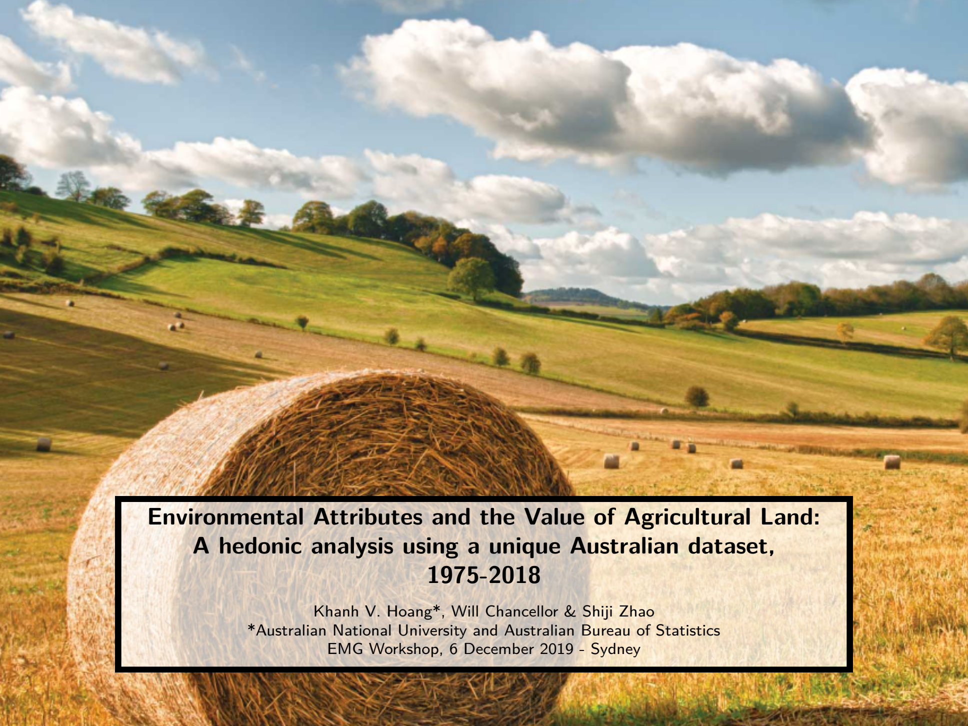

Environmental Attributes and the Value of Agricultural Land: A hedonic analysis using a unique Australian dataset, 1975-2018

Khanh V. Hoang*, Will Chancellor & Shiji Zhao
*Australian National University and Australian Bureau of Statistics
EMG Workshop, 6 December 2019 - Sydney

Motivation

- Land is a fundamental asset of a farm.
- ABS estimates land to account for over 60% of the productive capital stock in the agricultural industry.
- The emerging challenges facing the agriculture industry of climate change and environmental degradation mean that a better understanding of the determinants of farmland productivity and hence its value is more important than ever.

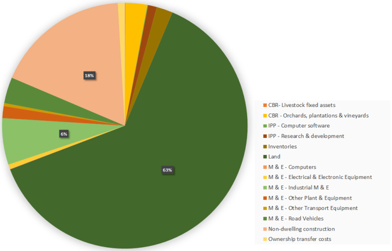

Figure 1: Share of Productive Capital Stock in Agricultural Industry by Asset Type, 2017-18

Motivation

- While current price values of the stock of agricultural land are available from the national balance sheet in the Australian System of National Accounts, there is no equivalent volume indicator.
- ABS only measures the land area with no adjustment made for quality
- As a consequence, soil degradation due to land management choices or exogenous factors such as climate and rainfall are ignored.

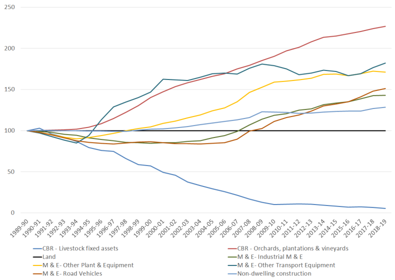

Figure 2: Agricultural Productive Capital Stock Growth, 1989-90 = 100

Overview

- Uniqueness of farm properties are due to many factors
- Location usually included in hedonic models at some aggregate level, not at individual property level.
- The data available for this study is unique containing a census of farm-level sales record from 1975 to 2018
- Geographic Information Systems (GIS) mapping has been used to integrate spatial data of the farm to an extensive range of characteristics of the land parcel.
- This study aims to quantify the link between the environmental attributes of the farmland and its value over time for Australia.
- Also assesses the suitability of different spatial hedonic price models to construct constant quality price indexes at the national level

The Data

- Base dataset is Corelogic property sales data in agricultural in Australia - around 700,000 transactions
- Corelogic data variables:
    - Geographic longitude and latitude
    - Sale price
    - Land area
    - Contract date
    - Property type
    - Land use
    - Bedrooms
    - Bathrooms
    - Multi Sale

Linking Spatial Data

- Environmental attributes of the property such as soil and climate at the farm level are spatially linked
- Corelogic data has been geocoded by ABARES to allow spatial datasets to be overlaid including topography data from:
  - Geoscience Australia Shuttle Radar Topographic Mission (SR) digital elevation model - slope of the land
  - Water Observations from Space - water coverage
  - Bureau of Meteorology AWAP grids - average rainfall and temperature by year
  - CSIRO National Landcare Program - soil quality

Big data - big job cleaning the data!

- Remove duplicates, missing or zero values, transactions prior to 1975 and after 2018
- Remove non-farm records
  - Transactions categorised by Core-logic as non-farm
  - Non agricultural activities such as mining, abboittors, urban corridor, hobby farms, residential properties etc.
  - Exclude land sold for less than $50 per hectare. We suspect these are transactions between family members
  - Remove multi-sale transactions where property parcels sold as a group
  - Any transactions where price exceeded $30,000 per hectare, and the land was less than 2 hectares were also excluded. This is to remove small hobby farms and suspected residential properties
- Remove extreme outliers where the sale price per hectare is above or below the Inter-quartile Range (IQR) by year at the State level. Usually 1.5 times the IQR: $Q1 - 1.5(IQR)$, $Q3 + 1.5(IQR)$ (Turkey Method)
- Total of around 130,000 farm land transactions in scope

Australia

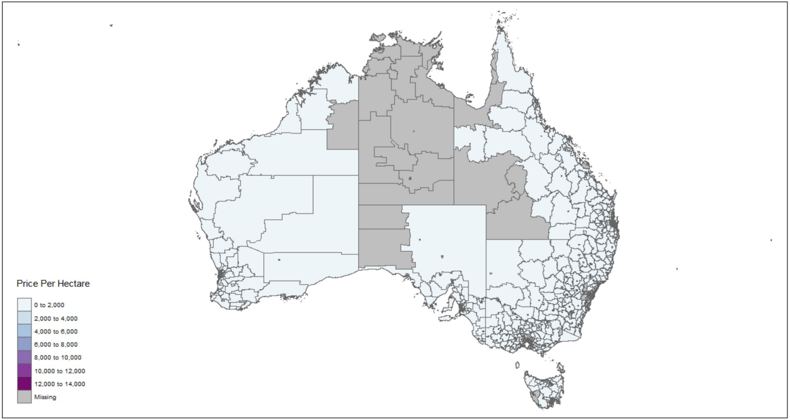

Figure 3: Average Price per Hectare by Statistical Area 2, 1975-2018

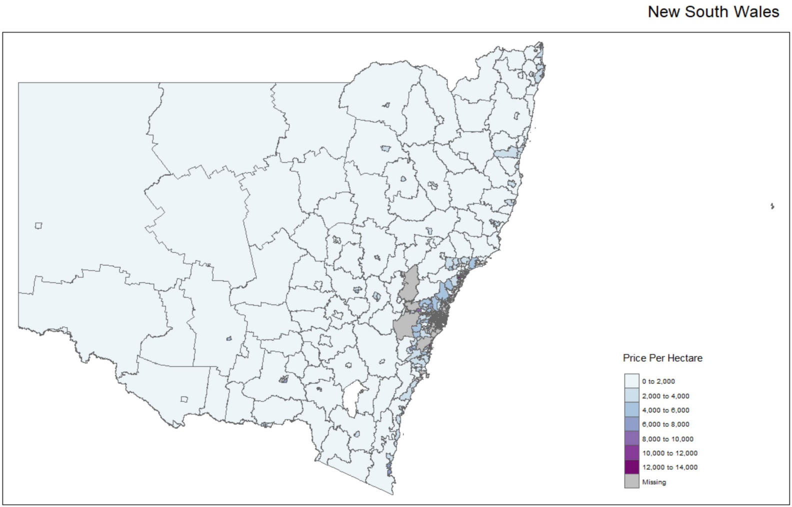

Figure 4: Average Price per Hectare by Statistical Area 2, 1975-2018

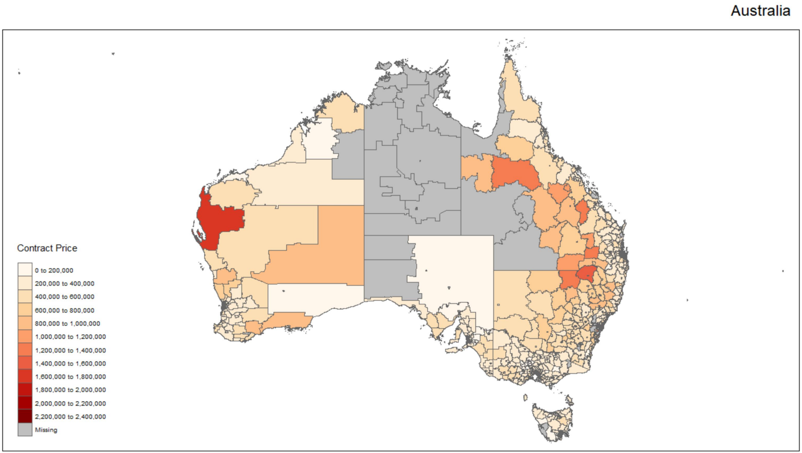

Figure 5: Average Contract Price by Statistical Area 2, 1975-2018

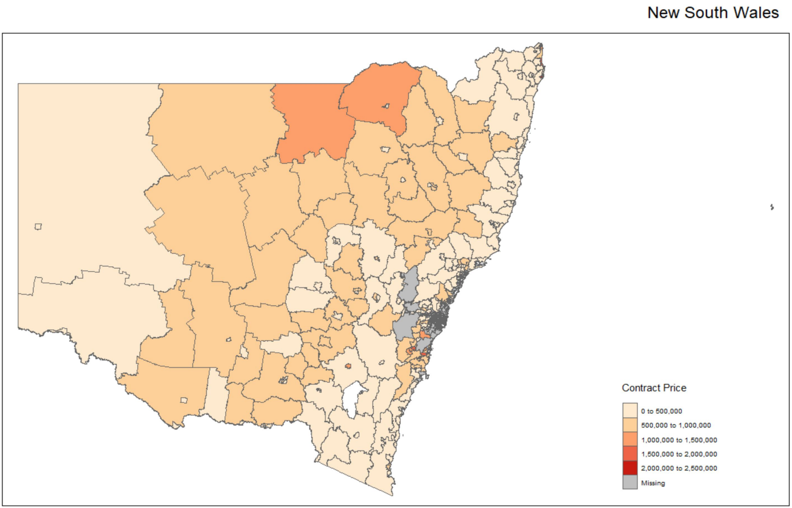

Figure 6: Average Contract Price by Statistical Area 2, 1975-2018

Data type	Variable	Description
Farm size	Hectares	Land size for transacted property
Location	Latitude and Longitude	Geospatial location coordinates
	ABS Statistical Regions	SA1, SA2, SA3, SA4
	ABARES and GRDC zone	High Rainfall Zone, Wheat Sheep Zone, Pastoral Zone GRDC Southern, GRDC Northern, GRDC Western
	Distance to nearest town	Distance from property to the nearest town with population of 10,000.
	Distance to road network	Distance from property to the closest road network
Structure	Number of bedrooms	Number of bedrooms if house is present on transacted property
	Number of bathrooms	Number of bathrooms if house is present on transacted property
	Residential building	Number of residential buildings points
	Farm building	Number of agricultural buildings points
Environmental Attributes	Primary Land Use	Dummy variable to identify grain and crops, livestock, mixed farming, dairy, vineyards, vacant, hobby and horticulture
	Land slope (flat)	Percentage of land parcel that is flat - SRTM digital elevation model
	Land slope (undulating)	Percentage of land parcel that is undulating, hilly or steep
	Rainfall	Annual rainfall assigned to farm by year of sale - BoM AWAP
	Maximum temperature	Average annual maximum temperature assigned to farm
	Minimum temperature	Average annual minimum temperature assigned to farm
	Soil Acid	Percentage of farmland with acidification risk
	Water Erosion	Percentage of farmland with water erosion risk
	Wind Erosion	Percentage of farmland with wind erosion risk
	Water cover 5%	Percentage of farmland with water cover based on Geoscience Australia Water Observations from Space data

Figure 7: Agricultural Land Characteristics

Options for Region Variable

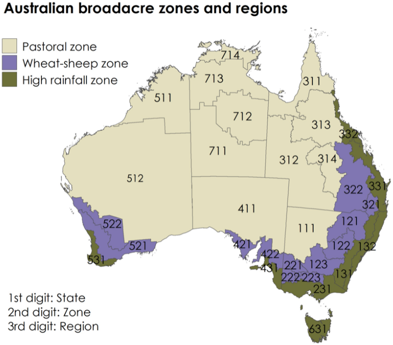

Figure 8: ABARES - AAGIS zones

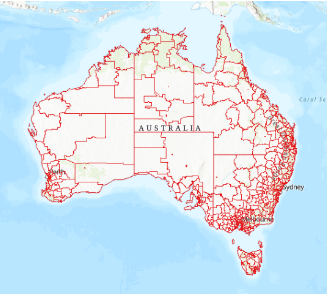

Figure 9: ABS Statistical Area 2

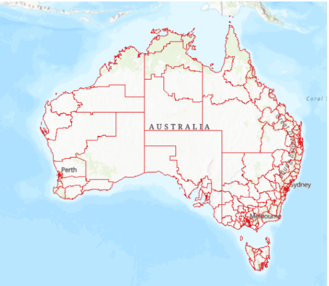

Figure 10: ABS Statistical Area 3

Method for Compiling Farmland Hedonic Price Indexes; Time Dummy Method

- The time-dummy method is the original hedonic method which typically uses the semi-log functional form (see Diewert (2003))

- A standard semi-log formulation is as follows:

$$y = X\beta + D\delta + \epsilon \tag{1}$$

$$\hat{P}^t_{TD} = exp(\hat{\delta}_t) \tag{2}$$

where

- y is a $Hx1$ vector of log-prices ($y_f = lnP_f$)
- X is an HxA matrix of physical characteristics
- D is an HxB matrix of time dummy variables
- ϵ is an $Hx1$ vector of random errors
- $\hat{P}^t_{TD}$ is a vector of constant quality prices

Time Dummy Model (cont)

Advantages:

- Its simplicity - the price index follows immediately from the estimated pooled time dummy regression equation
- No restrictions on parameters
- Functional form is neither continuous nor smooth with respect to time
- Add environmental attributes of the land
- Add categorical variables

Potential problems:

- (Multi)collinearity between land characteristics
- Heteroskedasticity
- Lack of flexibility, in that the shadow prices cannot evolve over time

Our Models

1. Semi-log with region dummies ("OLSD")

$$y = X\beta + D\delta + R\gamma + \epsilon \quad (3)$$

where:

- y is a $Hx1$ vector of log-price ($y_f = lnP_f$)
- X is an HxA matrix of physical characteristics
- D is an HxB matrix of time dummy variables
- R is an HxC matrix of region dummy variables
- ϵ is an $Hx1$ vector of random errors

The parameters to be estimated are:

- $Ax1$ vector of characteristic shadow prices β
- $Bx1$ vector of time-dummy shadow prices δ
- $Cx1$ vector of regional-dummy shadow prices γ
- The first element of β is the intercept where X consists of ones

Our Models (cont)

2. Semilog with geospatial spline ("GAM")

$$y = X\beta + D\delta + s(xlat, xlong) + \epsilon \tag{4}$$

Where:

- y is a H x 1 vector of log-price
- X is an H x A matrix of physical characteristics
- D is an H x B matrix of time dummy variables

The parameters to be estimated are:

- A x 1 vector of characteristic shadow prices β
- B x 1 vector of time dummy shadow prices δ
- s(xlat, xlong), the geospatial spline function defined on the latitudes and longitudes of the property

- The model is estimated with an optimal low rank approximation of a thin plate spline (which has n unknown parameters)
- The smoothing parameter is selected using Restricted Maximum Likelihood (REML).
- The spline is estimated using the GAM function from the R package mgcv.
- It introduces non-parametric terms to the estimation of land prices

Empirical results

Three models estimated at the Australian level:

1. OLSD - Variation across postcodes
2. GAM1 - Variation across individual properties using spatial co-ordinates
3. GAM2 - Use spline function to smooth variation across individual properties selected land characteristics and spatial co-ordinates

According to AIC as well as BIC: OLSD performs marginally better than the model using splines.

	Akaike Info. Criterion	Bayesian Info. Criterion
OLSD	196250	205231
GAM1	208725	209296
GAM2	208725	209296

Model Performance?

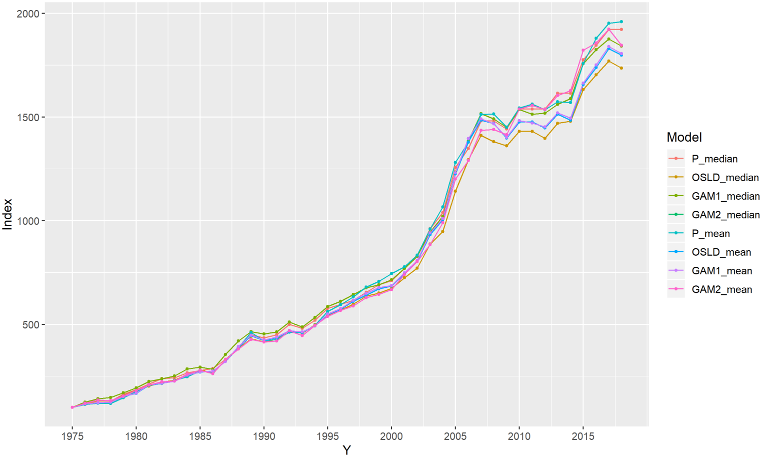

Figure 11: Australian Farmland Price Indexes, 1976 - 2018, Predicted vs Actual

Table 1: Model comparison of selected characteristics[1]

	OLSD	GAM1	GAM2
log (Hectare)	0.588***	0.551***	-
bath	0.059***	0.080***	0.059***
bed	−0.005**	0.012***	0.017***
acid pct	0.0004***	0.0002***	0.001***
wind pct	0.0003***	0.0001*	0.001***
water pct	0.0003***	0.0002***	0.001***
min temp	0.035***	0.061***	0.032***
max temp	−0.013***	−0.016***	0.005**
Rainfall	0.00004***	0.0001***	***
Building	0.027***	0.024***	0.026***
Residential	0.079***	0.082***	0.076***
Distance to road	−0.0001***	−0.0001***	***
Distance to town	−0.002***	−0.001***	***
No Slope	1.013***	1.017***	1.608***
Slope	0.929***	0.956***	1.533***
water 5%	0.030***	0.097***	0.097***
Constant	6.373***	5.517***	7.812***
SA2 Dummy	***	-	-
Time Dummy	***	***	***
Observations	133,724	133,724	133,724
Adjusted R^2	0.72	0.69	0.65

*p$<$0.1; **p$<$0.05; ***p$<$0.01

Note:Dummy variables and smoothed variables are not included in table

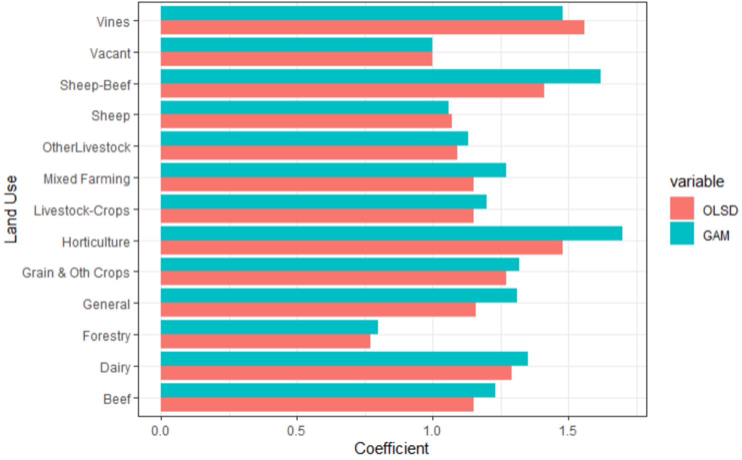

Figure 12: Comparison of Coefficient by Land Use types by Models

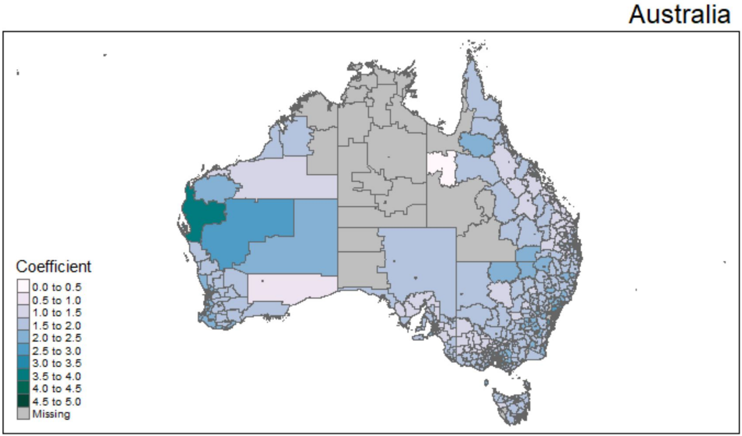

Figure 13: OSLD Model - SA2 region coefficients

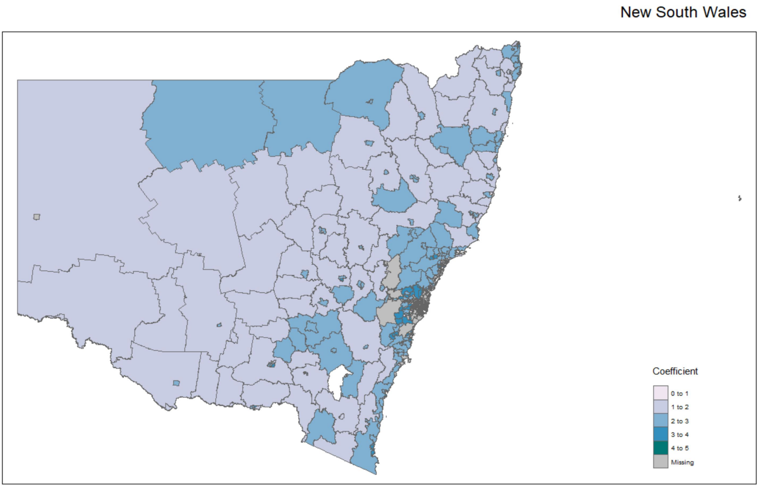

Figure 14: OSLD Model - SA2 region coefficients

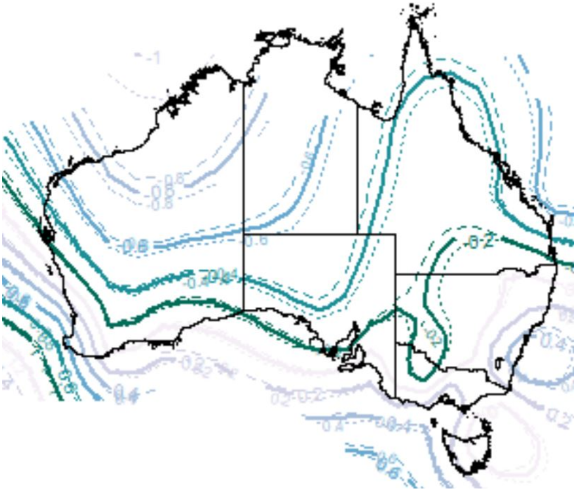

Figure 15: GAM1 Model - Smooth location coefficient of Longitude and Latitude

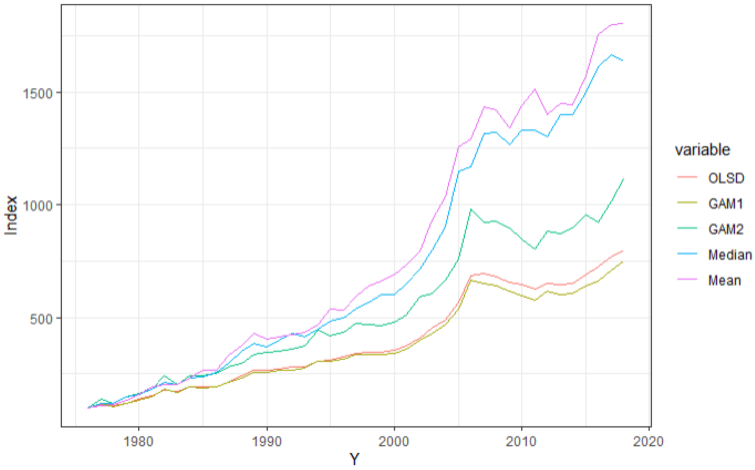

Figure 16: Constant Quality Land Price Indexes, 1975 - 2018

Conclusion

- Location is an important driver of farmland prices - significant variations in the value of farmland across location.
- We show that splines (or some other non-parametric method) provide a flexible way of incorporating geospatial data into a hedonic model. However, a suitable regional indicator is often just as effective
- Farm characteristics - Rainfall, temperature, farm size, land use, land gradient and structures on the farm all appear to impact farmland values
- Characteristics such as distance to nearest road or town, rainfall, water cover and soil characteristics have more nuanced relationships with farm land value.
- Land price indexes are volatile due to outliers and nonlinear relationship between farmland prices and land characteristics
- The constant quality price index revises downward the cumulative price change from 1975 to 2018 by between 60% and 140%, depending on how location are included.

Future Work

- Explore other non-parametric hedonic models
- Explore single or double hedonic imputation methods to create land price indexes
- Create regional and sub-regional price indexes
- Impact on agricultural productivity when constant quality price indexes for land is used
- Include more farmland characteristics, e.g. farm yield?
- Second PhD?

Thank You

Contact Author: Khanh V. Hoang | E: Khanh.hoang@abs.gov.au | P: (02) 6252 5986

Examples of Variables used to Explain Land Values

Variable	Reference
Agricultural returns - Monetary variables	- Market revenues (Carlberg 2002; Barnard *et al.* 1997; etc.) - Returns to land (Goodwin *et al.* 2005 & 2010; Weerahewa *et al.* 2008) - Net income (Devadoss & Manchu 2007) - Producer price of wheat (Goodwin & Ortalo-Magne 1992)
Agricultural returns – Non monetary variables	- Yield (Pyykkonen 2005; Devadoss & Manchu 2007; Latruffe *et al.* 2008) - Soil quality, temperature and precipitation, irrigation, presence of intensive crops (Barnard *et al.* 1997) - Fraction of cropland (Gardner 2002) - Proximity of a port (Folland & Hough 1991)
Government payments	- Total government payments (Devadoss & Manchu 2007; Vyn 2006; Henderson & Gloy 2008; Shaik et al. 2005) - One or multiple categories of government support (Goodwin *et al.* 2003 & 2005; Pyykkonen 2005)
Variables describing the market	- Pig density (Duvivier et al. 2005) - Manure and farm density (Pyykkonen 2005) - Average farm size (Folland & Hough 1991) - Size of the agricultural land market (Duvivier *et al.* 2005) - Dummy for a specific region
Macroeconomic factors	- Property tax rate, interest rate (Weerahewa *et al.* 2008; Devadoss & Manchu 2007) - Inflation rate (Alston 1986) - Multifactor productivity growth (Gardner 2002) - Debt to asset ratio, Credit availability (Devadoss & Manchu 2007) - Unemployment rate (Pyykkonen 2005)
Urban pressure indicators	- Total population (Devadoss & Manchu 2007) - Population growth, rurality (Gardner 2002) - Ratio of population to farm acres, urbanisation categories (Goodwin *et al.* 2010) - Dummy variables for metropolitan areas (Henderson & Gloy 2008) - Proportion of the labour employed in agriculture (Pyykkonen 2005)