



Australian Government
**Department of Agriculture
and Water Resources**
ABARES

farmpredict: A micro-simulation model of Australian farms

Neal Hughes, Wei Ying Soh, Chris Boulton, Kenton Lawson,
Manann Donoghoe, Haydn Valle and Will Chancellor¹

ABARES Working Paper
September 2019

¹All authors are in The Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES), Canberra, Australian Capital Territory. The views expressed in this article are the author's own and do not necessarily reflect the views of ABARES.

© Commonwealth of Australia 2019

Ownership of intellectual property rights

Unless otherwise noted, copyright (and any other intellectual property rights, if any) in this publication is owned by the Commonwealth of Australia (referred to as the Commonwealth).

Creative Commons licence

All material in this publication is licensed under a Creative Commons Attribution 3.0 Australia Licence, save for content supplied by third parties, logos and the Commonwealth Coat of Arms.



Creative Commons Attribution 3.0 Australia Licence is a standard form licence agreement that allows you to copy, distribute, transmit and adapt this publication provided you attribute the work. See the [summary of the licence terms](#) or the [full licence terms](#).

Inquiries about the licence and any use of this document should be emailed to copyright@agriculture.gov.au.

Cataloguing data

Hughes, N., Soh, W., Lawson, K., Boulton, C., Donoghoe, M., Valle, H. and Chancellor, W. 2019, farmpredict: A micro-simulation model of Australian farms, ABARES working paper, Canberra, November

Internet

This publication is available at <http://agriculture.gov.au/abares/research-topics/working-papers>.

Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES)

Postal address GPO Box 858 Canberra ACT 2601

Switchboard +61 2 6272 2010

Email info.abares@agriculture.gov.au

Web agriculture.gov.au/abares

Abstract

This paper presents *farmpredict* a data driven micro-simulation model of Australian broadacre farms based on the Australian Agricultural and Grazing Industry Survey (AAGIS). Farm production and financial data from AAGIS are combined with site-specific climate data including various measures of rainfall, temperature and soil moisture. A statistical model is estimated linking the production of outputs (e.g., wheat, beef cattle, wool etc.), the use of inputs (e.g., fuel, fertiliser, labour etc.) and changes in farm stocks (e.g., livestock and grain) with farm fixed inputs, input and output prices, climate variables and other control variables. The model is estimated using a non-parametric machine learning method, which combines a gradient boosted regression tree algorithm (*xgboost*) with multi-target stacking (two-stage regression). The resulting model can be used to forecast or simulate production, financial outcomes and stock changes for individual farms given scenarios for climate conditions and commodity prices. The performance of the model is evaluated via cross-validation. Simulation results are presented showing the types of farm responses to climate and price shocks produced by the model.

1 Introduction

Farm businesses face a high level of risk due mostly to variations in weather conditions and commodity markets. Both sources of risk are particularly high in Australia (Keogh 2012) given our export focused and unregulated commodity markets and our variable climate---with lower mean rainfall and higher variance than most other countries (Peel et al. 2004).

The effects of climate and price variation on farms are complex and farm specific. Models which can accurately measure these effects have a range of potential applications including forecasting farm production and profit (Nelson et al. 2007), pricing farm insurance products (Adeyinka et al. 2016) and supporting government drought policy programs (Nelson et al. 2007). Such models are also necessary to measure the effects of climate change on farms. This is important in Australia where trends in rainfall and temperature related to climate change (Cai & Cowan 2013) are already having significant effects on crop yields (Hochman et al. 2017) and farm productivity (Hughes et al. 2016).

This paper presents *farmpredict* a new data-driven micro-simulation model of Australian farms. *farmpredict* is based on data from the [Australian Agricultural and Grazing Industry Survey](#) (AAGIS) a long running national survey of Australian broadacre farms covering the major cropping, livestock (beef and sheep) and mixed farming industries. A sample of 40,733 observations over the period 1988-89 to 2017-18 are drawn from AAGIS and combined with a variety of location specific climate data including measures of rainfall, temperature and soil moisture.

The model comprises a series of input demand and output supply functions which predict the production of outputs, the use of inputs and changes in stocks conditional on prices, fixed inputs, climate conditions and other control variables. In this sense, the model structure is similar to 'dual' (reduced form) style production models common in econometrics (Mundlak 2001)--- which emerged as an alternative to 'primal' models due to concerns over endogeneity bias (Marschak & Andrews 1944).

This dual approach has been applied to estimate farm production systems in many Australian (Xayavong et al. 2011, Fisher & Wall 1990, Bell et al. 2007) and international (Segerson & Dixon 1999, Ball et al. 1997, Bouchet et al. 1989, Antle 1984) studies. These studies draw on economic profit maximisation theory to develop parametric models which are estimated via standard econometric methods.

The dual approach suffers from some widely acknowledged limitations (Mundlak 2001). Firstly, the parametric forms can be highly restrictive. For example, the common normalised quadratic profit function (Shumway 1983) assumes linear responses, such that all farms experience the same constant marginal changes in output or input. Secondly, for a variety of reasons (including data aggregation, risk and uncertainty, non-profit objectives and measurement error) observed data typically do not satisfy profit maximisation (see Mundlak 2001, Shumway 1995, Xayavong et al. 2011).

In this study, a machine learning approach to estimation is adopted (for an introduction to machine learning see Varian 2014, Einav & Levin 2014). Each function in the model is estimated via a gradient boosted regression tree algorithm, specifically *xgboost* (Friedman 2002, Chen & Guestrin 2016). A multi-target stacking (Spyromitros-Xioufis et al. 2012) framework is employed to account for interactions between functions, not dissimilar to two-stage least squares approaches in econometrics.

This non-parametric approach exploits the large sample sizes available and helps address some of the limitations of standard econometric methods. However, the model still draws on economic theory. In particular, only variables which can safely be assumed exogenous (with one-directional casual effects) are included as predictors: weather conditions, opening values of fixed inputs and prices (based on the standard assumption that farmers are price takers in competitive markets). These assumptions help make the model suitable not just for forecasting but for counterfactual simulation.

In terms of its structure and potential applications *farmpredict* is closely related to previous farm 'econometric process models' (Antle & Capalbo 2001), which combine spatially explicitly econometric models with bio-physical simulation models.

In particular, *farmpredict* is a successor to the Agricultural Farm Income Risk Model (AgFIRM) developed from AAGIS data by Kokic et al. (2007). AgFIRM combined a primal (structural) econometric farm production model (originally developed by Kokic et al. 1993) with two bio-physical models: the crop yield model of Potgieter et al. (2002) and the pasture growth model of Carter et al. (2000). Kokic et al. (2007) demonstrated how AgFIRM could be used to forecast farm incomes. Later Nelson et al. (2010) applied AgFIRM to project farm incomes under climate change scenarios.

In the United States, Antle & Capalbo (2001) introduced the concept of econometric process models via a dual form model of Montana grain farms. Later Stoorvogel et al. (2004) developed a similar model to examine trade-offs between agricultural and environmental outcomes for farming areas in northern Ecuador.

This study differs from the above models, in linking farm production with observed climate data rather than bio-physical models, to create a statistical model of the joint farm economic and bio-physical system. Machine learning feature selection methods are used to select a subset of climate variables for each equation in the model from a large combination of potential climate measures (e.g., rainfall, temperature, soil moisture etc.) and time scales / periods.

This data-driven approach has some obvious strengths and weaknesses: while it avoids the potential bias from using process based models (which are often calibrated for specific locations, crop / pasture types and farming systems) it also limits the ability of the model to extrapolate outside of historically observed ranges. While this means process based models might be better suited to climate change analysis (Antle & Capalbo 2001), statistical models can still be, and often are, usefully applied for this purpose (Lobell & Burke 2010). Further, the predictive accuracy achievable with machine learning opens a range of new practical applications including farm specific risk analysis of value in credit and insurance markets.

2 Methods

Data

Farm data

The [Australian Agricultural and Grazing Industry Survey](#) (AAGIS) collects detailed physical and financial information from around 1,600 broadacre (extensive non-irrigated cropping / livestock) farms across Australia each year. AAGIS data are all collected through face-to-face interviews with the owner or manager of the farm.

AAGIS includes farms across five industry categories (as defined by [ANZSIC 2006](#)): *Cropping specialists*, *Mixed cropping-livestock*, *Beef*, *Sheep* and *Sheep-Beef*. Within these broad categories, farms can produce a range of different crop and livestock outputs. For example, many cropping specialists farms still hold livestock, while many sheep (and some beef) farms also plant crops (see Table 2).

Results from the survey are presented as national, state or region level weighted averages (for 34 regions as detailed in Appendix A). The AAGIS farm level data are geocoded (with spatial coordinates identifying the approximate centre of the farm property). The spatial distribution of farms by industry is shown in Figure 1. Cropping activity occurs mostly within the 'Wheat-Sheep zone', with livestock production dominating in the coastal 'High-rainfall' zones (where rainfall is often too high for cropping) and the more in-land 'Pastoral' zones (where rainfall is often too low for cropping).

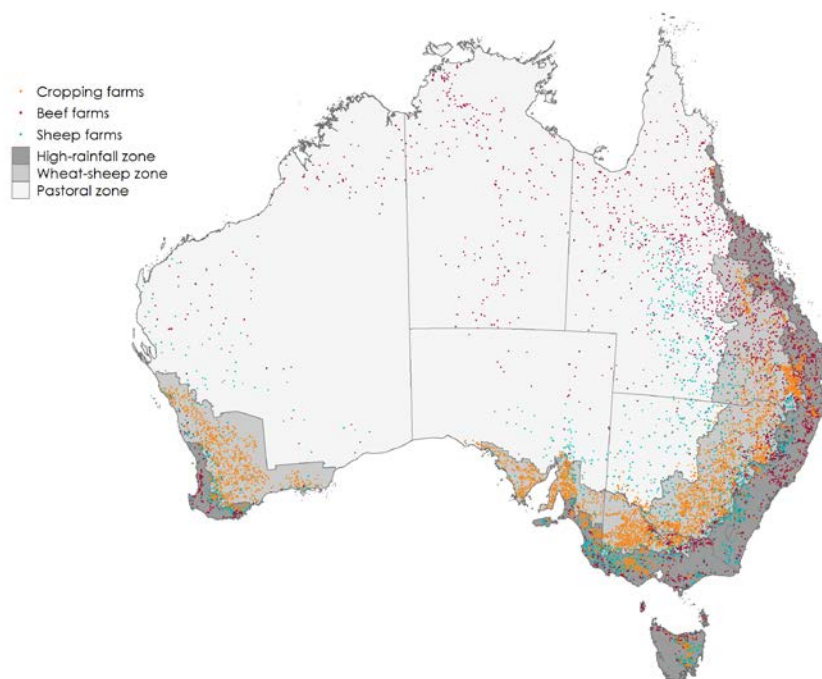


Figure 1: Spatial distribution of farms in the AAGIS sample by industry, 2004-05 to 2014-15

Note: Sheep farms include *Sheep* and *Sheep-Beef* industries, Cropping includes *Cropping specialists* and *Mixed cropping-livestock*.

While AAGIS data are available from 1977-78, the sample for this study is limited to the period 1988-89 to 2017-18 due to missing data in earlier years. AAGIS involves a rotating sample resulting in an unbalanced panel data set. Between 1988-89 and 2017-18 the data includes a

total of 43,959 observations for 13,117 distinct farm businesses, with farms spending an average of 3.4 years in the sample.

The complete set of variables are listed in Table 11, Table 12 and Table 14 in Appendix B. Key variables include:

Q_{jit}	the quantity of output j sold on farm i in year t
R_{jit}	the revenue received from output j on farm i in year t
A_{jit}	the area of crop j planted on farm i in year t
H_{jit}	the quantity of crop j harvested on farm i in year t
C_{vit}	expenditure on variable input v on farm i in year t
V_{vit}	a quantity index for variable input v ($\frac{C_{vit}}{P_{vt}}$) on farm i in year t
K_{kit}	the quantity of capital input k on farm i in year t
$S_{sit}^{op}, S_{sit}^{cl}$	the opening and closing quantities of stock s on farm i in year t
\mathbf{P}_t	annual price indexes for each output and input
\mathbf{Z}_{it}	a collection of control variables, including location and year

A total of 10 outputs and 8 variable inputs are defined for the model (Table 1), similar to the setup used by ABARES for its Total Factor Productivity (TFP) series (Zhao et al. 2012).

Table 1: Output, variable input and stock types

Outputs	Variable inputs	Stocks1
Beef cattle	Electricity	Beef cattle
Sheep	Fertiliser	Sheep
Lamb	Fuel	Wool
Wool	Chemicals	Wheat
Wheat	Other materials	Barley
Barley	Services	Sorghum
Oilseeds	Shearing labour	Oilseeds
Sorghum	Other	
Other crops		
Other		

For livestock outputs (beef cattle, sheep, and wool) price indexes are based on annual national median farm prices, for crops (wheat, barley, sorghum and oilseeds) they are based on Australian export prices (all other outputs and inputs use the price indexes of Zhao et al. 2012). For crop outputs both current and lag prices are constructed. For winter crops in particular (wheat, barley and oilseeds) lagged prices are a better indicator for expected prices, given crop planting decisions occur prior to the beginning of the survey year (see Figure 2). Price indexes and financial variables, such as R_{jit} and C_{vit} , are all normalised (adjusted for inflation) using the Consumer Price Index (CPI).

There are few missing values in the dataset post 1988-89, however some imputation is required for control variables Z_{edu} and Z_{age} (farmer education and age) and for crop pool payment

variables (which account for participation in centralised crop marketing schemes, see Appendix B and C). Lagged or next year values are used for imputation where available (for farms in the sample multiple years) and annual median values otherwise. Some livestock data is missing in 2004 (for beef and sheep sales, births and deaths) and is not imputed. After removing outliers (observations where one or more key variables exceed the 99.5th percentile) and observations with large changes in farm scale (differences between opening and closing land area), the final sample size for estimation is 40,202. Selected summary statistics are shown in Table 2.

Table 2: Sample percentiles for selected AAGIS variables by industry ('000 \$)

	0.05	0.25	0.5	0.75	0.95
Cropping and mixed farms					
Beef revenue, R_{beef}	0.0	0.0	0.0	34.9	315.4
Wheat revenue, R_{wheat}	0.0	22.2	126.0	343.7	986.5
Total revenue, R_{total}	95.6	282.8	561.4	1006.2	2129.8
Total costs, C_{total}	82.7	214.6	414.8	739.7	1638.7
Farm cash income, $R_{total} - C_{total}$	-129.7	24.2	120.2	288.5	741.0
Farm business profit	-318.2	-81.9	3.5	142.3	560.4
Beef farms					
Beef revenue, R_{beef}	30.7	114.1	316.5	838.4	3161.4
Wheat revenue, R_{wheat}	0.0	0.0	0.0	0.0	0.0
Total revenue, R_{total}	43.6	147.6	398.5	973.3	3330.8
Total costs, $*C_{total}$	43.6	121.8	304.5	756.3	2334.3
Farm cash income, $R_{total} - C_{total}$	-252.6	-7.0	63.9	236.5	1113.5
Farm business profit	-514.7	-94.1	-18.0	123.0	903.1
All farms					
Beef revenue, R_{beef}	0.0	0.0	42.7	217.8	1331.4
Wheat revenue, R_{wheat}	0.0	0.0	0.0	92.8	623.6
Total revenue, R_{total}	61.0	201.6	444.5	888.3	2227.7
Total costs, C_{total}	56.7	159.9	337.0	670.1	1710.1
Farm cash income, $R_{total} - C_{total}$	-154.6	7.8	84.6	236.9	745.0
Farm business profit	-340.6	-84.4	-11.2	111.7	572.6

Climate data

Climate data are obtained from a number of sources. Monthly rainfall and temperature data are sourced from the [Australian Water Availability Project \(AWAP\)](#) (Raupach et al. 2009). Soil moisture data are obtained from the Bureau of Meteorology (BoM) [Australian Water Resources Assessment Landscape model \(AWRA-L\)](#) (Frost et al. 2016). Daily rainfall and temperature data for Australian weather stations are obtained from the [Scientific Information for Land Owners \(SILO\) database](#) (Jeffrey et al. 2001). Data on hail storms were obtained from the BoM [Severe Storms Archive](#).

Both AWAP and AWRA-L provide spatial data for Australia on a 0.05 degree (around 5km) grid. Monthly data for rainfall, average maximum and minimum temperatures and soil moisture were matched to each farm using the spatial co-ordinates and land area. Daily rainfall and temperature data were matched to farms on the basis of nearest weather station.

A variety of variables are constructed from these sources, for a combination of different climate measures (Table 3) and time periods / seasons (Figure 2) of potential relevance to Australian cropping and livestock production (see Appendix B for more detail). Climate variables include measures of both total rainfall and rainfall volatility (the spread of daily rainfall totals across a period). Various measures of exposure to upper and lower temperature extremes are also included.

Table 3: Climate variable measures

Name	Description	Units	Source
<i>rain</i>	Rainfall volume	mm	AWAP
<i>tmax</i>	Average maximum temperature	degrees C	AWAP
<i>tmin</i>	Average minimum temperature	degrees C	AWAP
<i>moist</i>	Root zone soil moisture	index (0-1)	AWRA-L
<i>fr2</i>	Exposure to frost (days below 2C)	days	SILO
<i>gdd</i>	Heat accumulation (growing degree days)	degrees C	SILO
<i>hgdd</i>	Exposure to high temperatures	degrees C	SILO
<i>gni</i>	Rainfall volatility (Gini coefficient)	index	SILO
<i>pci</i>	Rainfall volatility	index	SILO
<i>hail</i>	Exposure to hail storms	index (0-1)	BoM

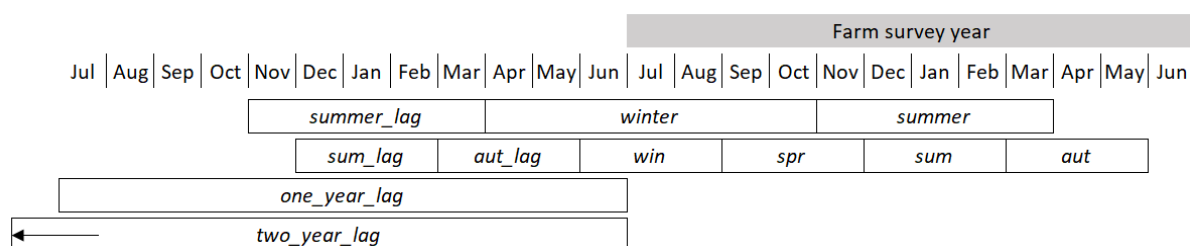


Figure 2: Climate variable time periods

The time periods (Figure 2) are defined with relevance to the survey (financial) year. Note that the *winter* cropping season begins prior to the start of the survey year. For example 2016-17 winter crop production data, refers to the winter crops planted in the autumn of 2016. Lagged climate variables (up to two years prior to the survey year) are relevant given their potential effects both on farmer expectations (and therefore behaviour) and on unmeasurable natural capital stocks (e.g. the condition of soil and livestock).

The set of climate variables is reduced by eliminating highly correlated terms, leaving a final set of 40 variables, as listed in Table 13. A subset of these climate variables are then selected for each equation in the statistical model, see [Estimation](#)).

Other spatial data

A range of other spatial variables are also matched to the sample farms (Table 14), including proximity to irrigation areas (based on land use map data *Land use of australia, 2010-11* 2016) and farm propensity for flooding, based on [Water Observations from Space \(WoFS\)](#) data (Mueller et al. 2016). In previous studies using AAGIS data (Hughes et al. 2016) propensity for flooding had significant and plausible effects on farm productivity (positive effects in dry years and negative in wet).

Estimation

The statistical model

The core of *farmpredict* is a statistical model which generates farm level predictions for multiple dependent (target) variables \mathbf{Y}_{it} (the 46 variables listed in Table 11) as a function of a large number of explanatory variables (predictors) $\mathbf{X}_{it} = \{\mathbf{P}_t, \mathbf{K}_{it}, \mathbf{S}_{it}^{op}, \mathbf{W}_{it}, \mathbf{Z}_{it}\}$ (as listed in Table 12, Table 13 and Table 14).

The key dependent variables in the model are defined below.

D_{jit}	Crop classification, = 1 if crop j sold on farm i in year t
A_{jit}	Proportion of farm land planted to crop j
H_{jit}	Yield for crop j , H_{jit}/A_{jit}
Q_{jit}	Proportion of stock sold for output j $= Q_{jit}/(H_{jit} + S_{jit}^{op} + S_{jit}^{purch})$, for crops and wool $= Q_{jit}/(S_{jit}^{op} + S_{jit}^{purch})$, for livestock
R_{jit}	Farm price received for output j , R_{jit}/Q_{jit}
V_{vit}	Quantity index for variable input v , C_{vit}/P_{vt}

For crops, the model predicts which crop types are planted D_{jit} , the proportion of farm land planted A_{jit} and crop yield H_{jit} . Given the potential for on-farm storage (and on-farm use of crops for livestock feed) crop sales Q_{jit} can differ from production H_{jit} . For farms with storage facilities, the model also predicts crop sales as a proportion of production and opening stocks: Q_{jit} .

Beef cattle and sheep sales are modelled as a proportion of opening stocks plus livestock purchases / transfers-in. Livestock is a complex commodity representing both an output and input (in the case of purchases) and a capital good / fixed input. In *farmpredict* livestock purchases / transfers-in are assumed exogenous (treated as a fixed input). In testing, livestock purchases displayed limited correlation with exogenous variables including climate and prices.

For most outputs, the model also predicts farm price received R_{jit} in-order to account for variation across farms in revenue per unit sold (for a given national market price P_{jt}). This variation can reflect a range of factors including differences in output quality or type (e.g., livestock weight/condition, livestock breed or crop variety etc.). While national price levels P_{jt} are considered exogenous---as they are determined by global commodity markets---farm prices received (R_{jit}/Q_{jit}) are assumed endogenous: as they can be affected by farmer decisions and are likely to be correlated with other predictor variables. For example, the types of crops and livestock produced vary by location, while crop and livestock quality can differ between wet and dry years.

For wheat, barley, sorghum and oilseeds, adjustments are also made to account for centralised marketing / pool payment schemes. Here R_{jit} is based on revenue received from crops sold within the financial year (even if that revenue is received in a subsequent year, see Appendix B and C for detail).

For inputs, predictions are only required for quantity indexes V_{vit} , given $C_{vit} = V_{vit} \cdot P_{vit}$ by definition. Other costs ($C_{othercosts}$) which include interest, rent payments and livestock purchases are treated exogenous and excluded from the statistical model.

Typically, dual form models estimate a complete set of price effects including all cross-price terms. However, as market prices P_{jt} typically contain limited cross-sectional variation (in our case none) estimation efficiency can be low (see Mundlak 2001). This study focuses on estimating own price effects (e.g., the effect of beef prices on beef output), with a limited number of cross-price terms included (see the [Machine learning](#) discussion below).

Further, noise is added to the year variable (Z_{year} in Table 12, see Appendix B) to help the model separate annual price effects P_{jt} from general time related changes (e.g., technology change). This noise effectively forces the machine learning algorithms to adopt smoother time trends and assign more of the annual variation to other model variables including prices. It is important to note that these co-linearity problems are less of a concern for other model predictors (including climate variables) given a high level of cross-sectional variation within each year.

Machine learning

The model applies the *xgboost* regression algorithm (Chen & Guestrin 2016) a popular implementation of the Gradient Boosted Regression Trees method (Friedman 2002). A version of multi-target regression stacking (Spyromitros-Xioufis et al. 2012) is applied to transform *xgboost* into a multiple target regression model. Similar to two-stage least squares the approach involves first regressing each target variable as an function of the inputs.

$$\check{Y}_{yit} = f_y^1(\mathbf{X}_{it}) \text{ for } y \in 1, \dots, 51$$

Predictions from this first stage \check{Y}_{it} are then used as predictors in second stage regressions. This process helps to account for interactions between target variables, which are important, given the joint multi-output nature of Australian broadacre farming. Following Spyromitros-Xioufis et al. (2012) we use out-of-sample predictions for \mathbf{Y}_{it} (collected via cross-validation) in the second stage. In order to exploit the panel structure of the data the second stage also incorporates lagged prediction errors $e_{yi,t-1}$, for those observations that were surveyed in a previous year:

$$\hat{Y}_{yit} = f_y^2(\mathbf{X}_{it}, \check{\mathbf{Y}}_{it}^{-y}, e_{yi,t-1})$$

$$e_{yi,t-1} = \begin{cases} \check{Y}_{yi,t-1} - Y_{yi,t-1}, & \text{if } Y_{yi,t-1} \text{ exists} \\ 0, & \text{otherwise} \end{cases}$$

To evaluate the model 10-fold cross-validation is employed. The cross-validation strategy involves 'blocking' by farm cross-sections: for each fold, each farm has all of its observations (across multiple years) grouped together in either the training or test sample. This strategy avoids an over-fitting problem where the model can identify the same farm appearing in multiple years, from predictors such as location and fixed inputs (which for many farms are relatively constant overtime). The hyper-parameters of the *xgboost* (including *nrounds*, *eta*, *min_child_weight* and *max_depth*) are tuned to minimise cross-validated Mean Absolute Error (MAE). The tuning procedure makes use of a model-based optimisation (MBO) algorithm (Bischl et al. 2017).

The model involves a large set of potential predictor variables (NNN variables in total for stage 1, see Appendix B). For each target variable a subset of predictors are selected by recursive feature elimination on the basis of the *xgboost* feature importance scores. A fixed subset of the predictors are deemed mandatory and forced to be included in each stage 1 function (see Appendix B). The stage 2 functions take all of the chosen stage 1 predictors as mandatory.

A number of other variable inclusions / exclusions are also made manually. Firstly, own price terms (e.g., effect of beef price on beef output) are mandatory. Secondly, given winter crop planting decisions are made prior to the start of the survey year, functions for D_{jit} and A_{jit} for wheat, barley and oilseeds exclude all climate variables except for lag terms and the *win* (June to August) period (see Figure 2).

Regression diagnostics

Regression diagnostics are presented in Appendix D. Table 15 and Table 16 compare the cross-validated performance of the *xgboost* model with standard Ordinary Least Squares (OLS) regression. On average (across the 39 continuous target variables) *xgboost* achieves an improvement (reduction) in mean absolute error of around 20 per cent relative to OLS. At the farm-level cross-validated R^2 averages 0.45 for the *xgboost* model and 0.27 for OLS. Note that the performance metrics (e.g., R^2) are all significantly higher in the simulation model (see the [Results](#) section), once variables are converted from ratios to levels and additional simulation assumptions are applied.

Predictive accuracy varies considerably across the target variables. Generally higher performance is achieved for the more common farm outputs (e.g., beef and wheat) and weaker performance for less common outputs (e.g., sorghum, lamb, and 'Other crops'). In some cases (R_{lamb_dot} , $R_{oilseeds_dot}$ and $V_{shearing}$) the simple OLS model obtains a strong fit, and the *xgboost* model offers limited gains. However for many key variables, particularly crop areas and yields and livestock turn-off, birth and death rates, OLS struggles and the *xgboost* model offers large gains in performance.

Performance for crop yield is generally better than livestock birth/death rates (e.g., H_{wheat_dot} has R^2 0.53 compared with S_{beef_births} 0.32). This is not surprising given the heterogeneity in livestock farming: where for example farms can specialise at different stages of the life-cycle, with some focused more on breeding and others on 'finishing' (even though dedicated feed-lot operations are excluded from the data-set).

In general, performance is relatively lower for crop and livestock turn-off rates Q_{jit} . Again this is not surprising given the potential sources of noise in these variables. As turn-off decisions affect stocks, they can be influenced longer-term planning decisions / expectations. Crop storage decisions are also affected by on-farm storage capacity, which is unobserved in the data. Further, very high livestock turn-off rates (e.g., farms selling / transferring the majority of their herd in a year) are likely to reflect random factors such as the rotation a herd between different properties or the sale of a farm business.

Table 17 (Appendix D) shows the relative contribution of the different explanatory variable types to each target variable (climate, prices, opening stocks, capital, location and other control variables). Climate conditions have a significant effect on all of the target variables, with larger effects on crop yields and livestock birth and death rates as would be expected. Further assessment of the model climate and price responses is provided in the [Results](#) section.

Simulation

farmpredict simulates production and financial outcomes at a farm level given scenarios for climate conditions, prices and other variables (Figure 3).

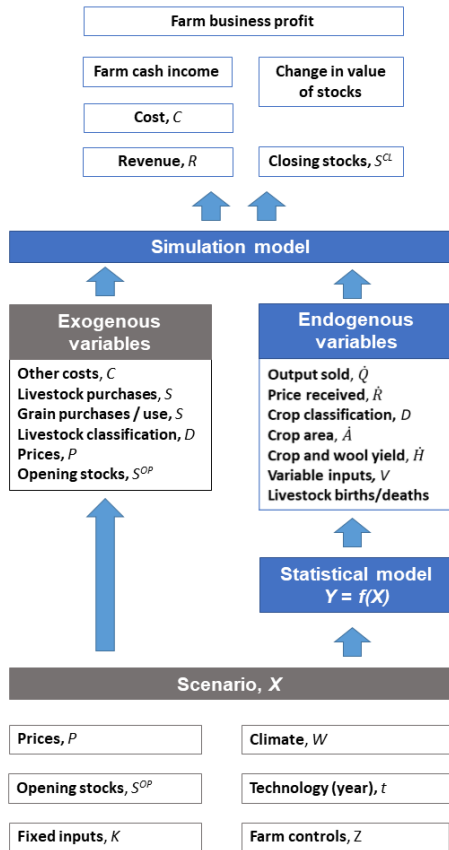


Figure 3: Generating simulation results using *farmpredict*

Model scenarios are defined by a data / assumptions for the predictors $\tilde{\mathbf{X}}_{it}$ (the *baseline* scenario takes the actual data, $\tilde{\mathbf{X}}_{it} = \mathbf{X}_{it}$). Alternative scenarios (see [Results](#)) involve assumptions for climate and price variables, usually taking farm fixed inputs, stocks and control variables as given by the survey data. For any scenario predicted values for the target variables $\hat{\mathbf{Y}}_{it}$ can be generated from the statistical model. A range of simulation results can then be generated including closing stocks and profit measures (see Appendix C for details). Profit measures produced by the model include *farm cash income*, *farm business profit* and *rate-of-return* all defined in keeping with [AAGIS definitions](#).

In producing these results, the simulation model takes into account a range of additional variables not included within the statistical model, including 'other costs' $C_{othercosts}$ (including interest, rent and livestock purchases), 'other revenue' R_{other} . These variables all remain exogenous and are held fixed at observed values in all scenarios (see Appendix C).

3 Results

Validation

Validation results are presented below in Table 4. More detailed results for a wider selection of model variables are presented in Appendix E.

Two validation scenarios are considered. The *baseline* scenario takes the out-of-sample regression predictions obtained through cross-validation in the model estimation stage (as described above). While out-of-sample predictions are used, some aspects of the simulation model make use of observed (in-sample) farm data in generating final profit measures (including the 'other' revenue and costs components, see Appendix C). As such, the *baseline* scenario may slightly overstate out-of-sample performance. To account for this we also include a *lagcast* scenario, which predicts farm outcomes using the previous years farm data and observed price and climate data (i.e., similar to a hindcast but with known prices and climate conditions). Note the *lagcast* scenario underestimates model performance at the region and national level, as it is subject to additional annual sampling error in the AAGIS data.

As shown in Table 4 performance improves as we move from farm, to regional to national scales. In general, results for the *lagcast* scenario show a similar pattern to the *baseline*, but with slightly weaker performance as would be expected. Model predictions for profit have limited skill at the farm level (R^2 of 0.38 for farm cash income and 0.23 for farm business profit), although farm level performance is much higher for other model variables including revenues and costs. Given the relatively low profit levels in the sample (with approximately half the farms experiencing losses) small errors in revenue or cost predictions can have large relative effects on simulated profits.

Table 4: R^2 for the *baseline* and *lagcast* scenarios, at the farm, regional and national levels

	<i>baseline</i>			<i>lagcast</i>		
	Farm level	Regional	National	Farm level	Regional	National
Wheat revenue, R_wheat	0.77	0.98	0.96	0.82	0.94	0.67
Beef revenue, R_beef	0.86	0.96	0.99	0.88	0.87	0.86
Wool revenue, R_wool	0.90	0.99	1.00	0.90	0.93	0.97
Lamb revenue, R_lamb	0.76	0.96	0.99	0.40	0.74	0.89
Total revenue, R_total	0.84	0.94	0.99	0.85	0.80	0.91
Fertiliser costs, C_fert	0.71	0.97	0.96	0.76	0.94	0.90
Fuel costs, C_fuel	0.76	0.94	0.98	0.81	0.84	0.83
Other materials costs, C_othermat	0.56	0.83	0.52	0.63	0.77	0.43
Services and labour costs, C_serv	0.80	0.93	0.94	0.84	0.83	0.72
Total costs, C_total	0.92	0.97	0.98	0.82	0.76	0.85
Farm cash income, FBP_fci	0.38	0.69	0.96	0.44	0.57	0.89
Farm business profit, FBP_fbp	0.23	0.65	0.92	0.37	0.58	0.80

Model skill in predicting profit improves considerably at higher scales with an R^2 of 0.57-0.69 at the regional level and 0.8-0.96 nationally (see Figure 4). The performance of the model by region is summarised in Figure 5 and Table 20 (Appendix E). Performance is strongest in south-eastern and south-western Australia and weakest in central WA and coastal NSW and QLD. These differences can be partly explained by sample size, with the best performing regions tending to be those with the most sample points (see Table 20).

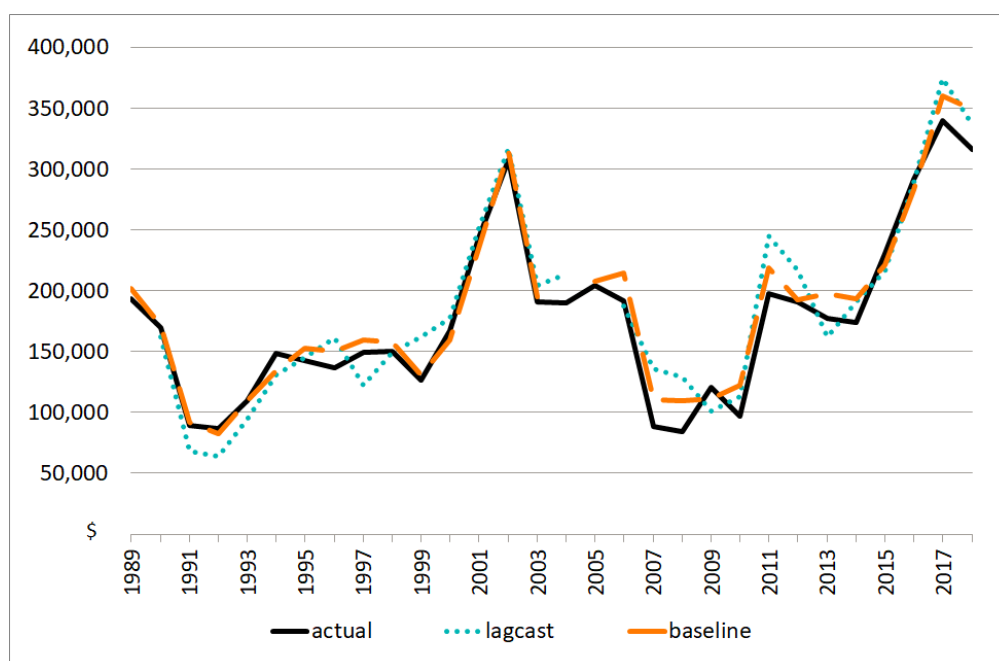


Figure 4: Average annual farm cash income actual and predicted, 1988-89 to 2017-18

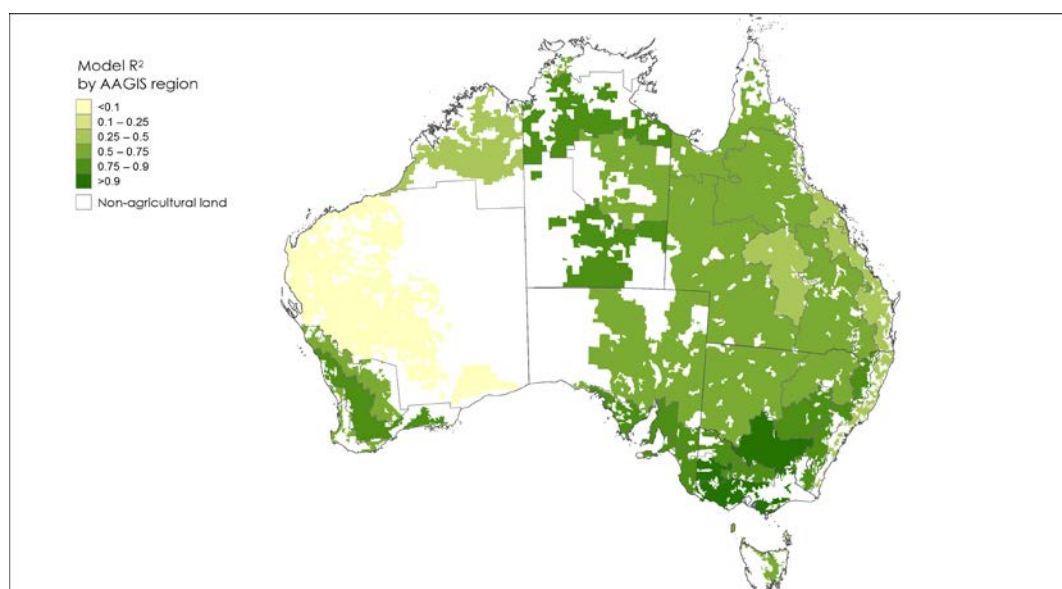


Figure 5: Regional R^2 for farm business profit under the *baseline* scenario

Some bias is observed in the predictions of costs due to end-point bias in the variable input regression models, such that costs are underestimated for very large farms. In the *baseline* scenario total costs are slightly underestimated on average (by 2.5 per cent) and farm business profit slightly overestimated (see Figure 4). This bias has limited relevance for scenario analysis, but would require some post-model bias-correction in forecasting applications.

The most affected input type is $C_{othermat}$, which is underestimated by nearly 20 per cent on average. While $C_{othermat}$ only accounts for around 10 per cent of total costs on average it is important in reflecting the effects of drought, because it includes the costs of purchased grain / hay for livestock. Improving the performance of this input is a subject for future research.

Price responses

Given the non-parametric approach employed, the responses of the model to specific variables can be complex: both non-linear and farm, time and location specific (depending on interaction effects with other variables). However, it is possible to estimate average farm responses to price changes through simulation.

Here the model is used to simulate a 10 per cent increase for a given output or input price relative to long run average prices (holding all other prices fixed). Long run average prices are computed as the 20-year average price of the input or output over the period 1988-89 to 2017-18. Table 5 shows the average own-price effects (e.g., effect of wheat prices on average wheat sales, production and stocks).

Table 5: Average own-price effect (%) for a 10% increase in output / input prices

	Production / births	Quantity sold / used	Closing stock
Output prices			
Wheat	2.0	1.9	1.5
Barley	-0.9	-0.8	-0.8
Oilseeds	1.7	1.7	-1.0
Sorghum	0.9	0.2	2.5
Beef	0.4	-0.0	0.3
Sheep	1.2	0.3	0.1
Lamb	0.0	0.2	0.0
Wool	-0.1	-0.0	-0.3
Input prices			
Fuel		-1.1	
Fertiliser		-1.9	
Chemicals		-2.2	

The average price responses are broadly consistent with economic theory: output price effects are positive (higher price, higher output) and input effects negative. A key exception is barley where the own-price effect is negative. This is due to a high level of co-linearity between barley and wheat prices in the sample period. Output supply responses are larger for crops than for livestock which is also plausible. Crop production can be increased in the short-term by increasing area planted and by applying more inputs. However, livestock production is somewhat constrained by the current herd, further increases in price provide an incentive to increase herd-size. As shown in Table 5 higher livestock prices also increase birth rates and closing stocks for beef and sheep.

Table 6 shows the average effects of a price change in a given output / input on farm total revenue, total cost and farm cash income. Here the model is able to simulate plausible cross price effects, with farm costs (input use) generally increasing in response to an output price

rises, while farm revenue (output) decreases following an increase in input prices. Beef prices have the largest effect on farm cash incomes on average, reflecting their larger share of average farm revenue.

Table 6: Average farm-wide effects (%) for a 10% increase in output / input prices

	Total revenue	Total cost	Farm cash income
Output prices			
Wheat	1.5	-0.0	6.6
Barley	0.5	0.1	1.9
Oilseeds	0.1	0.0	0.5
Sorghum	0.7	0.2	2.5
Beef	6.0	0.4	25.2
Sheep	0.4	0.0	1.6
Lamb	0.8	-0.2	4.2
Wool	0.9	0.1	3.9
Input prices			
Fuel	-0.0	0.3	-1.0
Fertiliser	-0.1	0.3	-1.3
Chemicals	-0.1	0.5	-2.0

Climate responses

Climate variable importance

Table 17 (Appendix D) shows the importance of climate on each model target variable including the relative importance of rainfall and temperature variables. Crop yields depend heavily on rainfall (accounting for around 90% of their climate effect), while key livestock variables (turn-off, birth and death rates) are relatively more dependent on temperature. For example, around 50% of the climate effect on cattle birth rate is due to temperature.

Table 18 lists the most important relationships between specific target and climate variables identified by the model. Not surprisingly the strongest climate relationship in the model is that between winter growing season rainfall and winter crop yields (wheat, barley and oilseeds). As would be expected sorghum production depends heavily on summer rainfall. A range of temperature related variables are found to have effects on livestock production, including lagged autumn maximum temperatures on cattle birth rates (which could reflect impacts on livestock mating / fertility).

Drought responses

To examine the model responses to climate further we define a *historical_climate* scenario, which takes a recent cohort of farms (2015-16 to 2017-18) and simulates the effect of observed

climate (at each farm location) for the period 1950-51 to 2018-19. This scenario reflects what would happen to current farms (with farm size, capital stocks, livestock holdings, technology and prices as observed between 2015-16 and 2017-18) under alternative historical climate conditions.

A detailed examination of the effects of historical and projected changes in climate on Australian farms remains an obvious future subject for the *farmpredict* model. For now we focus on the model's responses to drought conditions, using the recent 2018-19 drought as an example. Here we calculate percentage effects under 2018-19 conditions (relative to median conditions for the period 2000-01 to 2018-19) for farms in NSW (where the drought was most severe). Table 7 shows the effects on crop production, Table 8 livestock production, and Table 9 farm profits. Note these results show the effect of climate only and are not a forecast of 2018-19 farm outcomes (which would depend also on 2018-19 prices).

Table 7: Average effect (%) of 2018-19 drought conditions on crop production (NSW farms)

	Wheat	Barley	Oilseeds	Sorghum
Area planted	-7.8	-5.0	-13.3	-64.4
Yield	-36.9	-29.1	-29.1	-10.2
Harvest	-51.9	-44.5	-46.5	-70.7
Quantity sold	-49.9	-40.7	-46.8	-54.5
Closing stock	-32.8	-26.2	-26.2	-48.5
Price received	1.8	2.1	-1.6	-1.4
Crop revenue	-48.7	-37.9	-48.2	-54.3

Table 8: Average effect (%) of 2018-19 drought conditions on livestock production (NSW farms)

	Beef	Sheep	Lamb	Wool
Quantity sold	5.3	6.4	3.6	-6.2
Production				-6.9
Price received	-3.7	-0.7	-0.7	-1.9
Revenue	0.5	4.3	3.2	-7.4
Births	-5.7	-9.3		
Deaths	6.4	11.9		
Closing stock	-4.6	-7.6		-4.5

Table 9: Average effect (%) of 2018-19 drought conditions on revenue, cost and profit (NSW farms by industry)

	Cropping	Mixed	Sheep	Beef	Sheep-Beef
Wheat revenue	-47.6	-49.1	-57.4	-75.6	-34.3
Beef revenue	3.2	-0.2	-0.7	-0.1	2.2
Wool revenue	-7.2	-5.1	-10.4	-6.7	-4.8
Lamb revenue	2.0	4.8	2.6	-0.2	0.6
Total revenue	-29.0	-17.4	-5.4	-0.9	-0.4
Fertiliser costs	-23.1	-22.6	-13.6	-13.7	-14.5
Fuel costs	0.3	0.4	2.5	3.3	4.0
Other materials costs	4.6	9.9	14.0	13.4	13.8
Services and labour costs	-6.0	-5.0	-1.1	-2.4	-1.4
Total costs	-6.6	-4.8	-0.3	-0.4	-0.3
Farm cash income	-117.3	-46.7	-16.1	-2.2	-0.5
Wheat closing stock	-30.5	-34.9	-40.0	-47.9	-22.8
Beef closing stock	-5.2	-5.0	-3.5	-4.5	-5.0
Sheep closing stock	-7.0	-7.6	-8.8	-5.0	-6.6
Farm business profit	-328.8	-121.2	-86.9	-43.6	-42.9

As would be expected, drought leads to large reductions in crop areas planted, yield and production (Table 7). For grain crops these effects are offset slightly by an increase in prices received (reflecting improved grain quality / higher protein content). This effect is reversed for oilseeds---prices received are lower in drought---which is expected given oilseeds deteriorate in hotter and dry conditions. Ultimately drought leads to large reductions in revenue on cropping farms, offset only marginally by reductions in crop related input costs (such as fertiliser), resulting in negative average farm cash incomes and profits on average (greater than 100 per cent decreases, Table 9).

With livestock, drought leads to both an increase in turn-off rates and a decrease in prices received: reflecting negative effects of drought on livestock condition / quality (Table 8). As such, limited change in revenue is observed on livestock farms on average. Some increases in input costs are observed (related to higher livestock fodder requirements) although these are offset by reductions in cropping related costs. More significant are the effects on livestock holdings. Here we see lower birth rates and higher death rates and turn-off, leading to significant reductions in closing stocks, and in-turn farm business profits for both beef and sheep farms.

Note that under AAGIS definitions, farm business profit includes the value of changes in herd size, but not changes in value caused by prices (i.e., capital depreciation). If the drought-induced decline in livestock quality applies equally to all cattle (not just those sold) then the effective decline in farm profit would be larger than that shown in Table 9.

4 Conclusions

farmpredict is a data-driven bio-economic micro-simulation model of Australian broadacre farms. The model adopts a similar structure to previous reduced-form econometric farm models: with variable inputs and outputs modelled as functions of fixed inputs, prices and climate variables. A non-parametric machine learning approach is taken to estimating the mode, using a large farm-level panel dataset linked to location specific climate data. The resulting model is capable of simulating production and financial outcomes at a farm level across Australia for a diverse mix of cropping, livestock and mixed farm types.

Validation results show that the model has good skill in predicting production and revenue for major crop and livestock commodities. Model accuracy is limited in predicting profit at the individual farm level, although performance improves considerably at higher spatial scales. Ultimately, predicting profit at the farm level is a difficult task, given high levels of noise and heterogeneity in farm level data. This is compounded by the low profit levels of farms in the sample, which means that small errors in revenue or costs translate into large errors in profit.

Simulation results show that the model generates realistic responses to price and climate shocks consistent with economic and bio-physical fundamentals. Increases in crop prices lead to higher crop production and crop related input use; increases in livestock prices lead to higher birth rates and closing stocks; higher input prices lead to lower input use and lower farm output.

The simulation results demonstrate the differing effects of drought on cropping and livestock farms. Drought conditions lead to large decreases in crop areas planted, yield, production and revenue. As a result cropping farms experience large and immediate reductions in farm cash income. On livestock farms, climate effects are transmitted more through changes in livestock holdings, with drought leading to significant decreases in beef and sheep herds due to a combination of lower birth rates and higher turn-off and death rates.

farmpredict has a range of potential applications. Currently the model is being applied to assess the effects of recent and future potential changes in climate on the farm profitability. Plans are also underway to develop *farmpredict* for forecasting by linking the model with BOM seasonal climate outlooks. *farmpredict* could also be used to develop indicators of drought exposure and sensitivity which could help to inform government farm risk management and drought programs. Finally, *farmpredict* could have financial sector applications, including assessing farm lender exposure to climate change, and designing and testing weather insurance products.

Future research could improve *farmpredict* in a number of directions. In-particular, the model's treatment of livestock fodder costs is less than ideal (currently these costs are included in a broad 'other materials' variable, for which the model's performance is relatively low). This could be addressed by making farm grain and hay consumption (quantities) endogenous to the model. In time, this could also allow a partial equilibrium representation of domestic hay / grain markets to be added to the model. This is important in the context of the recent 2018-19 drought, where wide-spread dry conditions led to spikes in domestic grain prices, above international prices. Further, the model framework could easily be extended to Australian dairy farms, by making use of ABARES Australian Dairy Industry Survey (ADIS).

5 Appendix A: AAGIS regions

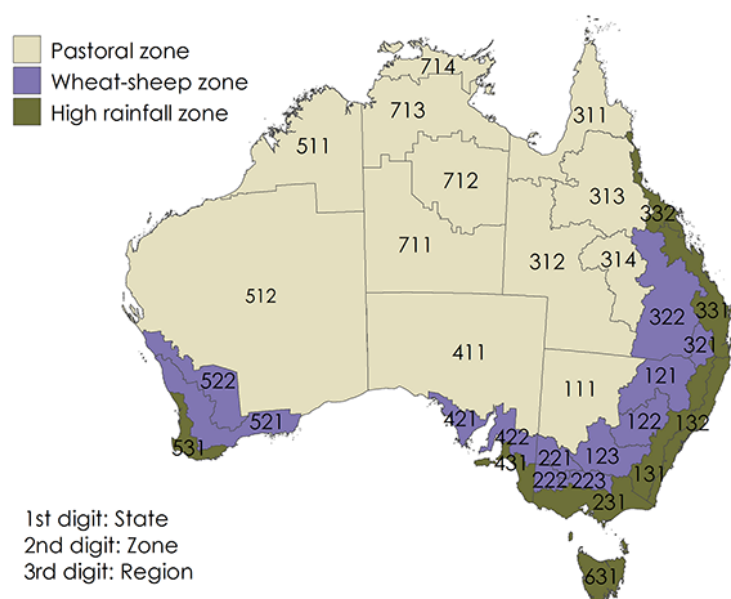


Figure 8: Map Australian Agricultural and Grazing Industries Survey (AAGIS) zones and regions

Table 10: AAGIS regions

State	Region name	Code
New South Wales (NSW)	Far West	111
New South Wales (NSW)	North West Slopes and Plains	121
New South Wales (NSW)	Central West	122
New South Wales (NSW)	Riverina	123
New South Wales (NSW)	Tablelands	131
New South Wales (NSW)	Coastal	132
Victoria (VIC)	Mallee	221
Victoria (VIC)	Wimmera	222
Victoria (VIC)	Central North	223
Victoria (VIC)	South-Eastern Victoria	231
Queensland (QLD)	Cape York and the Gulf	311
Queensland (QLD)	West and South West	312
Queensland (QLD)	Central North	313
Queensland (QLD)	Charleville - Longreach	314

ABARES working paper

Queensland (QLD)	Eastern Darling Downs	321
Queensland (QLD)	Darling Downs & Central Highlands	322
Queensland (QLD)	South Queensland Coastal	331
Queensland (QLD)	North Queensland Coastal	332
South Australia (SA)	North Pastoral	411
South Australia (SA)	Eyre Peninsula	421
South Australia (SA)	Murray Lands and Yorke Peninsula	422
South Australia (SA)	South East	431
Western Australia (WA)	The Kimberly	511
Western Australia (WA)	Pilbara and the Central Pastoral	512
Western Australia (WA)	Central and South Wheat Belt	521
Western Australia (WA)	North and East Wheat Belt	522
Western Australia (WA)	South West Coastal	531
Tasmania (TAS)	Tasmania	631
Northern Territory (NT)	Alice Springs Districts	711
Northern Territory (NT)	Barkly Tablelands	712
Northern Territory (NT)	Victoria River District - Katherine	713
Northern Territory (NT)	Top End Darwin and the Gulf	714

6 Appendix B: Model variables

Dependent variables

Table 11: Model target / dependent variables

	Description	Unit
A_barley_dot	Proportion of land planted to barley, A_barley / K_land	proportion
A_double_dot	Proportion of land double cropped, A_double / K_land	proportion
A_oilseeds_dot	Proportion of land planted to oilseeds, $A_oilseeds / K_land$	proportion
A_othercrops_dot	Proportion of land planted to othercrops, $A_othercrops / K_land$	proportion
A_sorghum_dot	Proportion of land planted to sorghum, $A_sorghum / K_land$	proportion
A_total_cropped_dot	Proportion of land cropped, $A_total_cropped / K_land$	proportion
A_wheat_dot	Proportion of land planted to wheat, A_wheat / K_land	proportion
D_barley	Barley planted, =1 if $A_barley > 0$	binary
D_double	Double cropping, =1 if $A_double > 0$	binary
D_oilseeds	Oilseeds planted, =1 if $A_oilseeds > 0$	binary
D_othercrops	Other crops planted, =1 if $A_othercrops > 0$	binary
D_sorghum	Sorghum planted, =1 if $A_sorghum > 0$	binary
D_wheat	Wheat planted, =1 if $A_wheat > 0$	binary
H_barley_dot	Barley yield, H_barley / A_barley	t / ha
H_oilseeds_dot	Oilseeds yield, $H_oilseeds / A_oilseeds$	t / ha
H_sorghum_dot	Sorghum yield, $H_sorghum / A_sorghum$	t / ha
H_wheat_dot	Wheat yield, H_wheat / A_wheat	t / ha
H_wool_dot	Wool yield, $H_wool / (S_sheep_op + S_sheep_purch + S_sheep_births)$	t / ha
Q_barley_dot	Proportion of barley sold, $Q_barley / (S_barley_op + H_barley + S_barley_purch)$	proportion
Q_beef_dot	Proportion of beef cattle sold, $Q_beef / (S_beef_op + S_beef_purch + S_beef_births)$	proportion
Q_lamb_dot	Proportion of lamb sold, $Q_lamb / (S_sheep_op + S_sheep_purch)$	proportion

ABARES working paper

Q_oilseeds_dot	Proportion of oilseeds sold, $Q_oilseeds / (S_oilseeds_op + H_oilseeds + S_oilseeds_purch)$	proportion
Q_othercrops	Quantity index for other crop output	index
Q_sheep_dot	Proportion of sheep sold, $Q_sheep / (S_sheep + S_sheep_purch + S_sheep_births)$	proportion
Q_sorghum_dot	Proportion of sorghum sold, $Q_sorghum / (Q_sorghum_op + H_oilseeds + S_sorghum_purch)$	proportion
Q_wheat_dot	Proportion of wheat sold, $Q_wheat / (S_wheat_op + H_wheat + S_wheat_purch)$	proportion
Q_wool_dot	Proportion of wool sold, $Q_wool / (S_wool_op + H_wool)$	proportion
R_barley_dot	Barley price received adjusted for pool payments, $(R_barley - R_barley_pool_pmt + R_barley_future_pool_pmt) / Q_barley$	\$ / t
R_beef_dot	Beef cattle price received, R_beef / Q_beef	\$ / no.
R_lamb_dot	Lamb cattle price received, R_lamb / Q_lamb	\$ / no.
R_oilseeds_dot	Oilseeds price received adjusted for pool payments, $(R_oilseeds - R_oilseeds_pool_pmt + R_oilseeds_future_pool_pmt) / Q_oilseeds$	\$ / t
R_sheep_dot	Sheep price received, R_sheep / Q_sheep	\$ / no.
R_sorghum_dot	Sorghum price received adjusted for pool payments, $(R_sorghum - R_sorghum_pool_pmt + R_sorghum_future_pool_pmt) / Q_sorghum$	\$ / t
R_wheat_dot	Wheat price received adjusted for pool payments, $(R_wheat - R_wheat_pool_pmt + R_wheat_future_pool_pmt) / Q_wheat$	\$ / t
R_wool_dot	Wool price received, R_wool / Q_wool	\$ / kg
S_beef_births_dot	Beef cattle birth rate, $S_beef_births / (S_beef_op + S_beef_purch)$	proportion
S_beef_deaths_dot	Beef cattle death rate, $S_beef_deaths / (S_beef_op + S_beef_purch)$	proportion
S_sheep_births_dot	Sheep birth rate, $S_sheep_births / (S_sheep_op + S_sheep_purch)$	proportion
S_sheep_deaths_dot	Sheep death rate, $S_sheep_deaths / (S_sheep_op + S_sheep_purch)$	proportion

V_chem	Quantity index for crop and pasture chemical inputs	index
V_fert	Quantity index for fertiliser input	index
V_fuel	Quantity index for fuel, oil and grease inputs	index
V_othermat	Quantity index for other material inputs	index
V_serv	Quantity index for service & labour inputs	index
V_shearing	Quantity index for shearing labour input	index
Z_conditions_4	Farmer assessment of seasonal conditions: drought	binary

Explanatory variables

Farm variables

Table 12: Model features (farm variables)

Name	Description	Unit	Mandatory
K_build	Quantity index for buildings and fixed improvements capital	index	Yes
K equip	Quantity index for plant and equipment capital	index	Yes
K_land	Total farm land area operated (average of opening and closing)	ha	Yes
K_otherlive	Quantity index for other livestock capital	\$	No
K_stock_op	Value of trading stocks on hand at 1 July	\$	No
S_barley_op	Barley on hand 1 July	t	No
S_beef_op	Beef cattle on hand at 1 July	no.	Yes
S_hay_op	Hay on hand 1 July	t	No
S_horses_op	Horses on hand 1 July	t	No
S_oilseeds_op	Oilseeds on hand 1 July	t	No
S_othercrops_op	Other crops on hand 1 July	t	No
S_pigs_op	Pigs on hand 1 July	t	No
S_sheep_op	Sheep on hand at 1 July	no.	Yes
S_sorghum_op	Sorghum on hand 1 July	t	No
S_wheat_op	Wheat on hand 1 July	t	Yes
S_wool_op	Wool on hand 1 July	t	Yes

Z_age	Age of operator-manager	years	Yes
Z_debt_op	Total debt at 1 July	\$	Yes
Z_edu_1	Highest education level is: No Schooling	binary	No
Z_edu_2	Highest education level is: Primary School	binary	No
Z_edu_3	Highest education level is: 1-4 years high school	binary	No
Z_edu_4	Highest education level is: 5-6 years high school	index	No
Z_edu_5	Highest education level is: trade completed	binary	No
Z_famlabour	Family weeks worked	wks	Yes
Z_ind_1	Industry: cropping	binary	No
Z_ind_2	Industry: mixed crop-livestock	binary	No
Z_ind_3	Industry: sheep	binary	No
Z_ind_4	Industry: beef	binary	No
Z_ind_5	Industry: beef-sheep	binary	No
Z_irrigcrop	Irrigated crops (1 =Yes, 0=No)	indicator	No
Z_land_price_op	Land value (farmer estimate) opening	\$ / ha	No
Z_lat	Latitude	degrees	Yes
Z_long	Longitude	degrees	Yes
Z_nonfarmincome	Total non-farm income	\$	Yes
Z_north	Farm in northern beef region (1=Yes, 0=No)	indicator	No
Z_state_1	State is NSW, =1 if Z_state=1	binary	No
Z_state_2	State is VIC, =1 if Z_state=2	binary	No
Z_state_3	State is QLD, =1 if Z_state=3	binary	No
Z_state_4	State is SA, =1 if Z_state=4	binary	No
Z_state_5	State is WA, =1 if Z_state=5	binary	No
Z_state_6	State is TAS, =1 if Z_state=6	binary	No
Z_state_7	State is NT, =1 if Z_state=7	binary	No
Z_year	Financial year with noise added, year ± 1	annual	Yes

Here Z_{year} is defined as:

$$Z_{year} = \begin{cases} t + 1, & \text{if } \sigma < 1/3 \\ t, & \text{if } 1/3 \geq \sigma \geq 2/3 \\ t - 1, & \text{if } \sigma > 2/3 \end{cases}$$

$$\sigma \sim U[0,1]$$

Climate variables

Table 13: Model features (climate variables)

Name	Description	Unit
W_winter_rain	Rainfall volume for the winter growing season (April to October)	mm
W_aut_tmax_L1	Average maximum temperature for the autumn season (March to May), previous year	degrees C
W_sum_rain	Rainfall volume for the summer season (December to February)	mm
W_winter_gdd	Heat accumulation (growing degree days) for the winter growing season (April to October)	degrees C
W_win_tmin	Average minimum temperature for the winter season (June to August)	degrees C
W_summer_gdd	Heat accumulation (growing degree days) for the summer growing season (November to March)	degrees C
W_aut_moist	Root zone soil moisture for the autumn season (March to May)	index
W_aut_tmin	Average minimum temperature for the autumn season (March to May)	degrees C
W_FY_rain	Rainfall volume for the financial year (July to June)	mm
W_sum_tmax_L1	Average maximum temperature for the Summer (December to February), previous year	degrees C
W_spr_rain	Rainfall volume for the spring season (September to November)	mm
W_aut_moist_L1	Root zone soil moisture for the autumn season (March to May), previous year	index
W_aut_rain	Rainfall volume for the autumn season (March to May)	mm
W_aut_gni_L1	Rainfall volatility (gini coefficient) for the autumn season (March to May), previous year	index
W_aut_tmax	Average maximum temperature for the autumn season (March to May)	degrees C
W_FY_rain_L1	Rainfall volume for the previous financial year	mm
W_FY_moist_L2	Root zone soil moisture for the average of two previous	index

ABARES working paper

	financial years	
W_sum_tmax	Average maximum temperature for the summer season (December to February)	degrees C
W_win_moist	Root zone soil moisture for the winter season (June to August)	index
W_FY_rain_L2	Rainfall volume for the average of two previous financial years	mm
W_win_rain	Rainfall volume for the winter season (June to August)	mm
W_aut_rain_L1	Rainfall volume for the autumn season (March to May), previous year	mm
W_spr_moist	Root zone soil moisture for the spring season (September to November)	index
W_win_gni	Rainfall volatility (gini coefficient) for the winter season (June to August)	index
W_winter_hgdd	Expsoure to extreme high temperature for the winter growing season (April to October)	degrees C
W_sum_rain_L1	Rainfall volume for the Summer (December to February), previous year	mm
W_summer_rain_L1	Rainfall volume for the summer growing season (November to March)	mm
W_summer_hgdd	Expsoure to extreme high temperature for the summer growing season (November to March)	degrees C
W_FY_moist_L1	Root zone soil moisture for the previous financial year	index
W_summer_moist	Root zone soil moisture for the summer growing season (November to March)	index
W_spr_gni	Rainfall volatility (gini coefficient) for the spring season (September to November)	index
W_FY_moist	Root zone soil moisture for the financial year (July to June)	index
W_aut_pci_L1	Rainfall volatility (gini coefficient) for the autumn season (March to May), previous year	index
W_spr_pci	Rainfall volatility (gini coefficient) for the spring season (September to November)	index
W_win_fr2	Exposure to frost (days below 2 C) for the winter season (June to August)	days
W_summer_moist_L1	Root zone soil moisture for the summer growing season (November to March)	index

W_aut_fr2_L1	Exposure to frost (days below 2 C) for the autumn season (March to May), previous year	days
W_win_pci	Rainfall volatility (gini coefficient) for the winter season (June to August)	index
W_spr_fr2	Exposure to frost (days below 2 C) for the spring season (September to November)	days
W_win_hail	Exposure to hail storms for the winter season (June to August)	index

Selected climate measures are defined below.

$$gdd = \sum_{t=1}^{t=n} \begin{cases} (tmax_t + tmin_t)/2 - 8, & \text{if } tmax_t < 32 \text{ and } tmin_t \geq 8 \\ 0, & \text{otherwise} \end{cases}$$

$$hgdd = \sum_{t=1}^{t=n} \begin{cases} tmax_t - 32, & \text{if } tmax_t > 32 \\ 0, & \text{otherwise} \end{cases}$$

$$hail = \max\{1 - dhail/50, 0\}$$

Where:

$tmax_t, tmin_t$ are the max. and min. temperatures on day $t \in 1, \dots, n$ for a given season

$dhail$ is the distance (in kilometers) to the nearest hail storm for a given farm in a given season

Other spatial data

Table 14: Other spatial predictors / explanatory variables

Name	Description	Unit
Z_flood1	Average frequency of flooding ignoring areas flooded more than 1.5% of the time	proportion
Z_flood20	Average frequency of flooding ignoring areas flooded more than 20% of the time	proportion
Z_flood5	Average frequency of flooding ignoring areas flooded more than 5.0% of the time	proportion
Z_irrig1	Relative distance to irrigation area, = $Z_irrig2/(K_land)^{0.5}$	index
Z_irrig2	Distance to nearest irrigation area	km
Z_rain_avg	Average annual rainfall, since 1977-78	mm

7 Appendix C: Simulation

This section details the simulation component of *farmpredict*, which takes predictions for target variables $\hat{\mathbf{Y}}_{it}$ (for a scenario $\tilde{\mathbf{X}}$) and then generates simulation results $\tilde{\mathbf{Y}}_{it}$ including estimates of farm revenue, cost, profit and changes in stocks. For convenience, below we mix symbolic notation (e.g., R_{jit}) with model code labels (e.g., R_wheat , R_beef etc., see Appendix B for variable definitions). Here we also introduce the suffix *hat* to separate model predicted / simulated variables from observed / exogenous variables.

To begin, predictions for livestock outputs (Q_beef_hat , Q_sheep_hat , Q_lamb_hat , H_wool_hat) and livestock births and deaths ($S_beef_births_hat$, $S_beef_deaths_hat$, $S_sheep_births_hat$, $S_sheep_deaths_hat$) are recovered by multiplying model predicted ratios (e.g., \hat{Q}_{jit}) by their relevant denominators (see Table 11).

For all of these variables, simulated output is set to zero if actual output is zero (i.e., decisions over whether or not to produce livestock outputs remain exogenous).

$$Q_beef_hat = \begin{cases} Q_beef_dot_hat \cdot (S_beef_op + S_beef_purch) & \text{if } Q_beef > 0 \\ 0 & \text{otherwise} \end{cases}$$

Next, crop areas planted are defined as:

$$\tilde{A}_{jit} = K_land \cdot \hat{A}_{jit} \hat{D}_{jit}$$

Here the farm decision to plant a crop is endogenous (determined by \hat{D}_{jit}). Next crop production (H_wheat_hat , H_barley_hat , $H_oilseeds_hat$, $H_sorghum_hat$) is defined as yield times area planted:

$$\tilde{H}_{jit} = \hat{H}_{jit} \tilde{A}_{jit}$$

Both crop and wool sales are then calculated as a proportion of total quantity available for sale (production plus opening stocks):

$$\tilde{Q}_{jit} = \hat{Q}_{jit} (\tilde{H}_{jit} + \tilde{S}_{jit}^{op})$$

Next revenue for livestock outputs (R_beef_hat , R_sheep_hat , R_lamb_hat , R_wool_hat) is defined as price received times quantity sold:

$$\tilde{R}_{jit} = \hat{R}_{jit} \cdot \tilde{Q}_{jit}$$

The equations for crop revenue (R_wheat_hat , R_barley_hat , $R_oilseeds_hat$, $R_sorghum_hat$) account for farm participation in crop pool payment schemes, where revenue from a given crop harvest may be spread over several years. For example:

$$R_wheat_hat = R_wheat_pool_pmt + R_wheat_dot_hat \cdot \tilde{Q}_{jit} \cdot (1 - R_wheat_pool_dot)$$

where $R_wheat_pool_pmt$ includes any crop pool payments received this year for crops sold in previous years and $R_wheat_pool_dot$ is the proportion of revenue from crops sold this year which will be received next year. Both $R_wheat_pool_pmt$ and $R_wheat_pool_dot$ are exogenous and both are zero for farms that do not participate in crop pools.

Variable costs (and other crop receipts) are then recovered by multiplying by their relevant price indexes:

$$\tilde{C}_{vit} = \hat{V}_{vit} \cdot \tilde{P}_{vt}$$

$$R_{othercrops_{hat}} = Q_{othercrops_{hat}} \cdot P_{othercrops}$$

For beef cattle and sheep, closing stocks are defined as:

$$S_{beef_cl_hat} = S_{beef_op} - Q_{beef_hat} + S_{beef_purch} + S_{beef_births_hat} - S_{beef_deaths_hat}$$

$$S_{sheep_cl_hat} = S_{sheep_op} - Q_{sheep_hat} - Q_{lamb} + S_{sheep_purch} + S_{sheep_births_hat} - S_{sheep_deaths_hat}$$

where S_{sheep_purch} and S_{beef_purch} are exogenous and all other variables are model predictions.

Crop closing stocks ($S_{wheat_cl_hat}$, $S_{barley_cl_hat}$, $S_{sorghum_cl_hat}$ and $S_{oilseeds_cl_hat}$) are defined as:

$$\tilde{S}_{sit}^{cl} = S_{sit}^{op} + \tilde{H}_{jit} \cdot \tilde{A}_{jit} - \tilde{Q}_{jit} + S_{sit}^{purch}$$

where S_{sit}^{purch} is exogenously determined and reflects net crop usage / purchase.

Wool stocks are defined as:

$$S_{wool_cl_hat} = S_{wool_op} + H_{wool_hat} - Q_{wool_hat}$$

Next $farm_cash_income$ can be calculated as:

$$farm_cash_income_hat = \sum^j \tilde{R}_{jit} + R_{other} - \sum^v \tilde{C}_{vit} - C_{othercosts}$$

where $C_{othercosts}$, R_{other} are exogenous (fixed at observed values). Finally, $farm_business_profit$ is defined as below, in accordance with standard AAGIS definitions:

$$farm_business_profit_hat = farm_cash_income_hat + \sum^s \tilde{P}_{st}^{\$} (\tilde{S}_{sit}^{cl} - \tilde{S}_{sit}^{op}) + FBP_fbp_resid + FBP_beef_rations \cdot \tilde{P}_{beef,t}^{\$} + FBP_sheep_rations \cdot \tilde{P}_{sheep,t}^{\$}$$

where FBP_fbp_resid includes other crop and livestock net gain in stocks less deductions for depreciation and the value of family labour. Here $\tilde{P}_{st}^{\$}$ is an average simulated unit price, at a state level for livestock and at a national level for crops, for example:

$$\tilde{P}_{st}^{\$} = \sum^i \tilde{R}_{sit} / \sum^i \tilde{Q}_{sit}$$

Finally, $farm_rate-of-return$ is defined as:

$$FBP_ror_hat = (FBP_fbp_hat + FBP_pfe_resid) / K_total_capital_op$$

where FBP_pfe_resid includes adjustments for interest and other financing costs (which in keeping with AAGIS translates $farm_business_profit$ into $profit\ at\ full\ equity$).

8 Appendix D: Regression results

Table 15: Cross-validated regression performance, xgboost stack versus least squares regression (ols)

	Sample	xgboost R^2	OLS R^2	xgboost RMAE	OLS RMAE	RMAE xgboost / OLS
A_barley_dot	13039	0.52	0.15	0.26	0.42	0.63
A_double_dot	2544	0.22	0.15	0.42	0.51	0.82
A_oilseeds_dot	7166	0.50	0.22	0.32	0.56	0.58
A_othercrops_dot	26763	0.46	0.07	0.17	0.25	0.67
A_sorghum_dot	2374	0.52	0.21	0.21	0.30	0.72
A_total_cropped_dot	35722	0.82	0.47	0.04	0.09	0.39
A_wheat_dot	16237	0.60	0.29	0.26	0.40	0.65
H_barley_dot	13039	0.45	0.23	0.27	0.32	0.85
H_oilseeds_dot	7169	0.39	0.14	0.28	0.34	0.83
H_sorghum_dot	2373	0.37	0.22	0.34	0.37	0.93
H_wheat_dot	16242	0.53	0.29	0.22	0.26	0.82
H_wool_dot	23485	0.33	0.11	0.20	0.24	0.84
Q_barley_dot	9368	0.24	0.04	0.23	0.24	0.95
Q_beef_dot	24723	0.22	0.11	0.23	0.25	0.95
Q_lamb_dot	10517	0.41	0.18	0.36	0.43	0.83
Q_oilseeds_dot	1731	0.17	-0.03	0.22	0.25	0.90
Q_othercrops	40125	0.54	0.28	0.71	1.27	0.56
Q_sheep_dot	20645	0.09	0.05	0.43	0.45	0.97
Q_sorghum_dot	1516	0.17	0.01	0.40	0.41	0.96
Q_wheat_dot	11100	0.19	-0.01	0.08	0.09	0.93
Q_wool_dot	8277	0.23	0.14	0.16	0.18	0.91
R_barley_dot	9457	0.46	0.37	0.15	0.16	0.90
R_beef_dot	25527	0.49	0.28	0.17	0.21	0.80
R_lamb_dot	10927	0.73	0.70	0.16	0.17	0.94

R_oilseeds_dot	5642	0.73	0.73	0.11	0.11	0.93
R_sheep_dot	21257	0.39	0.32	0.21	0.23	0.91
R_sorghum_dot	1530	0.11	0.15	0.19	0.18	1.04
R_wheat_dot	13290	0.39	0.29	0.14	0.16	0.89
R_wool_dot	23478	0.61	0.42	0.16	0.19	0.84
S_beef_births_dot	24259	0.32	0.14	0.21	0.24	0.87
S_beef_deaths_dot	22645	0.21	-0.01	0.56	0.58	0.95
S_sheep_births_dot	21863	0.39	0.18	0.24	0.29	0.83
S_sheep_deaths_dot	22885	0.19	0.09	0.60	0.63	0.95
V_chem	40125	0.73	0.50	0.49	0.85	0.58
V_fert	40125	0.71	0.49	0.46	0.74	0.63
V_fuel	40125	0.76	0.61	0.31	0.43	0.73
V_othermat	40125	0.55	0.44	0.48	0.62	0.78
V_serv	40125	0.80	0.67	0.29	0.40	0.73
V_shearing	40125	0.86	0.82	0.30	0.41	0.73
Average	18914	0.45	0.27	0.28	0.36	0.81

Table 16: Cross-validated regression performance (binary targets), xgboost stack versus least squares

	Sample	xgboost accuracy	LS accuracy	accuracy xgboost / LS
D_barley	40125	0.89	0.83	1.07
D_double	40125	0.94	0.94	1.00
D_oilseeds	40125	0.91	0.86	1.06
D_othercrops	40125	0.85	0.80	1.06
D_sorghum	40125	0.97	0.96	1.01
D_wheat	40125	0.92	0.89	1.03
Z_conditions_4	34995	0.88	0.85	1.04

Table 17: xgboost feature importance score, by feature type (stack 1)

	Climate (rain)	Climate (temp.)	Prices	Stocks	Capital	Location	Other Controls
A_barley_dot	9.4	2.8	0.3	14.0	43.5	12.1	18.0
A_double_dot	8.5	10.0	2.7	11.0	31.7	4.2	31.9
A_oilseeds_dot	9.0	4.9	2.4	16.7	27.0	8.8	31.3
A_othercrops_dot	11.1	4.8	2.3	5.6	46.1	6.8	23.3
A_sorghum_dot	10.4	3.1	2.4	9.3	44.9	7.4	22.6
A_total_cropped_dot	1.6	0.4	0.1	5.6	23.5	1.3	67.6
A_wheat_dot	8.6	3.5	1.0	9.2	35.9	5.8	36.0
D_barley	7.7	3.7	1.1	44.3	5.9	11.9	25.4
D_double	23.8	9.9	4.6	5.3	14.7	10.1	31.7
D_oilseeds	5.9	4.9	2.7	14.3	15.7	16.6	39.8
D_othercrops	7.9	4.8	1.3	14.6	31.4	8.7	31.3
D_sorghum	10.6	5.0	0.3	22.4	5.0	28.8	27.8
D_wheat	6.6	3.0	0.3	38.2	7.6	4.5	39.7
H_barley_dot	49.1	5.4	7.8	3.8	6.8	4.7	22.4
H_oilseeds_dot	52.2	6.9	5.6	3.4	5.9	6.0	20.1
H_sorghum_dot	24.9	17.0	2.5	2.0	5.8	6.0	41.8
H_wheat_dot	51.4	4.5	5.7	2.3	5.6	6.5	24.0
H_wool_dot	14.4	7.0	2.4	13.3	10.5	12.3	40.1
Q_barley_dot	21.1	5.3	6.7	21.5	11.0	19.4	15.0
Q_beef_dot	12.0	6.6	0.6	43.0	5.0	18.0	14.7
Q_lamb_dot	7.5	6.0	2.9	34.2	7.9	11.0	30.5
Q_oilseeds_dot	25.5	10.3	6.3	24.5	10.5	9.4	13.6
Q_othercrops	3.1	2.1	3.6	7.9	27.8	6.2	49.3
Q_sheep_dot	28.9	7.6	5.9	17.8	8.8	7.6	23.3
Q_sorghum_dot	30.5	7.6	9.8	27.2	9.5	4.6	10.7
Q_wheat_dot	9.8	10.4	8.0	38.7	11.6	14.0	7.6

ABARES working paper

Q_wool_dot	8.9	6.4	5.6	60.2	7.0	1.7	10.2
R_barley_dot	13.6	3.2	61.3	1.4	2.7	7.3	10.4
R_beef_dot	6.9	11.7	26.6	6.3	6.0	9.6	32.9
R_lamb_dot	2.5	2.5	77.7	4.4	2.5	6.2	4.2
R_oilseeds_dot	6.1	5.1	10.3	0.7	2.1	16.0	59.7
R_sheep_dot	3.2	1.7	78.2	4.1	2.6	4.6	5.6
R_sorghum_dot	31.9	14.7	20.5	4.7	7.0	4.9	16.2
R_wheat_dot	20.9	3.5	48.3	3.1	4.1	8.4	11.7
R_wool_dot	13.8	11.3	50.6	5.5	3.3	6.4	9.1
S_beef_births_dot	12.3	11.1	6.7	30.4	6.6	11.4	21.4
S_beef_deaths_dot	32.7	17.0	2.8	7.7	9.4	5.6	24.9
S_sheep_births_dot	12.4	3.5	7.6	19.0	7.2	15.8	34.5
S_sheep_deaths_dot	26.0	6.1	11.2	11.1	6.3	6.7	32.7
V_chem	3.4	1.4	4.5	3.8	51.0	4.1	31.8
V_fert	4.1	3.4	2.0	4.9	46.5	16.2	22.9
V_fuel	1.6	1.5	2.2	20.6	56.6	3.9	13.7
V_othermat	7.8	4.1	4.1	55.2	7.4	4.1	17.3
V_serv	2.6	1.7	1.3	45.3	31.3	3.5	14.3
V_shearing	1.0	0.6	1.2	92.4	1.3	0.9	2.7
Z_conditions_4	52.0	9.3	15.1	2.2	4.3	11.1	6.1

Table 18: Most important relationships between specific model targets (columns) and specific climate variables (rows)

Climate variable	Target 1	Target 2	Target 3	Target 4
<i>W_winter_rain</i>	H_wheat_dot	H_barley_dot	H_oilseeds_dot	S_beef_deaths_dot
<i>W_aut_tmax_L1</i>	H_wool_dot	Q_oilseeds_dot	S_beef_births_dot	S_beef_deaths_dot
<i>W_sum_rain</i>	R_sorghum_dot	Q_sorghum_dot	H_sorghum_dot	Q_oilseeds_dot
<i>W_winter_gdd</i>	R_wool_dot	S_beef_deaths_dot	D_double	H_barley_dot
<i>W_win_tmin</i>	D_double	Q_sheep_dot	Q_wheat_dot	S_beef_deaths_dot
<i>W_summer_gdd</i>	H_sorghum_dot	Q_oilseeds_dot	R_sorghum_dot	A_double_dot
<i>W_aut_moist</i>	Q_sheep_dot	Q_oilseeds_dot	S_beef_deaths_dot	D_double
<i>W_aut_tmin</i>	H_sorghum_dot	S_beef_deaths_dot	Q_sorghum_dot	D_double
<i>W_FY_rain</i>	Q_beef_dot	Q_sorghum_dot	H_sorghum_dot	R_sorghum_dot
<i>W_sum_tmax_L1</i>	R_sorghum_dot	Q_beef_dot	R_wheat_dot	H_sorghum_dot
<i>W_spr_rain</i>	R_sorghum_dot	Q_sorghum_dot	S_sheep_deaths_dot	D_double
<i>W_aut_moist_L1</i>	R_sorghum_dot	S_beef_deaths_dot	Q_barley_dot	Q_sheep_dot
<i>W_aut_rain</i>	Q_sorghum_dot	R_sorghum_dot	Q_sheep_dot	S_beef_deaths_dot
<i>W_aut_gni_L1</i>	R_sorghum_dot	S_beef_deaths_dot	H_oilseeds_dot	H_sorghum_dot
<i>W_aut_tmax</i>	Q_wheat_dot	S_beef_deaths_dot	Q_sheep_dot	H_oilseeds_dot
<i>W_FY_rain_L1</i>	Q_barley_dot	D_double	S_sheep_deaths_dot	R_barley_dot
<i>W_FY_moist_L2</i>	Q_oilseeds_dot	S_beef_deaths_dot	Q_sorghum_dot	A_barley_dot
<i>W_sum_tmax</i>	Q_barley_dot	Q_sorghum_dot	R_sorghum_dot	R_barley_dot
<i>W_win_moist</i>	R_sorghum_dot	D_double	S_beef_deaths_dot	S_sheep_deaths_dot
<i>W_FY_rain_L2</i>	H_sorghum_dot	H_oilseeds_dot	S_beef_deaths_dot	R_sorghum_dot

Figure 9 shows the average model responses to growing season rainfall (*W_winter_rain*) and lagged autumn average maximum temperature (*W_aut_tmax_L1*). Note that farm level marginal responses will differ from these averages (depending on farm type / location etc.). As would be expected, wheat yield is increasing in winter rainfall. Beef birth (death) rates are generally increasing (decreasing) in winter rainfall, and decreasing (increasing) in autumn temperature. Wheat area planted is on average increasing in autumn maximum temperatures, up to around 25 degrees, and decreasing thereafter (although the effects of climate on crop areas are relatively small in comparison with yields).

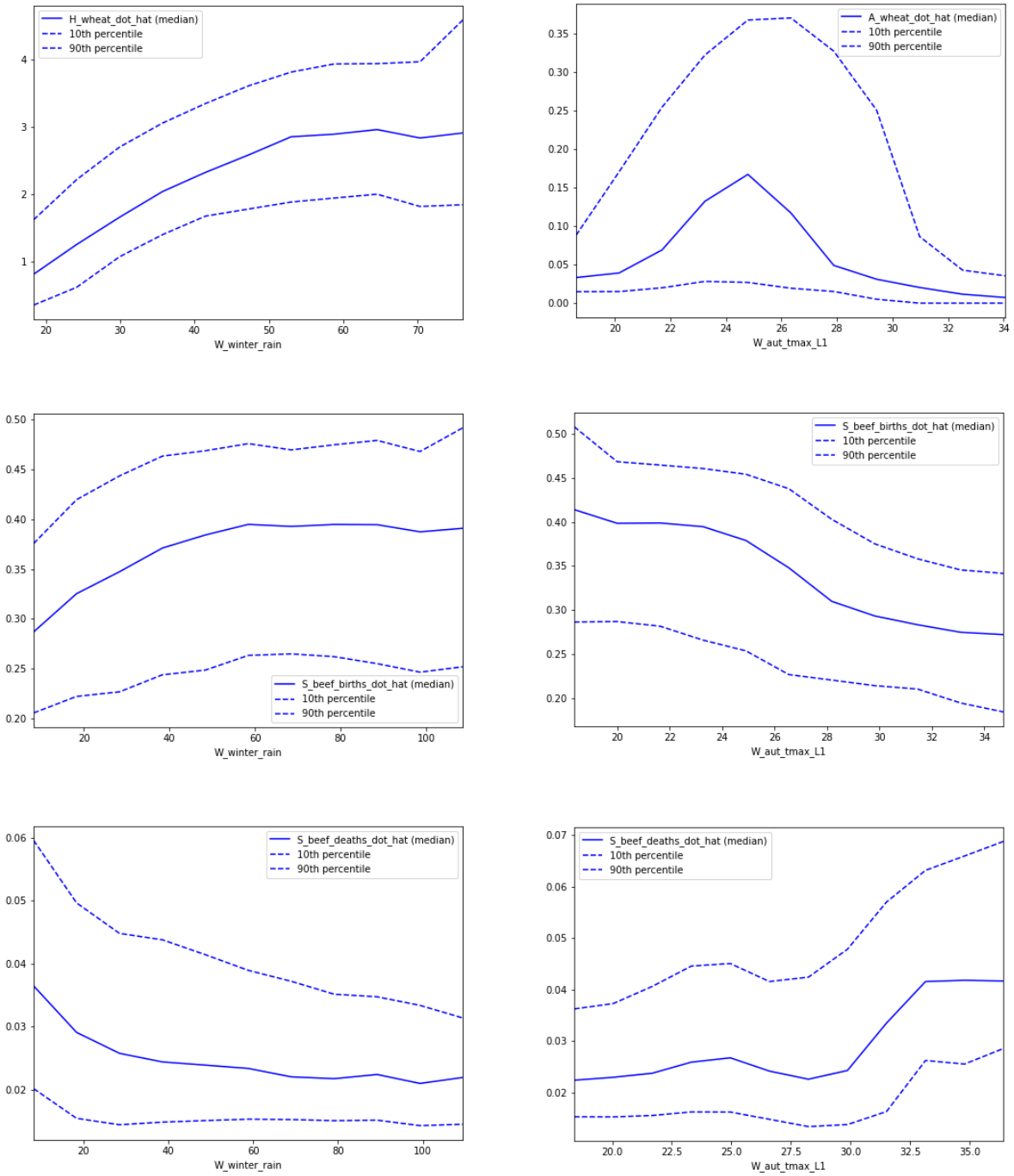


Figure 9: Average response of selected target variables to W_{winter_rain} and $W_{aut_tmax_L1}$

9 Appendix E: Validation results

Table 19: Farm level cross-validated performance of the *baseline* scenario

	Farm level	Regional	National
Wheat gross receipts, R_wheat	0.77	0.98	0.96
Barley gross receipts, R_barley	0.63	0.95	0.92
Oilseeds gross receipts, R_oilseeds	0.56	0.93	0.97
Sorghum gross receipts, R_sorghum	0.57	0.95	0.67
Other crop gross receipts, R_othercrops	0.55	0.82	0.56
Beef cattle receipts, R_beef	0.86	0.96	0.99
Sheep gross receipts, R_sheep	0.60	0.94	0.98
Wool gross receipts, R_wool	0.90	0.99	1.00
Prime lamb net receipts, R_lamb	0.76	0.96	0.99
Expenditure on fertiliser, C_fert	0.71	0.97	0.96
Expenditure on fuel, oil and grease, C_fuel	0.76	0.94	0.98
Expenditure on crop and pasture chemicals, C_chem	0.71	0.97	0.98
Other materials, C_othermat	0.56	0.83	0.52
Total expenditure on services and labour, C_serv	0.80	0.93	0.94
Hired labour shearing, C_shearing	0.85	0.97	0.87
Wheat area sown, A_wheat	0.81	0.98	0.96
Barley area sown, A_barley	0.60	0.94	0.93
Oilseeds area sown, A_oilseeds	0.46	0.92	0.96
Sorghum area sown, A_sorghum	0.36	0.93	0.87
Wheat production (harvest), H_wheat	0.76	0.97	0.98
Barley production (harvest), H_barley	0.58	0.94	0.95
Oilseeds production (harvest), H_oilseeds	0.54	0.93	0.98
Sorghum production (harvest), H_sorghum	0.41	0.94	0.71
Beef number sold, Q_beef	0.87	0.96	0.96
Sheep number sold, Q_sheep	0.67	0.89	0.86

Prime lamb number sold, Q_lamb	0.75	0.96	0.97
Wool produced, Q_wool	0.91	0.99	0.99
Beef cattle births, S_beef_births	0.87	0.95	0.98
Beef cattle deaths, S_beef_deaths	0.53	0.84	0.82
Sheep births, S_sheep_births	0.86	0.97	0.97
Sheep deaths, S_sheep_deaths	0.52	0.86	0.88

Table 20: Regional level R^2 of profit measures under the *baseline* scenario

	Sample size	Farm business profit	Farm cash income
NSW: Riverina	2428	0.92	0.89
VIC: Southern and Eastern Victoria	2667	0.92	0.92
VIC: Wimmera	1566	0.89	0.83
NSW: Central West	2012	0.87	0.83
VIC: Central North	1782	0.87	0.89
NSW: Tablelands	1896	0.84	0.86
SA: Murray Lands & Yorke Peninsula	1664	0.83	0.82
WA: Central and South Wheat Belt	1887	0.83	0.88
SA: Eyre Peninsula	1046	0.79	0.81
NT: Alice Springs Districts	338	0.78	0.84
NT: Victoria River District	336	0.76	0.71
SA: South East	1694	0.73	0.60
QLD: Darling Downs and Highlands	2561	0.71	0.75
VIC: Mallee	1297	0.69	0.65
QLD: Central North	931	0.68	0.66
NT: Barkly Tablelands	171	0.66	0.53
TAS: Tasmania	1812	0.65	0.70
WA: North and East Wheat Belt	1099	0.62	0.51
NSW: Far West	798	0.60	0.67
QLD: Eastern Darling Downs	1479	0.60	0.69
NSW: North West Slopes and Plains	1920	0.60	0.40

ABARES working paper

SA: North Pastoral	635	0.58	0.69
WA: The Kimberly	199	0.57	0.50
QLD: West and South West	838	0.55	0.50
NT: Top End Darwin and NT Gulf	202	0.53	0.56
QLD: South Coastal	1519	0.51	0.65
QLD: Cape York and the Gulf	368	0.49	0.30
NSW: Coastal	883	0.45	0.48
WA: South West Coastal	1103	0.44	0.25
QLD: Charleville - Longreach	936	0.42	0.70
QLD: North Coastal	579	0.32	-0.03
WA: Pilbara and the Central	357	-0.07	0.43

10 References

- Adeyinka, AA et al. 2016, 'The viability of weather-index insurance in managing drought risk in rural australia', *International Journal of Rural Management*, vol. 12, no. 2, pp. 125–142.
- Antle, JM 1984, 'The structure of uS agricultural technology, 1910–78', *American journal of agricultural economics*, vol. 66, no. 4, pp. 414–421.
- Antle, JM & Capalbo, SM 2001, 'Econometric-process models for integrated assessment of agricultural production systems', *American Journal of Agricultural Economics*, vol. 83, no. 2, pp. 389–401.
- Ball, VE et al. 1997, 'Cap reform: Modelling supply response subject to the land', *Agricultural Economics*, vol. 17, pp. 277–288.
- Bell, R et al. 2007, *The responsiveness of Australian farm performance to changes in irrigation water use and trade*, Productivity Commission.
- Bischl, B et al. 2017, 'MlrMBO: A modular framework for model-based optimization of expensive black-box functions', *arXiv preprint arXiv:1703.03373*.
- Bouchet, F, Orden, D, & Norton, GW 1989, 'Sources of growth in french agriculture', *American Journal of Agricultural Economics*, vol. 71, no. 2, pp. 280–293.
- Cai, W & Cowan, T 2013, 'Southeast australia autumn rainfall reduction: A climate-change-induced poleward shift of ocean–atmosphere circulation', *Journal of Climate*, vol. 26, no. 1, pp. 189–205.
- Carter, JO et al. 2000, 'Aussie GRASS: Australian grassland and rangeland assessment by spatial simulation', in *Applications of seasonal climate forecasting in agricultural and natural ecosystems*, Springer, pp.329–349.
- Chen, T & Guestrin, C 2016, 'Xgboost: A scalable tree boosting system', in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, ACM, pp.785–794.
- Einav, L & Levin, J 2014, 'Economics in the age of big data', *Science*, vol. 346, no. 6210, p. 1243089.
- Fisher, BS & Wall, CA 1990, 'Supply response in the australian sheep industry: A profit function approach', *Australian Journal of Agricultural Economics*, vol. 34, no. 2, pp. 147–166.
- Friedman, JH 2002, 'Stochastic gradient boosting', *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378.
- Frost, A, Ramchurn, A, & Smith, A 2016, *The bureau's operational aWRA landscape model*, Bureau of Meteorology.
- Hochman, Z, Gobbett, DL, & Horan, H 2017, 'Climate trends account for stalled wheat yields in australia since 1990', *Global Change Biology*, vol. 23, no. 5, pp. 2071–2081.
- Hughes, N, Lawson, K, & Valle, H 2016, *Farm performance and climate: climate adjusted productivity on broadacre cropping farms*, Australian Bureau of Agricultural; Resource Economics; Sciences.

- Jeffrey, SJ et al. 2001, 'Using spatial interpolation to construct a comprehensive archive of Australian climate data', *Environmental Modelling & Software*, vol. 16, no. 4, pp. 309–330.
- Keogh, M 2012, 'Including risk in enterprise decisions in Australia's riskiest businesses', in *56th annual conference of the Australian agricultural and resource economics society, Perth, Australia*, pp.8–10.
- Kokic, P et al. 1993, *Australian broadacre agriculture: forecasting supply at the farm level*.
- Kokic, P et al. 2007, 'From rainfall to farm incomes: transforming advice for Australian drought policy. I. Development and testing of a bioeconomic modelling system', *Crop and Pasture Science*, vol. 58, no. 10, pp. 993–1003.
- Land use of Australia, 2010-11* 2016, Australian Bureau of Agricultural; Resource Economics; Sciences.
- Lobell, DB & Burke, MB 2010, 'On the use of statistical models to predict crop yield responses to climate change', *Agricultural and Forest Meteorology*, vol. 150, no. 11, pp. 1443–1452.
- Marschak, J & Andrews, WH 1944, 'Random simultaneous equations and the theory of production', *Econometrica, Journal of the Econometric Society*, pp. 143–205.
- Mueller, N et al. 2016, 'Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia', *Remote Sensing of Environment*, vol. 174, pp. 341–352.
- Mundlak, Y 2001, 'Production and supply', in BL Gardner & G C. Rausser (eds.), *Handbook of agricultural economics: Volume 1A agricultural production*, Elsevier, Elsevier, pp.4–85.
- Nelson, R, Kokic, P, & Meinke, H 2007, 'From rainfall to farm incomes: transforming advice for Australian drought policy. II. Forecasting farm incomes', *Crop and Pasture Science*, vol. 58, no. 10, pp. 1004–1012.
- Nelson, R et al. 2010, 'The vulnerability of Australian rural communities to climate variability and change: Part II: Integrating impacts with adaptive capacity', *Environmental Science & Policy*, vol. 13, no. 1, pp. 18–27.
- Peel, MC, McMahon, TA, & Finlayson, BL 2004, 'Continental differences in the variability of annual runoff—update and reassessment', *Journal of Hydrology*, vol. 295, no. 1, pp. 185–197.
- Potgieter, AB, Hammer, GL, & Butler, D 2002, 'Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO', *Crop and Pasture Science*, vol. 53, no. 1, pp. 77–89.
- Raupach, M et al. 2009, 'Australian water availability project: CSIRO marine and atmospheric research component: Final report for phase 3', *Centre for Australian weather and climate research*, vol. 67.
- Segerson, K & Dixon, BL 1999, *Climate change and agriculture: The role of farmer adaptation*, Cambridge University Press, Cambridge, United Kingdom; New York, NY, USA.
- Shumway, CR 1983, 'Supply, demand, and technology in a multiproduct industry: Texas field crops', *American Journal of Agricultural Economics*, vol. 65, no. 4, pp. 748–760.
- Shumway, CR 1995, 'Recent duality contributions in production economics', *Journal of Agricultural and Resource Economics*, pp. 178–194.

Spyromitros-Xioufis, E et al. 2012, 'Multi-label classification methods for multi-target regression', *arXiv preprint arXiv*, vol. 1211.

Stoorvogel, J et al. 2004, 'The tradeoff analysis model: Integrated bio-physical and economic modeling of agricultural production systems', *Agricultural systems*, vol. 80, no. 1, pp. 43–66.

Varian, HR 2014, 'Big data: New tricks for econometrics', *The Journal of Economic Perspectives*, pp. 3–27.

Xayavong, V, Islam, N, & Salim, R 2011, 'Estimating production response of broadacre farms in western australia: The nexus of empirics and economics revisited', *Economic Analysis and Policy*, vol. 41, no. 3, pp. 217–232.

Zhao, S et al. 2012, 'Measuring productivity of the australian broadacre and dairy industries: Concepts, methodology and data', *Productivity Growth in Agriculture: An International Perspective*, pp. 73–108.