
	One Storey / Lowset	
	Semi Detached	
	Standard	
	Terrace	
	Two Storey / Highset	
Land	General	
	Government	
	Industrial	
	Office/Retail	
	Parks / Reserves	
	Res Acreage	
	Res Development	
	Res House	
	Rural Acreage	
Storage Unit	Car Space	
Unit	Highrise	
	Lowrise	
	Penthouse	
	Quadraplex	
	Standard	
	Studio	
	Townhouse/Villa	
	Triplex	
Blank		

Table A6 Constructed identifiers using CL_PRIMARY_LAND_USE

Constructed Identifier	**Cl_primary land use contains:**
Cropping	“Crop”, “Cropping”, "Crops", "Grain", "Grains", "Cereals"
Livestock	"Cattle", "Pastoral", "Sheep", "Beef", "Livestock", "Grazing", "Wool", "Pigs", "Poultry", "CAMEL", "Mutton", "Goats"
Mix	"Mix", "Mixed", "Cereals and Sheep", "Cereals and Cattle"
Dairy	"Dairy", "Milk", "Cream"
Vineyard	"Vines", "Vineyard", "Vineyards", "Vinyard", "Vinyards"
Sugar	"Sugar"
Lifestyle	"Lifestyle", "House", "Housesite", "Dwelling", "Residential"
Horticulture	"Orchards", "Vegetables", "Citrus", "Pineapples", "Fruits”, "Groves", "Cotton", "Peanuts", "Pineapples", "Pome", "Almonds" "Berry"
	Cl_primary land use equals:
Irrigated	"Small Crops & Fodder - Irrigated", "Vines - Irrigated", "Farming-Mixed-Part Irrigated", "Vines - Irrigated", "Vines and Others - Irrigated", "Farming-Dairy-Part Irrigated", "Cattle-Dairy - Irrigated Pasture", "Farming-Dairy-Part Irrigated", "Grazing/Pastoral-Part Irrigate", "Citrus and Others - Irrigated", "Vegetables - Irrigated", "Citrus - Irrigated", "Stone Fruits - Irrigated", "Stone and Pome Fruits - Irrigated", "Cotton", "Peanuts", "Pineapples", "Vegetables - Irrigated", "Almonds - Irrigated"

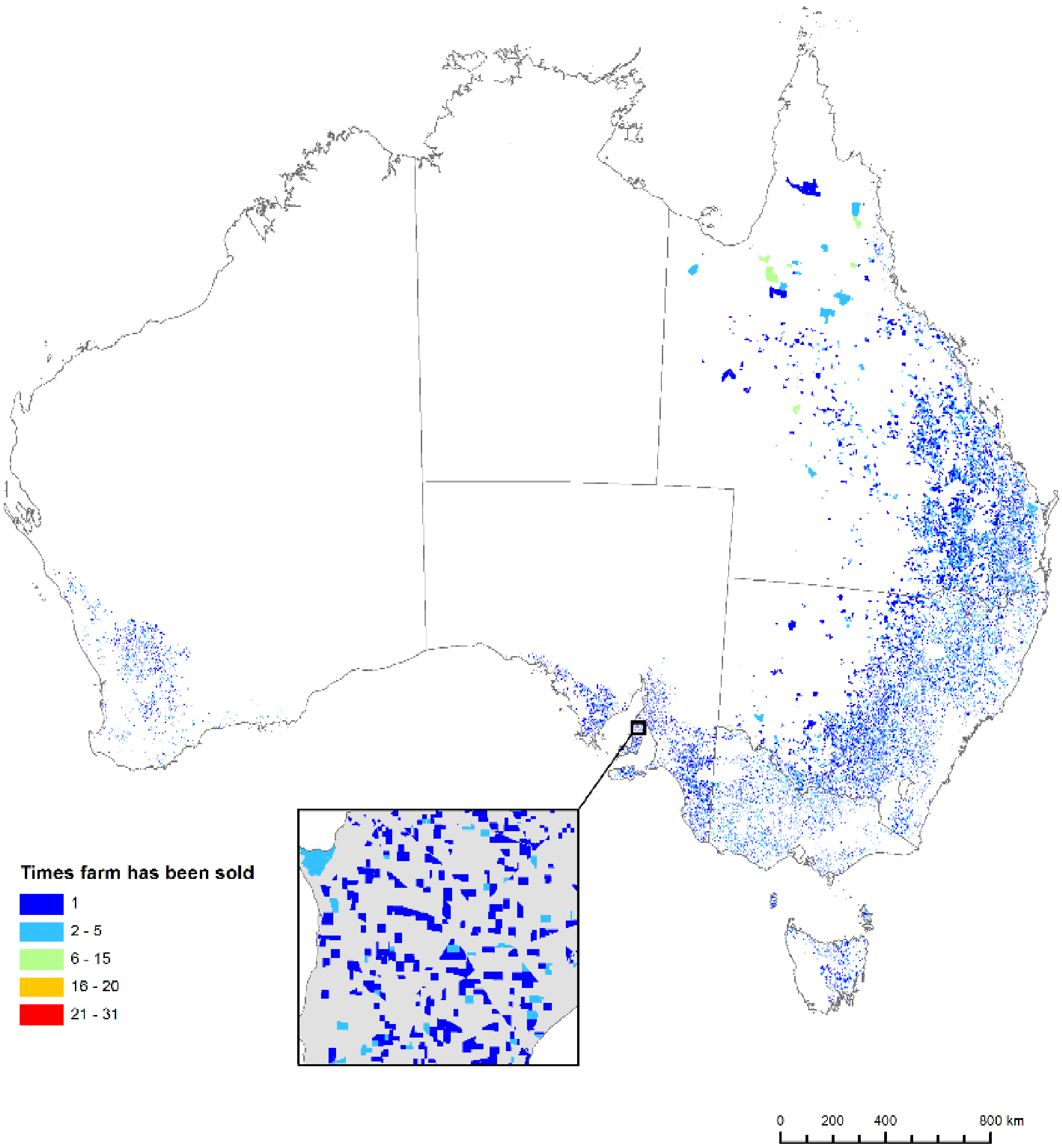
Appendix B: Geospatial statistical summaries

Map A2 Farm size in clean dataset 1975-2018



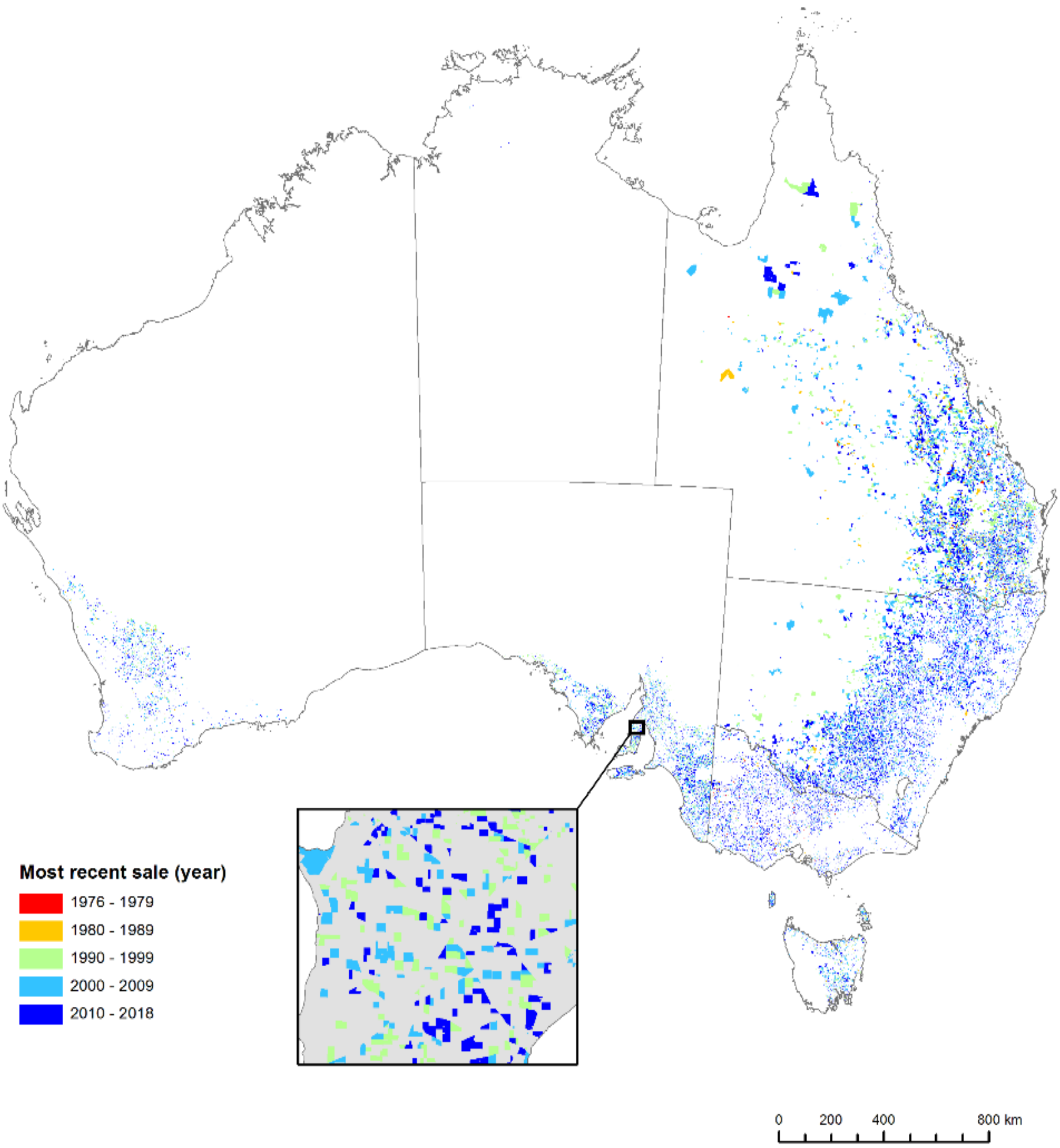
Source: Authors estimates

Map A3 Count of sales 1975-2018



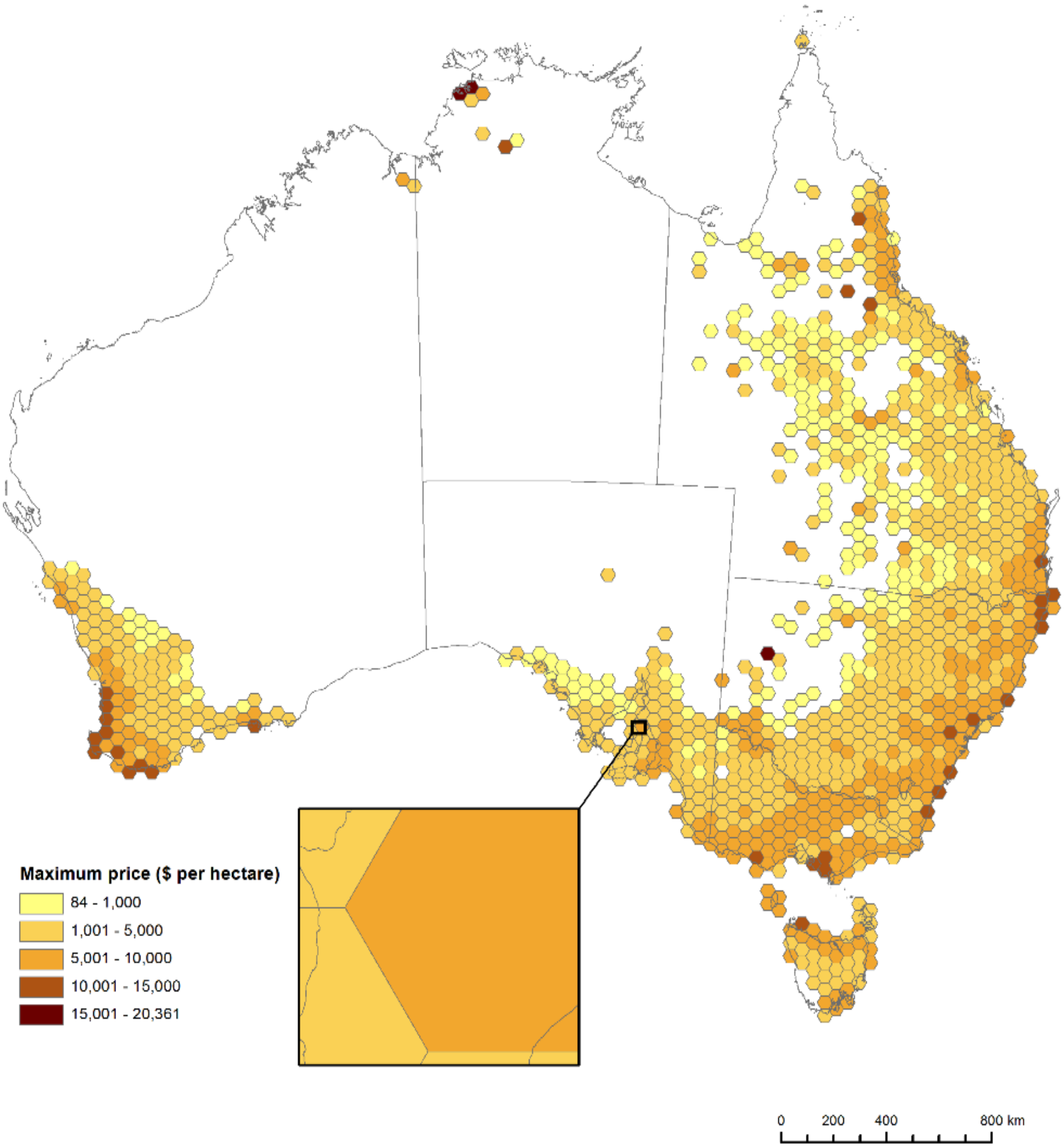
Source: Authors estimates

Map A4 Latest year of sale 1975-2018



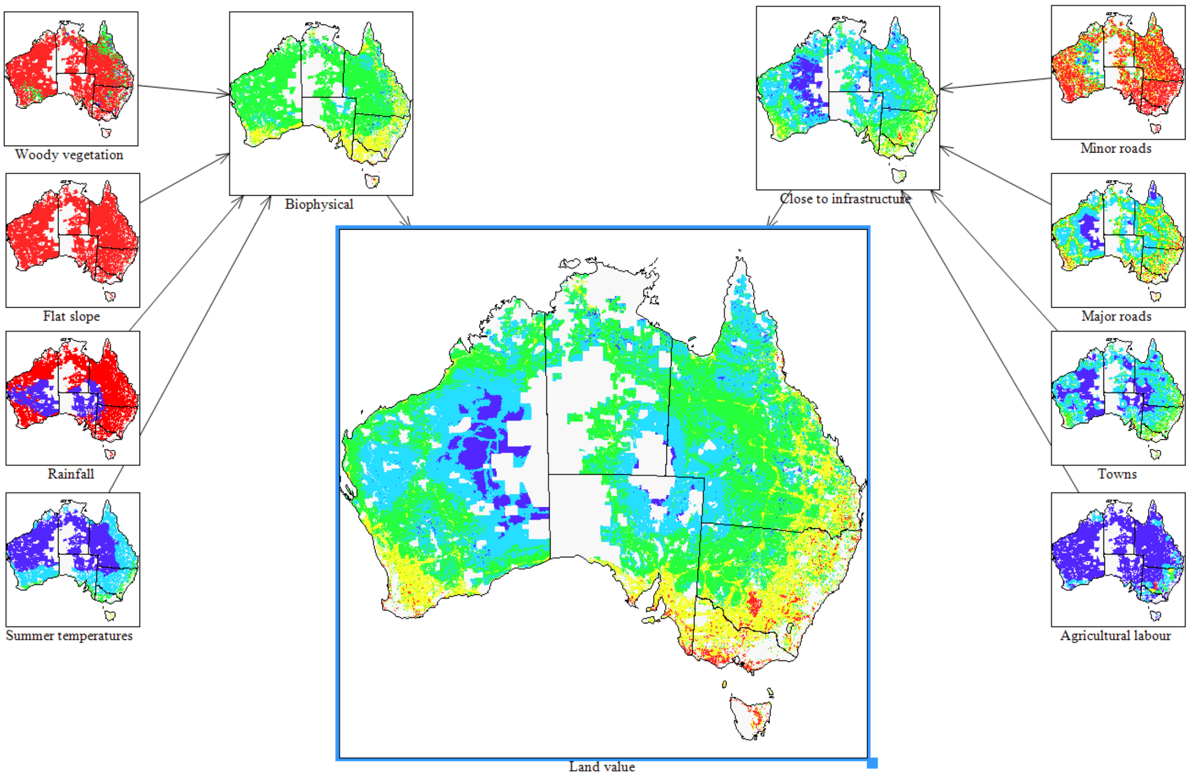
Source: Authors estimates

Map A5 Price per hectare (maximum) by parcel in clean dataset 1975-2018 (as hexbins)



Source: Authors estimates

Map A6 Experimental example of land value modelling using MCAS-S



Source: Authors estimates

Note: Once the direction and magnitude of the explanatory variables in relation to the dependent variable is known; the spatial datasets for these explanatory variables can be weighted and overlaid. This allows us to identify 'optimal areas' based on the explanatory variables (see red areas in the central map). It may be possible to compare these modelled values to actual values (in the CoreLogic dataset) to identify farmland that is either undervalued or overvalued. **The modelling results in Figure A10 are for demonstrative purposes only and do not reflect the results in this paper.**

References

Australian Bureau of Statistics (ABS) 2017, Australian Statistical Geography Standard (ASGS): Volume 4 - Significant Urban Areas, Urban Centres and Localities, Section of State, July 2016 (ABS Cat no. 1270.0.55.004)

Australian Bureau of Statistics (ABS) 2019,- Australian National Accounts: National Income, Expenditure and Product, Dec 2018, Table 37. Industry Gross Value Added, Chain volume measures, Annual https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5206.0Dec%202018?OpenDocument Annual 1975 to 2018 (ABS Cat no. 5206.0)

Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) 2019, Catchment scale land use of Australia – Update December 2018, retrieved on 20/06/2019 from http://www.agriculture.gov.au/abares/aclump/Pages/land-use/Catchment-scale-land-use-of-Australia-2018.aspx

Beaton, A. E., and Tukey J. W. 1974, The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* 16: 147–185.

Bureau of Meteorology (BoM) 2019a, Gridded Daily Rainfall Metadata, retrieved on 20/06/2019 from http://www.bom.gov.au/climate/austmaps/metadata-daily-rainfall.shtml

Bureau of Meteorology (BoM) 2019b Daily maximum temperature for Australia, retrieved on 20/06/2019 from www.bom.gov.au/jsp/awap/temp/index.jsp

Chicoine, D. 1981, Farmland values at the urban fringe: An analysis of sale prices. land *Economics*, 57(3), 353-362. doi:10.2307/3146016

Cho, M. 1996, House price dynamics: a survey of theoretical and empirical issues, *Journal of Housing Research*, 7(2), pp 145–172.

Conniffe D. and D. Duffy 1999, Irish house price indices – methodological issues, *Economic and Social Review*, 30(4), pp 403–423

De Haan, J. and Diewert, W. E., 2013, *Handbook on Residential Property Prices Indices (RPPIs)*, Methodologies and working papers. Eurostat, Luxembourg

Dent, J. and Ward, M. B. 2014, Food bowl or folly? The economics of irrigating Northern Australia, Monash Business School, Department of Economics

Department of Environment and Energy 2018, National Forest and Sparse Woody Vegetation Data (Version 3, 2018 Release), retrieved on 20/06/2019 from https://data.gov.au/data/dataset/national-forest-and-sparse-woody-vegetation-data-version-3-2018-release

Department of Environment and Energy 2019, National Forest and Sparse Woody Vegetation Data (Version 3, 2018 Release), available via https://data.gov.au/data/dataset/national-forest-and-sparse-woody-vegetation-data-version-3-2018-release

Drescher, K., Henderson, J. and McNamara, K. 2001, Farmland prices determinants, In *American Agricultural Economics Association Annual Meeting*, Chicago, IL, August 2001

Eves, C. 2016, The analysis of NSW rural property: 1990-2014, 22nd Annual Pacific-Rim Real Estate Society Conference, Sunshine Coast, Australia, 17-20 January 2016

Geoscience Australia 2011, Digital Elevation Data, retrieved on 20/06/2019 from http://www.ga.gov.au/scientific-topics/national-location-information/digital-elevation-data

Geoscience Australia 2019, Water Observations from Space, retrieved on 20/06/2019 from https://www.ga.gov.au/scientific-topics/community-safety/flood/wofs

GEODATA 2006, GEODATA TOPO 250K Series 3 (Shape file format). Commonwealth of Australia (Geoscience Australia).dataset.
http://pid.geoscience.gov.au/dataset/ga/64058

Grains Research & Development Corporation (GRDC) 2019, Growing Regions, retrieved on 24/07/2019 from https://grdc.com.au/about/our-industry/growing-regions

Hamilton, L.C. 2012, *Statistics with Stata*: version 12, Cengage Learning.

Hansen J (2006), Australian house prices: a comparison of hedonic and repeat-sales measures, Reserve Bank of Australia Research Discussion Paper No 2006-03.

Hill, R.J., Melser, D. and Reid, B. 2010, Hedonic imputation with geospatial data: An application of splines to the housing market, Mimeo.

Hill, R.J 2011, Hedonic price indexes for housing, *OECD Statistics Working paper* 2011/1, OECD Publishing, Paris.

Hill, R.J. and Scholz, M., 2014. Incorporating geospatial data in house price indexes: A hedonic imputation approach with splines, *Graz Economics Pa*pers, 5.

Huang, H., Miller, G.Y., Sherrick, B.J. and Gomez, M.I. 2006, Factors influencing Illinois farmland values, *American Journal of Agricultural Economics*, 88(2), pp.458-470.

Huber, P. J. 1964, Robust estimation of a location parameter, *Annals of Mathematical Statistics* 35: 73–101.

King, D.A. and Sinden, J.A. 1988, Influence of soil conservation on farm land values, *Land Economics*, 64(3), pp.242-255.

Leys, J., Chappell, A., Mewett, J. and Barson, M. 2017, Wind erosion assessment for National Landcare Program, *NSW Office of Environment and Heritage, Australia*.

Ma, S. and Swinton, S.M. 2011, Valuation of ecosystem services from rural landscapes using agricultural land prices, *Ecological Economics*, 70(9), pp.1649-1659.

Martin, P. and Topp, V. 2019, Farm performance: broadacre and dairy farms, 2016-17 to 2018-19, retrieved on 06/09/2019 from http://www.agriculture.gov.au/abares/research-topics/surveys/farm-performance#overview

McFadden, D. 1977, *Quantitative methods for analysing travel behaviour of individuals: Some recent developments*. Institute of Transportation Studies, University of California.

McKenzie N.J., Hairsine P.B., Gregory L.J., Austin J., Baldock J.A., Webb M.J., Mewett J., Cresswell H.P., Welti N., Thomas M. 2017, Priorities for improving soil condition across Australia's agricultural landscapes, Report prepared for the Australian Government Department of Agriculture and Water Resources. CSIRO, Australia.

Mendelsohn, R., Nordhaus, W., & Shaw, D. 1994, The Impact of Global Warming on Agriculture: A Ricardian Analysis, *The American Economic Review*, 84(4), 753-771. Retrieved from http://www.jstor.org/stable/2118029

Palmquist, R.B. and Danielson, L.E. 1989, A hedonic study of the effects of erosion control and drainage on farmland values, *American Journal of Agricultural Economics*, 71(1), pp.55-62.

Pyykkönen, P. 2006, Factors affecting farmland prices in Finland, *Pellervo Economic Research Institute*, Publication No. 19, Helsinki

Rural Bank 2016, Australian farmland values 2015, Volume 1, retrieved on 06/10/2017 from https://www.ruralbank.com.au/assets/responsive/pdf/publications/farm-land-values-2015.pdf

Ryan, P. 2014, pctile – Create variable containing percentiles, viewed on 30/07/2019 from https://www.stata.com/manuals13/dpctile.pdf

Sheng, Y. and Chancellor, W. 2019, Exploring the relationship between farm size and productivity: Evidence from the Australian grains industry, *Food Policy*, *84*, pp.196-204.

Soriano, F. 2008, Modelling the relationship of energy efficiency attributes to hours price: the case of detached houses sold in the Australian Capital Territory in 2005-06, full statistical report to the *Energy Efficiency Rating and House Price in the ACT*, National Framework for Energy Efficiency, Department of the Environment, Water, Heritage and the Arts (from https://www.buildingrating.org/document/energy-efficiency-rating-and-house-price-act).

Stata 2014a, qreg – Quantile regression, viewed on 30.07/2019 from https://www.stata.com/manuals13/rqreg.pdf

Stata 2014b, Bootstrap sampling and estimation, from https://www.stata.com/manuals13/rbootstrap.pdf#rbootstrap

Reserve Bank of Australia (RBA) 2019 Money and Credit Statistics Rural Debt by Lender – D9 from https://www.rba.gov.au/statistics/tables/

Triplett, J. 2004, Handbook on hedonic indexes and quality adjustments in price indexes: Special application to information technology products, *OECD Science, Techology and Industry* Working Papers, 2004/9, OECD Publishing.

Waltl, S.R., 2019. Variation across price segments and locations: A comprehensive quantile regression analysis of the Sydney housing market. *Real Estate Economics*, *47*(3), pp.723-756.