# Hedonic Estimation of Australian Boardacre Farmland Values

Khanh V. Hoang*, Will Chancellor & Shiji Zhao

Australian National University and Australian Bureau of Statistics

*EMG Workshop - Sydney*

November 30, 2018

# Motivation

- The valuation of farmland in Australia is often based on the purchase price and general expectations such as local market conditions
- But the value of the land is the product of many factors other than land size
- Exact longitude and latitude of each parcel of farmland now enable linking with other spatial data to produce more characteristics of farmland, for example:
  - distance to amenities including to the nearest railway station or port
  - use of land for different types of productive and non-productive purposes
  - regional economic condition characteristics
  - indicator of soil quality
- Hedonic methods are widely used to estimate residential real estate values; however, their use in valuing Australian farmland is much less common

# Our Data Set

Property sales data collected from the New South Wales (NSW) Land Registry Services for FY 2009/10 and FY 2010/11

Data variables:

- Sale price
- Contract date
- Land area in square metres
- Address of property
- Geographic longitude and latitude

ABARES has mapped over 50 additional variables to the Corelogic dataset using satellite imagery and location information

# Dealing with non-farm records

We have a total of 1768 transactions after removing non-farm records

- We define farmland as broadacre agricultural land with 50% land use for either grazing and / or cropping
- Non broadacre activities such as horticulture, mining, urban corridor etc. are out of scope
- Extreme outliers where the sale price per hectare is extremely low or extremely high are removed
- Exclude land sold for less than $50 per hectare. We suspect these are transactions between family members
- Any transactions where price exceeded $20,000 per hectare, and the land was less than 2 hectares were also excluded. This is to remove small and prestigious hobby farms and residential properties

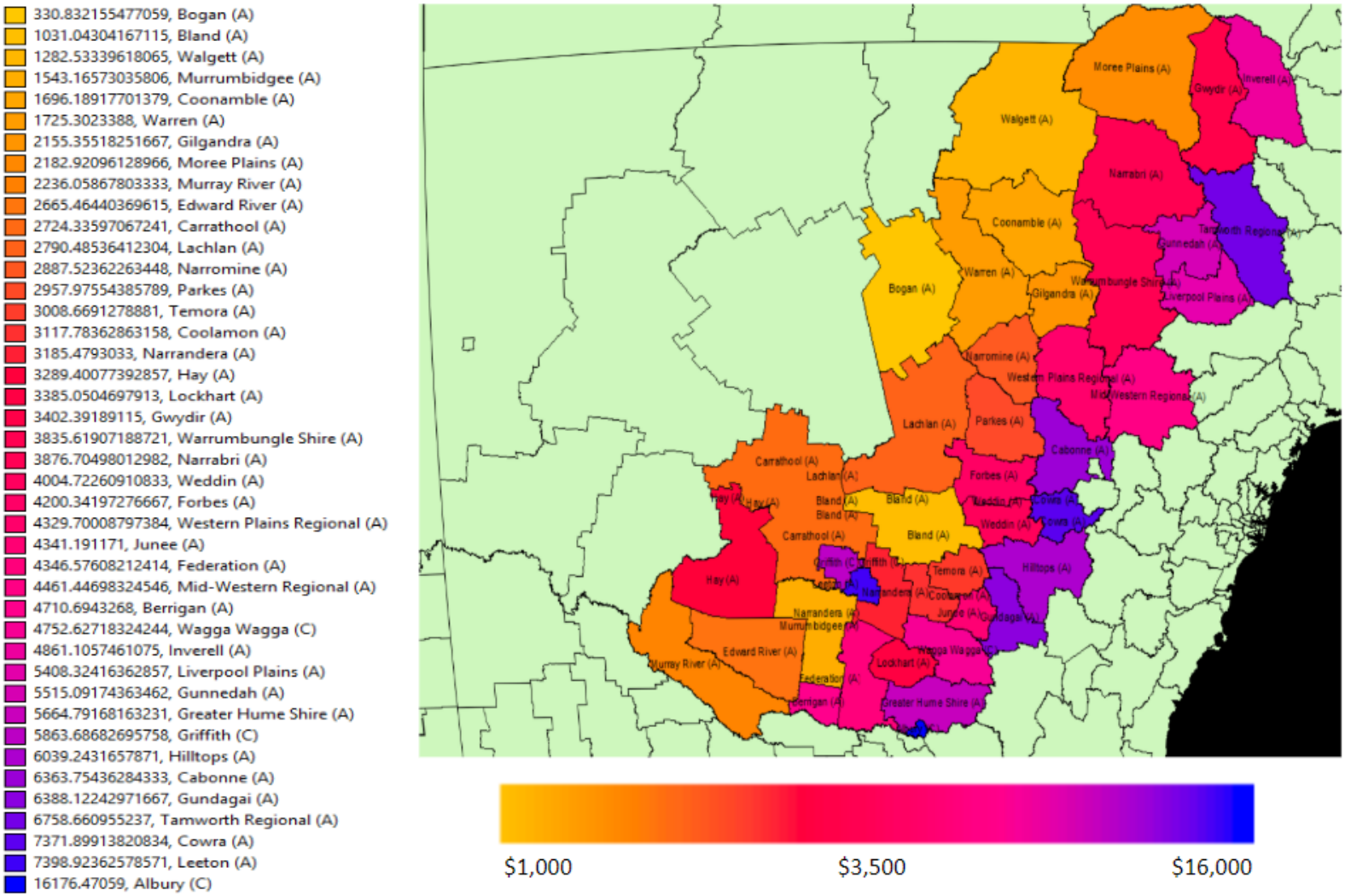



Figure 1: Average Price per Hectare by Local Government Area

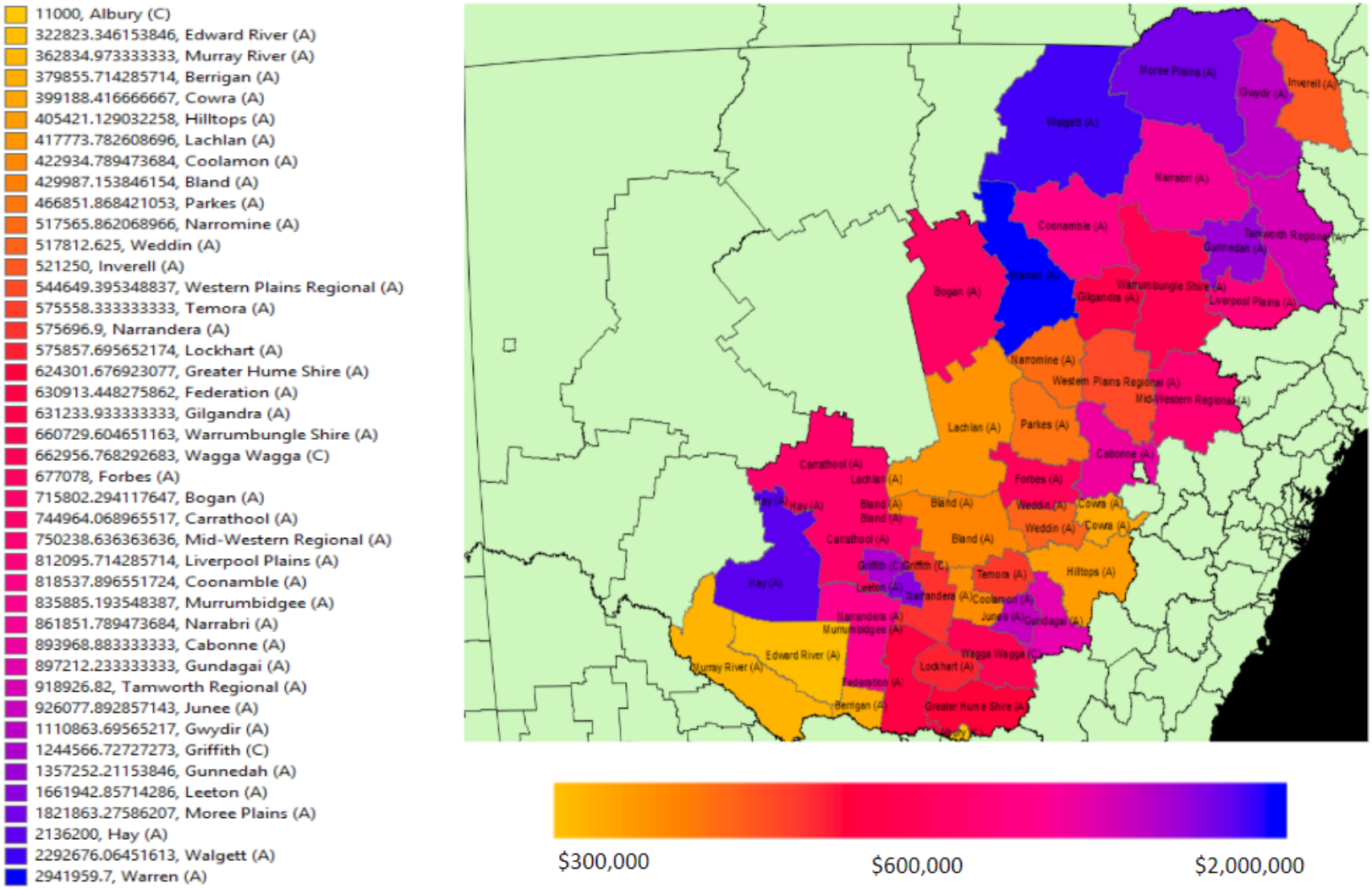



Figure 2: Price by Local Government Area

# Examples of Variables used to Explain Land Values

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

# Multicollinearity

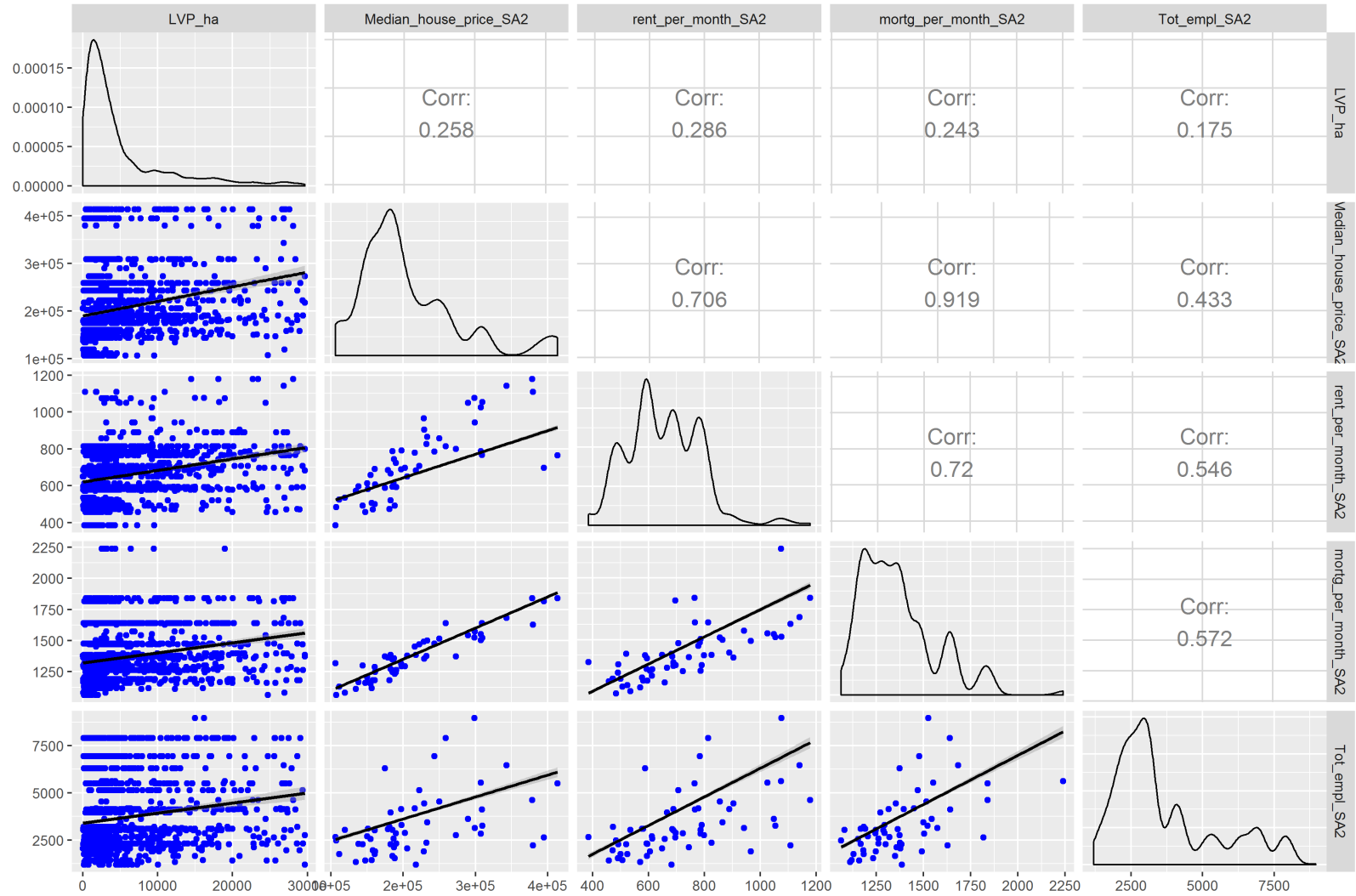



Figure 3: Correlation between Price Per Hectare and Regional Economic Conditions

# Our Data Set (cont)

The characteristics selected for the hedonic models are:

- **Local Government Area**
- **Geographic longitude and latitude**
- **Land Use** - *Modelled by integrating ABS agricultural commodity data, satellite imagery and other land use information. Categorised into land use for: farming; conservation; and non farming*
- **Distance to transport and amenities** - *The direct distance of the farm parcel to port by rail or road*
- **Average rainfall** - *The farm spatial coordinates and land area are matched with monthly rainfall and temperature data sourced from the Australian Water Availability Project*
- **Soil quality dummy indicator** - *A "1" indicate risk of dryland salinity by 2050. The spatial file is from the National Land and Water Resources Audit*
- **Erosion dummy indicator** - *Derived using spatial data from NSW Government Land transaction. A "1" indicate there was erosion in the polygon*
- **Population density** - *Total population divided by total area in square kilometres. Based on ABS Census data at the statistical area 1*
- **Number of agriculture businesses in region** - *Based on ABS Census data at the statistical area 2*
- **Total employment in region** - *Based on ABS Census data at the statistical area 2*

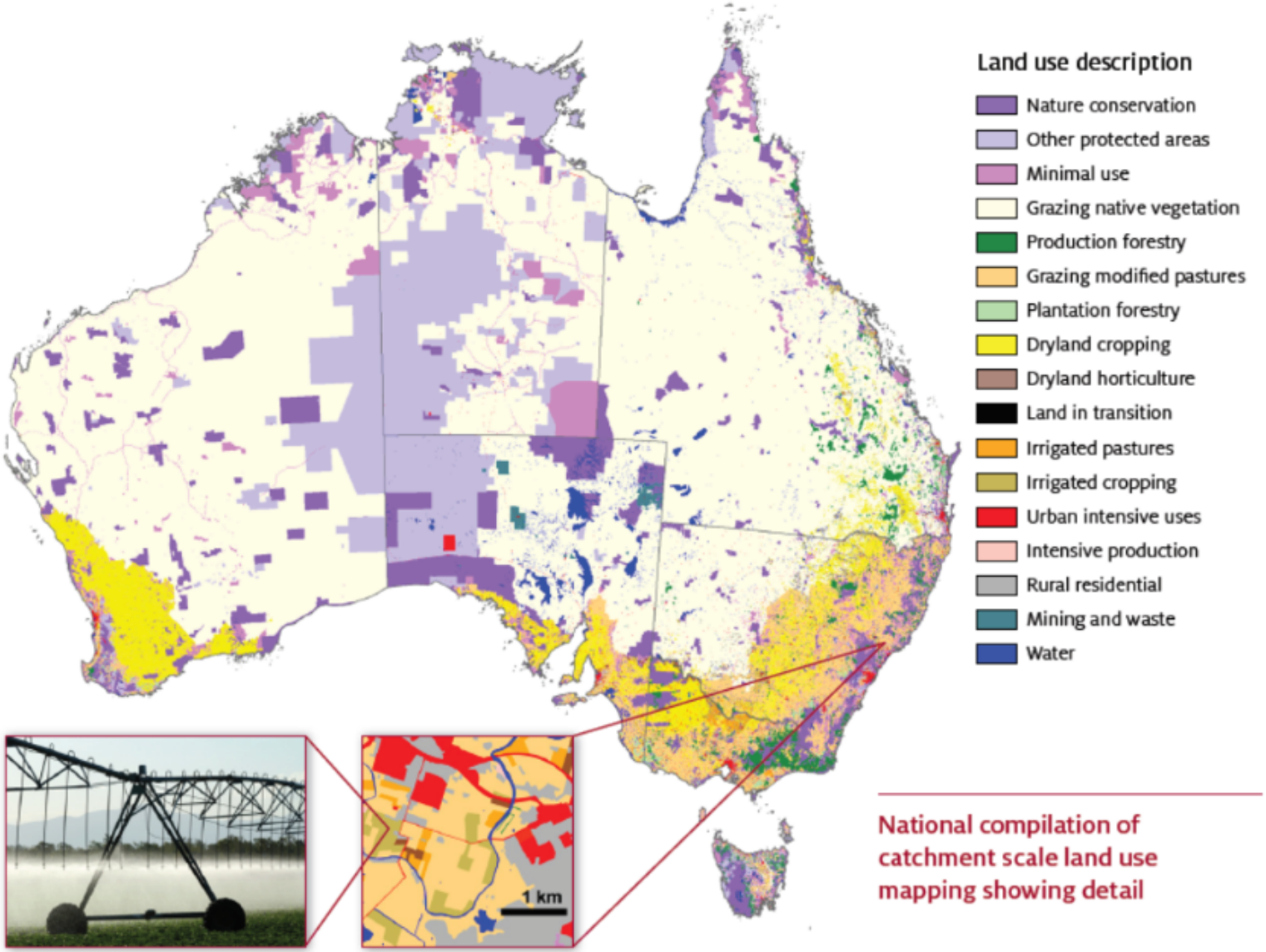



Figure 4: Land Use in Australia

# Hedonic Model Selection

- For continuous variables, probably the most common type of hedonic model is ordinary least squares (OLS) regression
- Hill and Scholz (2014) used a hedonic imputation approach with splines to produce house price indexes
- Standard hedonic models measures average movements in average locations belonging to average price segments
- Waltl (2016) used quantile regression to address the variation across price segments and location of the Sydney housing market

# Our Models

## Conditional Mean Functions

(i) semilog with regional dummies

$$E(y|X, D) = X\beta + D\delta$$

(ii) semilog with geospatial spline

$$E(y|X, lat, long) = X\beta + s(xlat, xlong)$$

- y is a H x 1 vector of log-price per hectare
- X is an H x B matrix of physical characteristics
- D is an H x C matrix of regional dummies
- The parameters to be estimated are:
  - B x 1 vector of characteristic shadow prices $\beta$
  - C x 1 vector of regional shadow prices $\delta$
  - s(xlat, xlong ), the geospatial spline function defined on the latitudes and longitudes

# The Spline Function

- s(xlat,xlong) functional form is not determined beforehand but driven by the data
- 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 Random Effects Maximum Likelihood (REML).
- The spline is estimated using the GAM function from the R package mgcv.

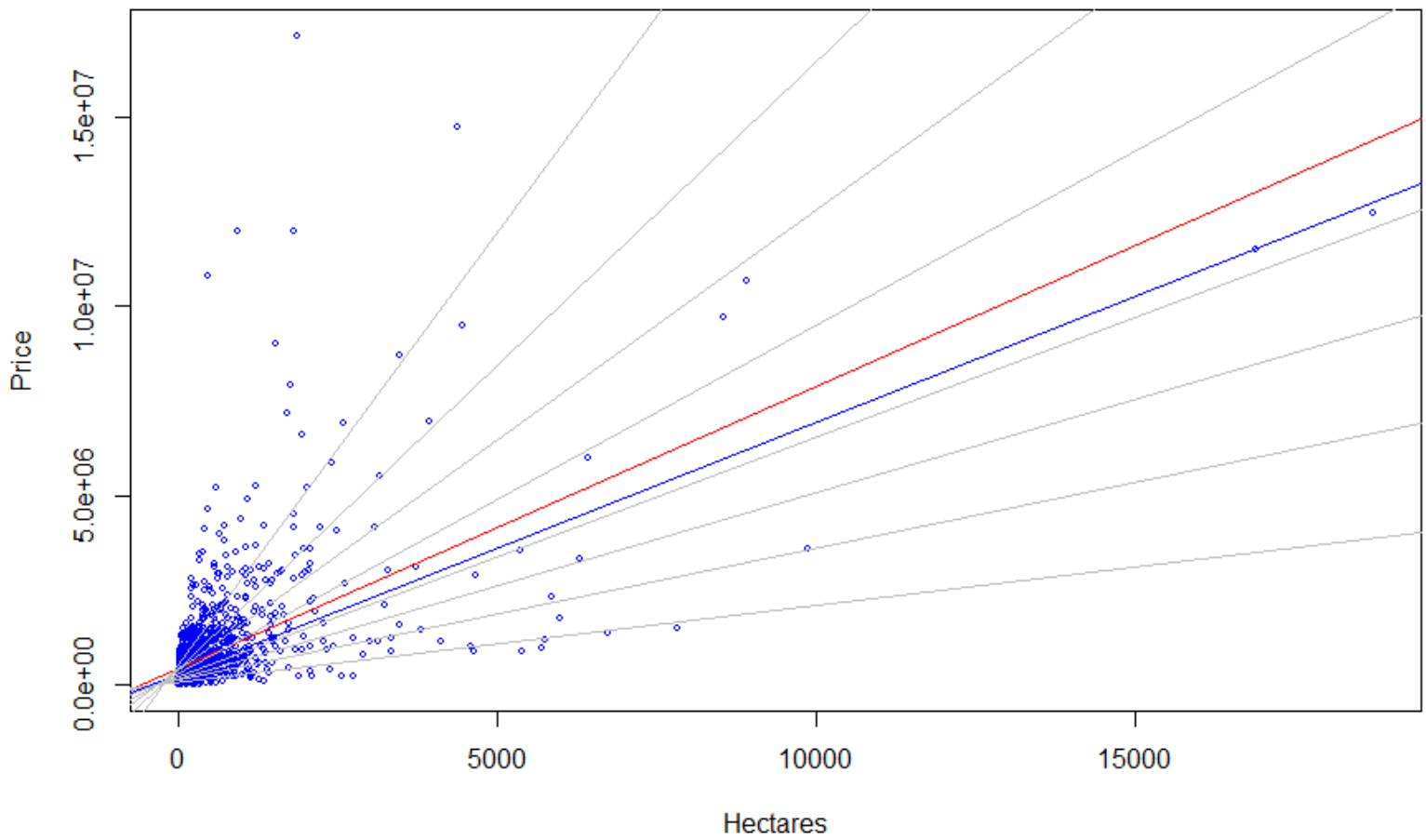



Figure 5: Plot of Price by Hectares

Red=OLS, Blue = $50^{th}$ Quantile, Grey = $10^{th}$ - $90^{th}$ Quantile

# Our Models (continued)

## Conditional Quantile functions - Koenker & Bassett (1978)

(iii) quantile regression

$$Q_y(\tau|X) = X\beta(\tau)$$

(iv) quantile regression with geospatial spline

$$Q_y(\tau|X) = X\beta(\tau) + g^{\tau}(xlat, xlong)$$

- $\tau \in (0,1)$ denotes a specific quantile level. Setting $\tau = 0.5$ yields an estimate for the conditional median
- $g^{\tau}$'s function form is not determined beforehand. It has been estimated using a Schwartz Information Criterion penalization approach and the triogram method developed by Hansen *et al.* (1998).
- The spline is estimated using the function rqss from the R package quantreg (Koenker 2013)

# Table 1: Results for Semilog models

| | *Dependent variable:* | |
|---|---|---|
| | Log-Price Per Hectare | |
| | No Regional Dummies | Regional Dummies |
| Farm | $0.18^{***}$ (0.06) | $0.26^{***}$ (0.06) |
| Residence | $0.83^{***}$ (0.08) | $0.82^{***}$ (0.07) |
| Land use for farming | $-4.24^{***}$ (0.60) | $-3.59^{***}$ (0.57) |
| Land use for conservation | $-6.29^{***}$ (0.64) | $-5.20^{***}$ (0.62) |
| No of agriculture businesses | $-0.001^{***}$ (0.000) | $-0.001^{***}$ (0.000) |
| Soil quality indicator | $0.34^{***}$ (0.07) | $0.21^{***}$ (0.07) |
| Population density | $0.01^{***}$ (0.001) | $0.005^{***}$ (0.001) |
| Employment in Region | $0.000^{***}$ (0.000) | $0.000^{***}$ (0.000) |
| Erosion indicator | $-0.49^{***}$ (0.06) | $-0.50^{***}$ (0.05) |
| Average rainfall | $0.03^{***}$ (0.003) | $0.02^{***}$ (0.01) |
| Distance to port by rail/road | $-0.002^{***}$ (0.000) | $-0.01^{***}$ (0.001) |
| LGA dummies* | | *** |
| Constant | $11.22^{***}$ (0.62) | $13.53^{***}$ (1.15) |
| Observations | 1,794 | 1,794 |
| $R^2$ | 0.33 | 0.45 |
| Adjusted $R^2$ | 0.32 | 0.43 |
| Residual Std. Error | 0.97 (df = 1781) | 0.88 (df = 1740) |
| F Statistic | $71.85^{***}$ (df = 12; 1781) | $26.99^{***}$ (df = 53; 1740) |
| *Note:* | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

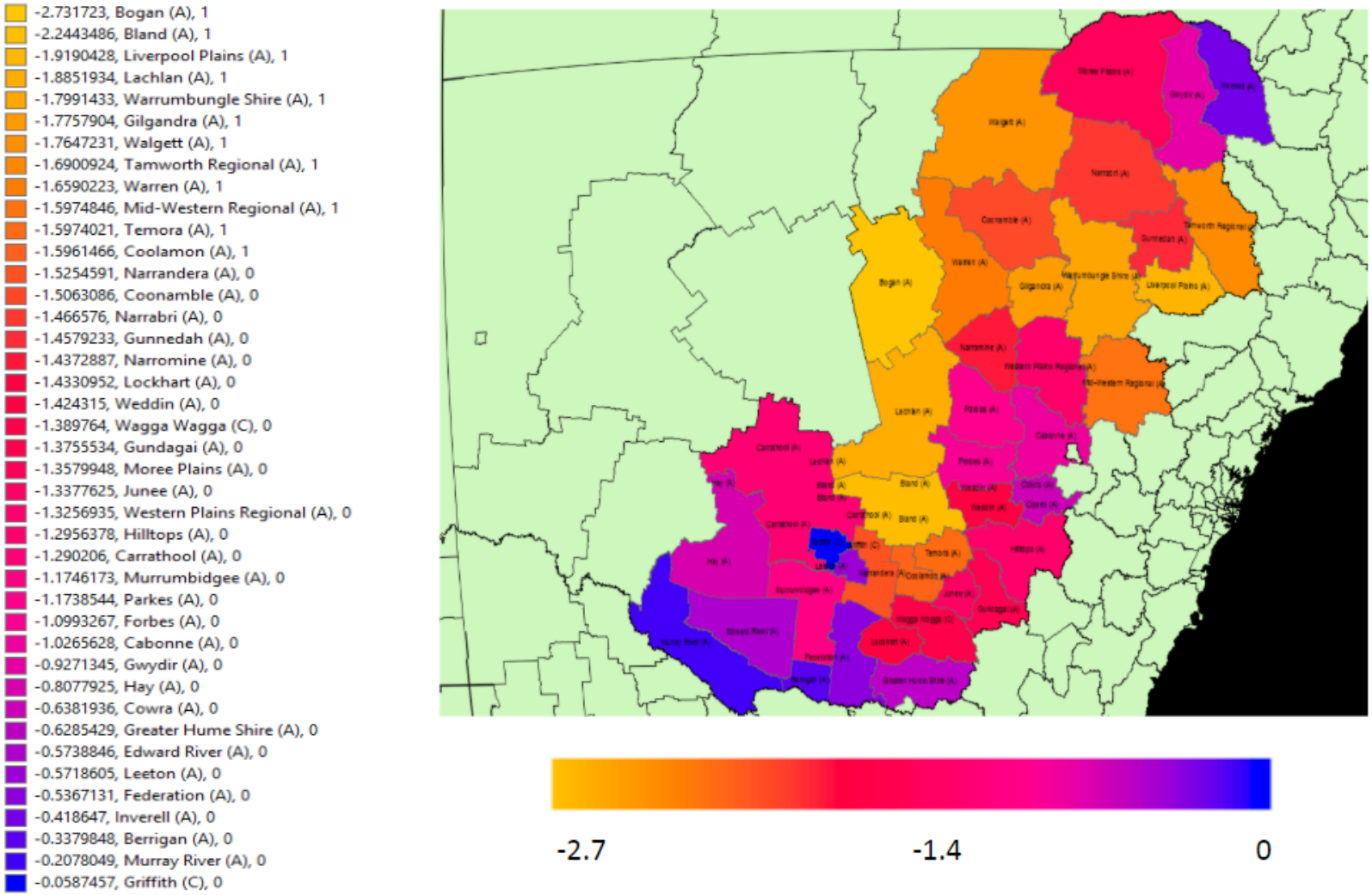
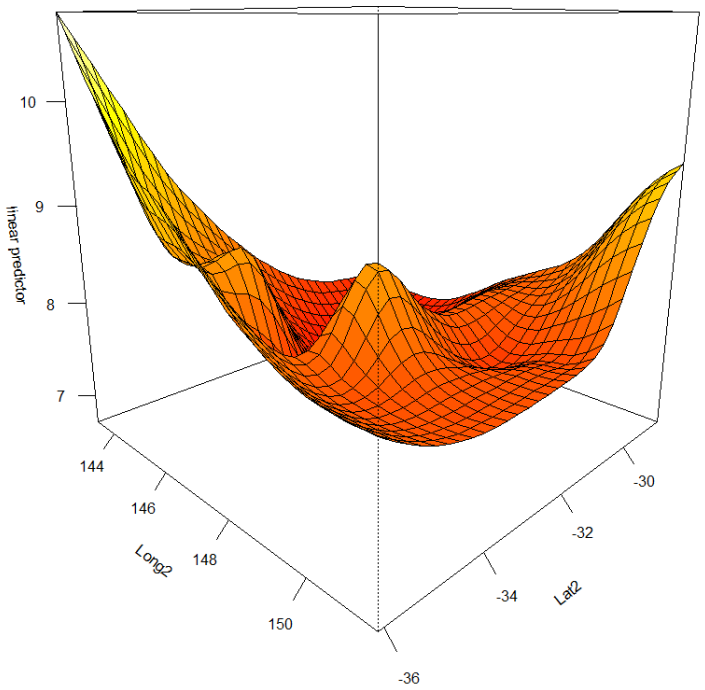



Figure 6: Coefficient by Local Government Area

## Table 2: Results Semilong with geospatial spline

| | *Dependent variable:* |
|---|---|
| | Log - Price per hectare |
| Farm | $0.27^{***}$ (0.06) |
| Residence | $0.85^{***}$ (0.07) |
| Land use for farming | $-3.79^{***}$ (0.57) |
| Land use for conservation | $-5.47^{***}$ (0.61) |
| No of agricultural businesses | $-0.001^{***}$ (0.000) |
| Soil quality indicator | $0.21^{***}$ (0.07) |
| Population density within 2km | $0.01^{***}$ (0.001) |
| Total employment | $0.000^{***}$ (0.000) |
| Erosion indicator | $-0.52^{***}$ (0.06) |
| Average rainfall | $0.02^{***}$ (0.01) |
| Distance from port by rail/road | $-0.01^{***}$ (0.001) |
| Constant | $12.67^{***}$ (0.70) |
| | Approx. significant of smooth terms |
| s(Latitude,Longitude) | $24.22^{***}$ (0.06) |
| Observations | 1,794 |
| Adjusted $R^2$ | 0.43 |
| Log Likelihood | $-2,355.30$ |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |



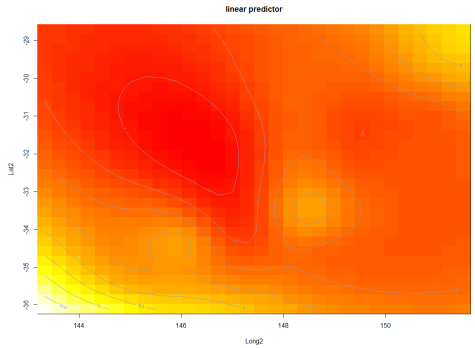




Figure 7: Coefficient of Longitude and Latitude

## Table 3: Results for Quantile Regression

| | *Dependent variable:log-price per hectare* | | | | | |
|---|---|---|---|---|---|---|
| | QR | | | QR with Lat,Long Spline | | |
| | $20^{th}$ | $50^{th}$ | $80^{th}$ | $20^{th}$ | $50^{th}$ | $80^{th}$ |
| Farm | $0.33^{***}$ | $0.07$ | $-0.02$ | $0.37$ | $0.14$ | $-0.03$ |
| Residence | $0.81^{***}$ | $0.83^{***}$ | $0.78^{***}$ | $0.82^{***}$ | $0.93^{***}$ | $0.79^{***}$ |
| Land use - farming | $-3.92^{***}$ | $-5.17^{***}$ | $-5.43^{***}$ | $-2.72^{***}$ | $-3.98^{*}$ | $-5.05^{***}$ |
| Land use - conservation | $-6.79^{***}$ | $-7.56^{***}$ | $-6.98^{***}$ | $-8.49^{***}$ | $-5.99^{***}$ | $-6.61^{***}$ |
| No of ag. businesses | $-0.001^{***}$ | $-0.001^{***}$ | $-0.0^{**}$ | $-0.001^{***}$ | $-0.00^{*}$ | $-0.0^{*}$ |
| Soil quality indicator | $0.42^{***}$ | $0.30^{***}$ | $0.31^{***}$ | $0.24^{***}$ | $0.13$ | $0.17^{*}$ |
| Pop. density | $0.004^{***}$ | $0.01^{**}$ | $0.01^{***}$ | $0.004^{***}$ | $0.005$ | $0.01$ |
| Total employment | $0.00^{***}$ | $0.00^{***}$ | $0.00^{***}$ | $0.00^{***}$ | $0.00^{**}$ | $0.00^{**}$ |
| Erosion indicator | $-0.32^{***}$ | $-0.32^{***}$ | $-0.52^{***}$ | $-0.47^{***}$ | $-0.38^{***}$ | $-0.48^{***}$ |
| Average rainfall | $0.02^{***}$ | $0.03^{***}$ | $0.03^{***}$ | $0.00$ | $0.009$ | $0.01$ |
| Distance - port to rail | $-0.004^{***}$ | $-0.002^{***}$ | $-0.002^{***}$ | $-0.004^{***}$ | $-0.004^{***}$ | $-0.004^{*}$ |
| Constant | $10.76^{***}$ | $11.96^{***}$ | $13.03^{***}$ | $10.37^{***}$ | $12.48^{***}$ | $14.00^{***}$ |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

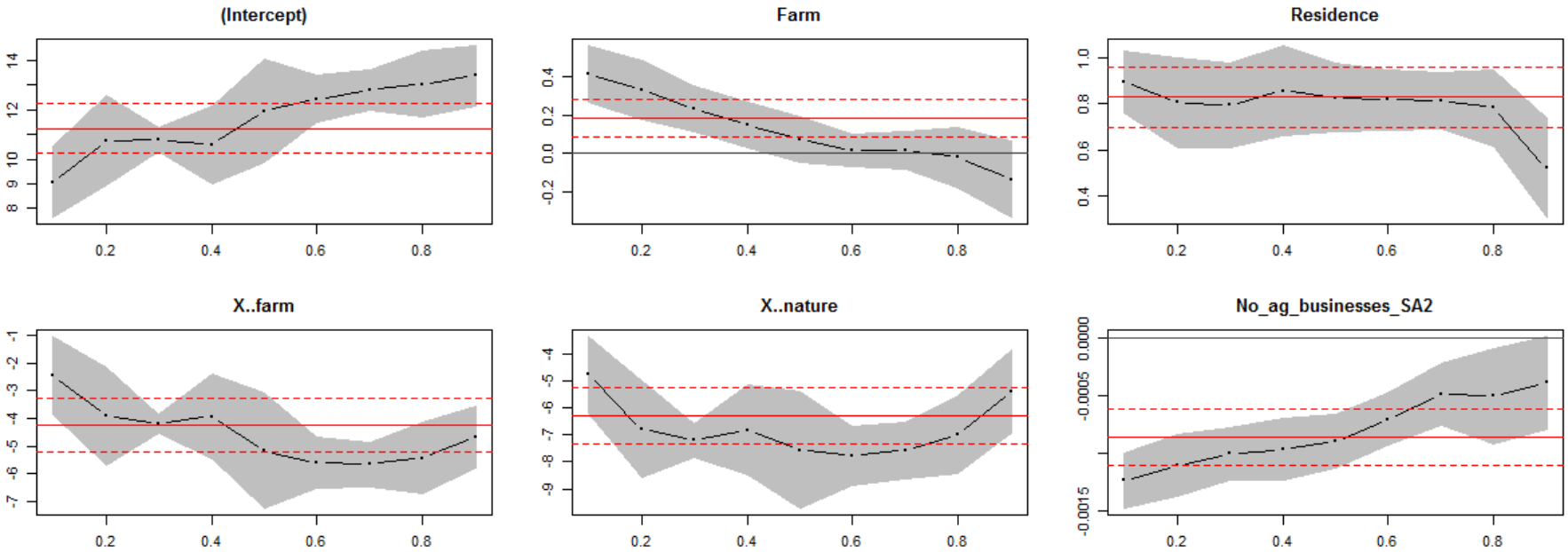



Figure 8: Quantile Regression at each $10^{th}$ percentile for Log - Price per Hectare

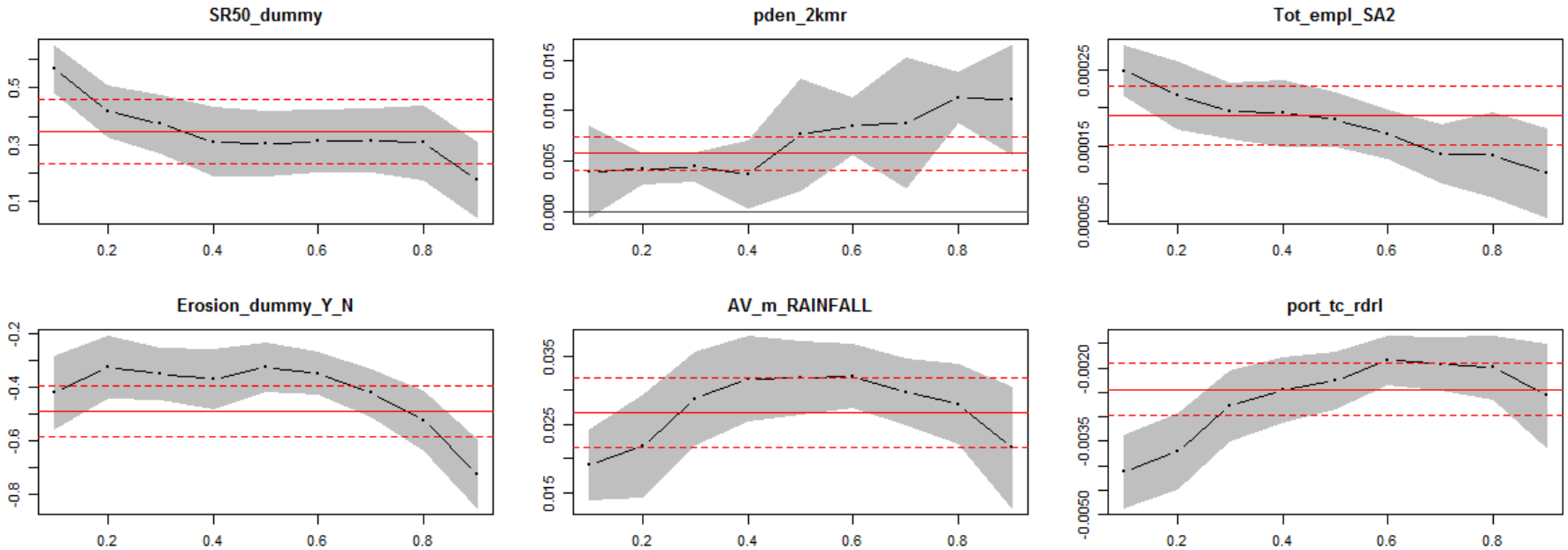



Figure 9: Quantile Regression at each $10^{th}$ percentile for Log - Price per Hectare

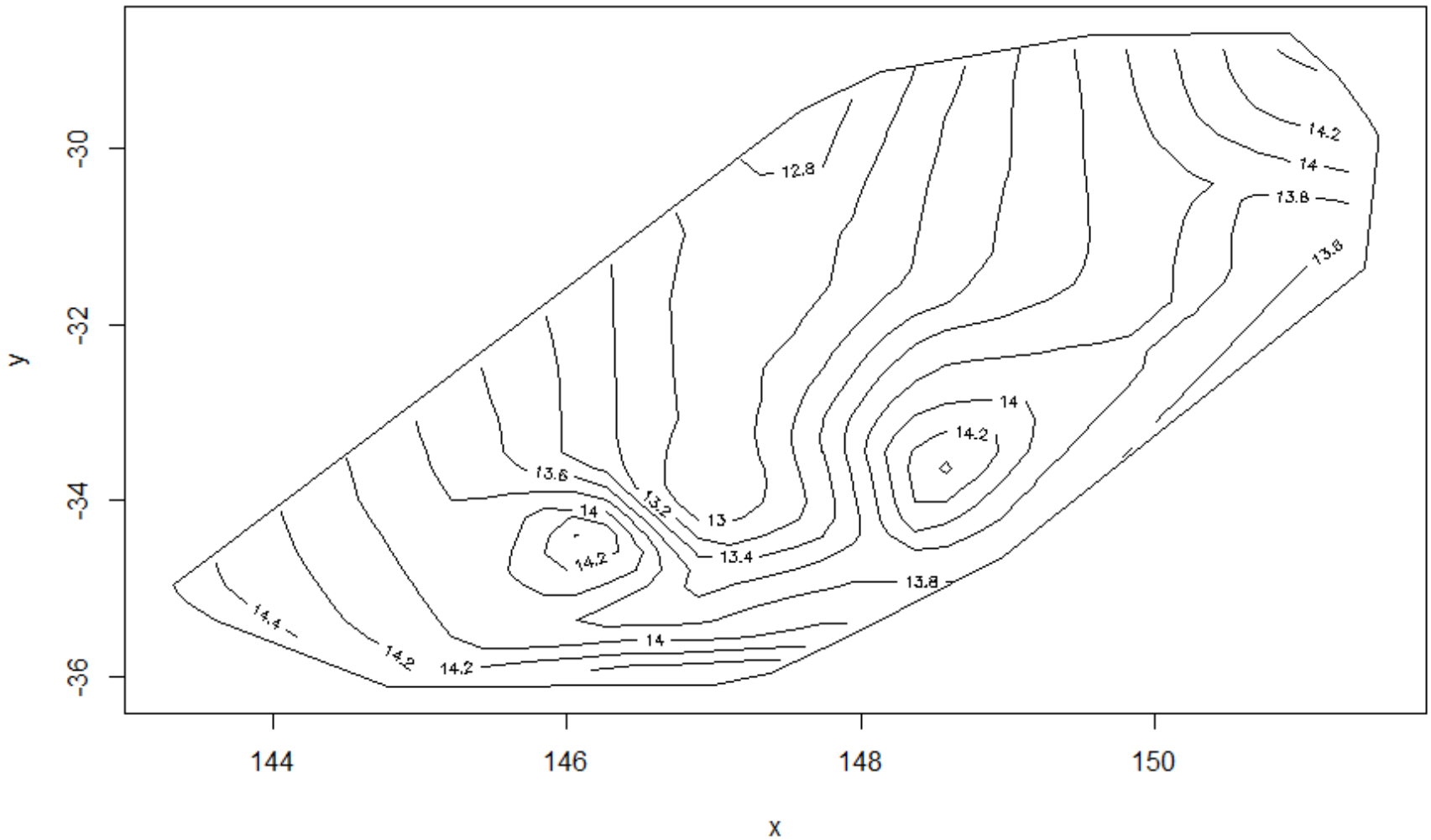



Figure 10: Triogram of Longitude and Latitude

# Comparing the Performance of our Models

Table 4: Semilog Models

| | Akaike Info. Criterion | Bayesian Info. Criterion | Gen. Cross Validation |
|---|---|---|---|
| Semilog no regional dummies | 4982.4 | 5053.8 | 2525.1 |
| Semilog with regional dummies | 4703.8 | 5000.4 | 2380.2 |
| Semilog with Lat,Long spline | 4711.9 | 4922.4 | 2415.4 |

- Including location in the model (whether regional or lat, long spline) is important
- Interestingly, the spline model marginally outperforms its regional dummies counterpart based on the BIC but not based on AIC or GCV

# Conclusion and future work

- Splines (or some other non-parametric method) provide a flexible way of incorporating geospatial data into a hedonic model
- Applying these approaches to a sample of agricultural land sales in NSW we find that there are variations in the value of farmland across location and price segments
- The quantile regression approach with the geospatial spline is better suited in case of a large number of regional entities. It also achieved lower standard errors in this analysis
- In sum, this leads to a preference of using splines over the regional dummies models
- We recently gained access to a much larger longitudinal transaction data series, spanning 20 years and covering all of Australia
- Explore other hedonic models to account for variation in the price of farmland across location, time and price segments

# Thank You