Modelling Rural Land Use in New Zealand

- A Discrete Choice Perspective

Levente Timar*

AARES 2011

Methods, Data and Results Sections

WORKING DRAFT. ALL FINDINGS AND METHODS ARE PRELIMINARY. PLEASE DO NOT CITE OR CIRCULATE.

* Levente is a Research Economist at Motu Economic and Public Policy Research

Author contact details

Levente Timar
Motu Economic and Public Policy Research
levente.timar@motu.org.nz

Motu Economic and Public Policy Research

PO Box 24390
Wellington
New Zealand

Email	info@motu.org.nz
Telephone	+64 4 9394250
Website	www.motu.org.nz

1. Methods

The economic model I use to describe landowners' land use decisions is a standard discrete choice random utility maximization model.[1] Land is of heterogeneous quality, and suitability for the various uses depends on (multiple dimensions of) quality. Therefore, at any given time, potential benefits derived from each parcel vary by use. As economic conditions change, production technologies advance and the farmer accumulates human capital, the relative desirability of land use alternatives may change on any parcel. When the top-ranked alternative changes due to these forces, the farmer converts the parcel to a different use. The observed pattern of land use therefore represents a snapshot of outcomes from a dynamic process.

I model land use choices based on such a snapshot, a single cross section of data, by assuming that farmers are profit maximisers.[2] The desirability of a land use alternative is thus directly related to its profitability, and farmers follow the decision rule

$$R_{ij} > R_{ik} \;\; \forall \, j \neq k,$$

where R_{ij} represents instantaneous net returns to land use alternative j on parcel i. By choosing a land use, farmers maximize the present value of the future stream of net returns to their piece of land. Land use decisions are inherently more complex and dynamic than this: past land use choices, price expectations, option values, risk attitudes and other preferences may differ across the population, and they all play an important role in determining outcomes. Conversion costs between the different uses and financial constraints introduce another layer of complexity. Modelling these issues in a discrete choice framework is not at all a trivial task and my analysis, like most others, abstracts from them.

Adopting a static framework implies that land use responses to changing input and output prices or other macro-economic variables are not represented in the data. It would be especially important to understand such effects from the perspective of policy analysis, but the framework is ill-adapted to provide insight into them. Despite these limitations, a single-period discrete choice model can still be useful in investigating how factors that vary within the cross section affect land use decisions. There exist geophysical and climatic limitations on most land use types. Also, the vast majority of parcels stay in the same use from one year to the next, and overall land use shares tend to change only slowly over time. Therefore, a good understanding of the location-specific determinants of land use goes a long way toward explaining farmers'

[1] The terms 'farmer' and 'landowner' are used interchangeably to mean 'decision maker' throughout this paper, although in some cases they may not be the same person.
[2] Utility maximisation and profit maximisation are equivalent in this framework, the difference is purely semantic.

choices. The static model provides a characterisation of the role of these factors. Its results are, of course, conditional on the levels of unobserved variables, including those that vary over time only.

Focusing on a cross section of data also allows the researcher to skirt issues of endogeneity arising in dynamic settings. The profitability of plantation forestry on a certain plot, for example, depends on the location of sawmills and other relevant infrastructure. New facilities can be built over time, however, and their chosen sites would presumably be affected by land use outcomes, that is, the profitability of the various alternatives. A cross sectional analysis is not subject to these types of endogeneity problems because sawmill locations are fixed and unaffected by the modelled land use choices.

Returning to the farmer's decision rule, not all factors that affect profitability can be observed. It will prove convenient to break up R_{ij} into two additive terms: a component (V_{ij}) that is observed, and a component (ε_{ij}) that is known to the farmer, but is not observed by the researcher. This term is assumed to have a known distribution, and is treated as a random error. Because of the error, ex ante we can only make probabilistic statements about the choice of land use. Denoting by P_{ij} the probability that parcel i will be found in land use j,

$$P_{ij} = \text{Prob}(R_{ij} > R_{ik}) \;\; \forall\, j \neq k$$

$$P_{ij} = \text{Prob}(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik}) \;\; \forall\, j \neq k$$

$$P_{ij} = \text{Prob}(\varepsilon_{ij} - \varepsilon_{ik} > V_{ik} - V_{ij}) \;\; \forall\, j \neq k$$

The exact functional form this probability takes is determined by the distributional assumptions placed on the error term, which differ by discrete choice model. Regardless, the parameters of the model are estimated by maximizing the sample log-likelihood (or simulated log-likelihood) function

$$\text{LL}(\beta) = \textstyle\sum_i \sum_j y_{ij} \ln P_{ij},$$

where β denotes the parameters, and y_{ij} is an indicator variable whose value is equal to one if the observed land use choice on parcel i is j, and zero otherwise. The estimated parameters will be those that, given the assumed structure of the model, are most likely to have resulted in the sample data.

I specify the observed component of profitability as a linear function of a vector of explanatory variables (X_i) specific to the parcel, and a coefficient vector to be estimated (β_j) that varies over the land use alternatives.

$$R_{ij} = \beta_j{}'X_i + \varepsilon_{ij}$$

If one of the elements of X_i equals one, V_{ij} will include a constant term for each land use. These choice-specific constants play a role similar to that of the constant in an ordinary least squares estimation: they non-parametrically control for all factors (unmeasured or unobservable) that are omitted from the specification, but affect the attractiveness of the land use.

Explanatory variables include factors based on terrain, soil characteristics and climate. In some specifications, X_i also includes distance-based variables derived from the spatial location of the parcel, and regional profitability in the various land uses. All of the independent variables exhibit some level of spatial correlation, and the same can likely be said of omitted variables of importance. These may lead to spatial autocorrelation in the error term, violating standard assumptions on its distribution. I employ two common strategies (the spatially derived explanatory variables and systematic subsampling) to partially address the issue.

2. Data

My dependent variable is the choice of land use in 2002 from a set of four options: dairy farming, sheep or beef farming, plantation forestry and scrub (essentially, uncultivated or non-productive). These are the four primary rural uses in New Zealand; they collectively cover almost half of the land area of the country. All other uses, including horticultural, urban and Department of Conservation (DOC) land are exogenous to the model. Land use decisions regarding public and urban uses are fundamentally different from those taking place on private rural plots, so these can fairly safely be considered exogenous. Other rural uses are not included because they generally have less information available about them, and they tend to be relatively unimportant compared to the modelled uses in New Zealand (though horticulture has become increasingly significant over time).

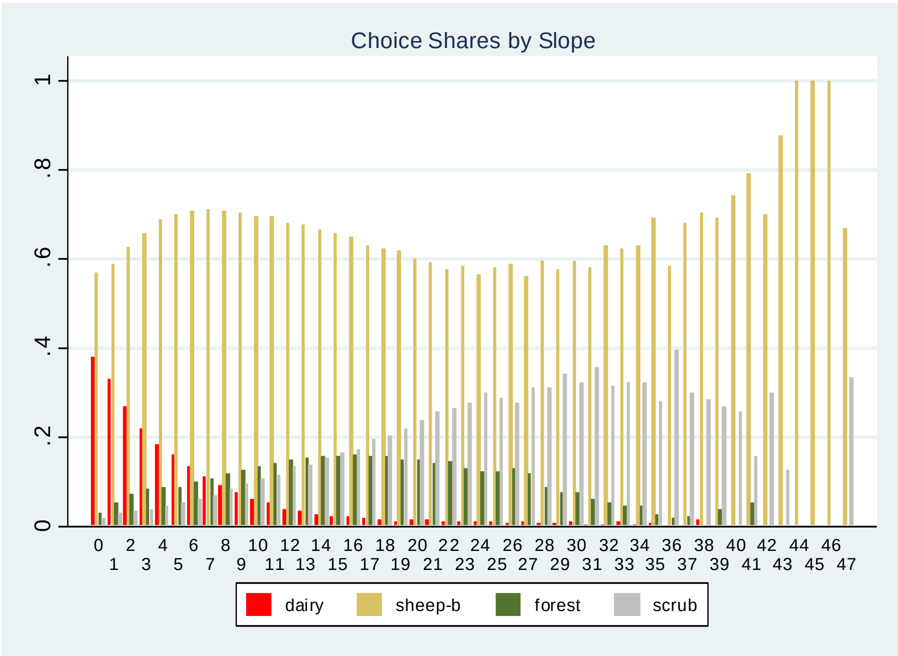
The land use data reflects the revealed preferences of landowners. Information from several sources was combined by Motu Research to establish land use on privately held parcels in 2002. These include remote sensing satellite observations of land cover (Ministry for the Environment, 2005) and land use information derived from AgriBase, a large database of rural properties (AsureQuality, 2005). Maps of Department of Conservation land (Department of Conservation, 2005) and ownership (Landcare Research, 2008) were also utilized in the construction of this layer.

Geographic land use data of complete national coverage is available for a single year, 2002, only. The primary difference between land cover and land use is the ability to separate dairy farms from sheep and beef farms in the land use dataset – both farm types show up as pasture in satellite imagery. This ability is paramount because dairy farms differ from other livestock farms in important ways, as discussed above.

I model the choice of land use as a function of variables that are expected to affect the relative profitability of the various uses in the choice set. These include geophysical and climatic factors, spatial variables specific to the location of the plot, as well as socio-economic factors controlling for the effects of land governance and mean regional profitability. Each observation represents a 25-hectare (500m-by-500m) pixel of privately owned land. Summary statistics for these candidate explanatory variables, by observed land use, are shown in Table 1. The variables are described in the flowing paragraphs.

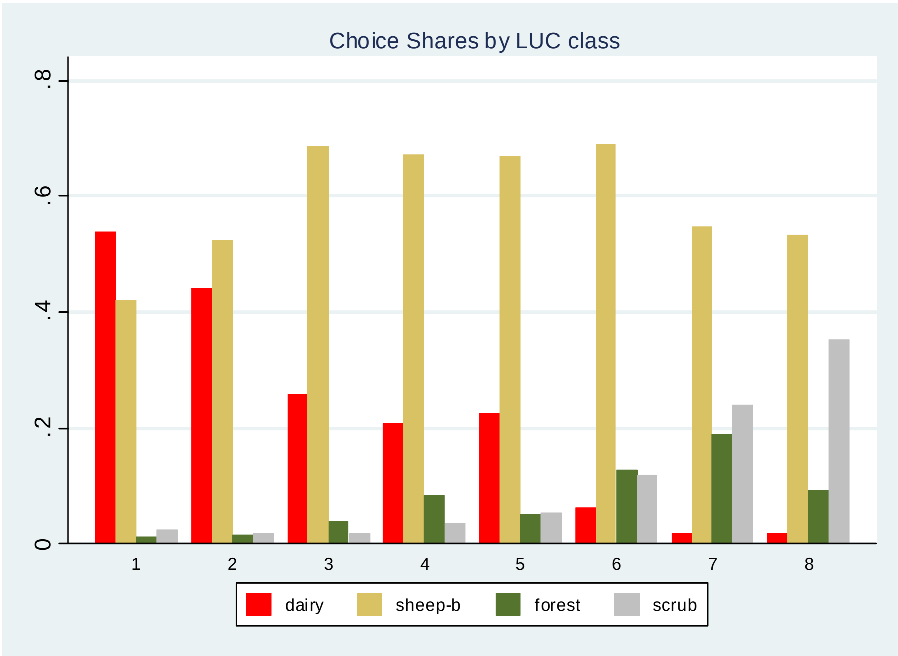
Terrain obviously affects the uses to which a piece of land can be put. In New Zealand, steep slopes sometimes make more intensive uses infeasible or uneconomic. Slope may affect profitability through productivity as well as through harvest costs. Dairy farming, for example, is found almost exclusively in flat regions of the country. Accordingly, one of the explanatory variables considered is slope (in degrees). Figure 1 depicts the observed relationship between the land use alternatives and slope, where shares within each slope class add up to one. Plains (areas with slopes of less than 3 degrees) make up more than a quarter of the sample land area, while areas of extreme slope (more than 25 degrees) are rare. The slope layer is part of the Land Environments of New Zealand dataset (Landcare Research, 2004).

Figure 1



Land Use Capability (LUC) class is a summary measure of land quality, constituting part of the New Zealand Land Resource Inventory (Landcare Research, 2002). It is not a cardinal measure of land quality, but merely an ordinal indicator: an "assessment of the land's capability for use, while taking into account its physical limitations and its versatility for sustained production" (Lynn et. al, 2009). The assessment is based on an interpretation of physical information compiled from a field assessment of rock types, soils, landforms and slopes, erosion types and severities and vegetation cover. This is sometimes supplemented with information on climate, flood risk, erosion history and the effects of past practices. Land is classified into eight classes, with limitations to use increasing from LUC class 1 to LUC class 8. In general, classes 1-4 are suitable for multiple land uses including arable cropping. Classes 5-7 are unsuitable for arable cropping, but are suitable for pastoral grazing and forestry use. Limitations to use reach a maximum with class 8, which is best managed as conservation land. Higher-intensity land uses therefore generally require land of a lower LUC class. The observed distribution of land use types within LUC classes confirms this relationship. This is shown in figure 2. Nearly 90% of the land in the sample is in LUC classes 3, 4, 6 and 7, while the combined share of classes 1, 5 and 8 is only 5%. Although slope is one of the several geophysical inputs into the LUC classification, Todd and Kerr (2009) found that slope has explanatory power beyond its effect through LUC class.

Figure 2



The land use capability classification is primarily about limitations to use. Another, in many cases more relevant, measure of land quality is its biological productivity – productivity is what ultimately leads to economic returns in agricultural and forestry uses. I employ two "Storie Index"-type variables: one for pastoral farming, and another one for exotic forestry (Baisden, 2006a). Both variables are based on factors representing climate and soil properties that affect plant growth (namely growing degree days, soil moisture deficit and soil particle size at the location), and both were calibrated via logarithmic regression to satellite observations of net primary productivity (NPP) in New Zealand (Baisden, 2006b).[3] The productivity indices are measured in kilograms of biomass per square metre per year. They have different absolute values (that is, the amount of grass biomass that can be grown on a piece of land is different from the amount of wood biomass that can be grown on the same piece of land), but pasture productivity and exotic forest productivity are highly correlated with a correlation coefficient of $r = 0.973$. Baisden (2006b) found that the relationship between NPP and the calibrated pasture productivity index was much stronger than that between NPP and the calibrated exotic forestry productivity index.

It is important to discuss the relationship between LUC class and pasture productivity (or alternatively, forest productivity). Both are proxy measures of land quality determined by

[3] Satellite observations of NPP measure productivity under current land use. The calibration procedure is necessary for estimating productivity under alternative uses.

underlying geophysical and climatic factors. Although they are derived differently, one might naively expect to find some correlation between the two. On the contrary (and perhaps surprisingly), they are nearly orthogonal. Land use capability is about limitations to use and not about production per se. It does not necessarily relate to the amount of grass or wood that a piece of land will grow.[4] Additional insight into this claim can be gained by examining the geographic distribution of the two variables. While LUC classes are fairly evenly distributed across larger regions, pasture productivity exhibits a strong north-south gradient. This suggests that the former variable is mostly driven by local geophysical differences in rock and soil type, and the latter by climatic conditions: the LUC classification is affected by climate only when it imposes a significant limitation on use, while the calibration procedure for the productivity variables masks large-scale variation in soil quality. They therefore capture different dimensions of "land quality", which explains the low correlation between them.

The next variable provides information on the ownership and governance structure of the land. It is an indicator variable whose value equals zero for land under private ownership with no restrictions, and one for land under Maori status. Maori land is subject to a range of unique restrictions and protections, and these restrictions can create difficulties for owners wishing to develop or use their land. Thus the variable represents limitations on use and governance – it is not a socio-economic indicator for the owner. The Audit Office (2004) estimates that nearly 40 percent of Maori land is underdeveloped. Ownership data was compiled by Landcare Research (2008) using information on governance from the Maori Land Information Base.

Rural support networks, the local transport and financial infrastructure, region-specific regulations, and individuals' socio-economic characteristics may all play a significant role in determining profitability and hence the choice of land use. These factors are often unobserved or difficult to quantify, and they often give rise to spatial autocorrelation. Systematically accommodating spatial effects (whether in the explanatory variables or in the error term) in multinomial discrete choice models is challenging, and, at the current stage of the research, I opted for a workaround that is often exercised in the literature: including various spatially derived variables in the utility specification. The effectiveness of this strategy will, of course, depend on how well these variables capture the underlying spatial phenomenon. Processing

[4] Another way of saying this is that there is no inherent reason that land without any limitations should grow more grass than land with a limitation that does not affect grass growth, but makes it less than ideal for tillage. This argument is precisely why the productivity indices were developed in the first place (Baisden, 2010).

facility locations were acquired from the Ministry of Agriculture and Forestry (2010), and distances were calculated in ArcGIS.

Distances to the nearest dairy, meat or wood processing facilities are clearly related to the profitability of the relevant land use type, and may also partially control for other spatial influences. A region with a well-developed support infrastructure for dairying, for example, is likely to have more dairy processing facilities thereby reducing the average distance to the nearest plant. As already discussed, these variables are endogenous in the sense that facilities will tend to be built near farms that supply them with raw materials. However, at any point in time, plant locations are given, and the distance variables can be considered exogenous to the land use decision.

Another spatial variable, distance to the nearest supermarket, is used as a proxy for access to local produce and factor markets and other amenities typically provided by population centres. Distance calculations were carried out in ArcGIS using supermarket locations (Zenbu, 2011). All distances are measured in straight-line kilometres.

Finally, I also employ regional mean profits (for forestry, revenues) in the various uses as explanatory variables. Regional variation in profits is not helpful for identifying land use reactions to changing profits over time, but it may capture other region-specific unobserved factors. That is, if differences in regional mean profits are not driven exclusively by differences in regional mean "land quality", then the variable will have additional explanatory power.[5] The profit variables are measured in thousands of dollars per hectare.

Table 1

Land use / Variable	Obs	Mean	Std. Dev.	Min	Max
Dairy					
Slope	61800	2.99	4.07	0.00	38.00
LUC class	61800	3.55	1.52	1.00	8.00
Pastoral productivity	61800	1.39	0.30	0.31	2.01
Forest productivity	61800	1.26	0.09	0.89	1.43
Maori status	61800	0.01	0.11	0.00	1.00
Distance - dairy	61800	28.75	23.03	0.00	198.72
Distance - sheep/beef	61800	26.21	19.06	0.00	191.23
Distance - forestry	61800	15.50	9.81	0.00	95.59
Distance - market	61800	11.54	7.29	0.00	77.35
Mean profit - dairy	61800	2.73	0.33	1.93	3.12

[5] Land quality is used in the broadest sense here, to include all explanatory variables that affect profitability.

Mean profit - sheep/beef	61800	0.54	0.22	0.00	0.76
Mean revenue - forest	61800	0.34	0.15	-0.34	0.70
Sheep or beef					
Slope	268457	8.83	7.23	0.00	47.00
LUC class	268457	4.98	1.63	1.00	8.00
Pastoral productivity	268457	1.06	0.35	0.05	2.02
Forest productivity	268457	1.15	0.13	0.37	1.43
Maori status	268457	0.02	0.14	0.00	1.00
Distance - dairy	268457	58.50	33.61	0.00	203.11
Distance - sheep/beef	268457	38.38	25.27	0.00	195.89
Distance - forestry	268457	22.97	13.88	0.00	134.29
Distance - market	268457	17.95	10.49	0.00	85.87
Mean profit - dairy	268457	2.81	0.26	1.93	3.12
Mean profit - sheep/beef	268457	0.42	0.22	0.00	0.76
Mean revenue - forest	268457	0.23	0.19	-0.34	0.70
Forestry					
Slope	43282	11.13	6.76	0.00	41.00
LUC class	43282	5.81	1.28	1.00	8.00
Pastoral productivity	43282	1.22	0.29	0.24	1.97
Forest productivity	43282	1.21	0.10	0.81	1.42
Maori status	43282	0.07	0.26	0.00	1.00
Distance - dairy	43282	52.57	34.97	0.50	183.74
Distance - sheep/beef	43282	38.10	21.24	0.50	163.69
Distance - forestry	43282	20.28	12.09	0.00	77.27
Distance - market	43282	18.73	10.98	0.00	57.49
Mean profit - dairy	43282	2.75	0.28	1.93	3.12
Mean profit - sheep/beef	43282	0.43	0.23	0.00	0.76
Mean revenue - forest	43282	0.29	0.18	-0.34	0.70
Scrub					
Slope	42204	14.64	7.61	0.00	47.00
LUC class	42204	6.21	1.15	1.00	8.00
Pastoral productivity	42204	1.18	0.34	0.07	2.01
Forest productivity	42204	1.19	0.12	0.43	1.42
Maori status	42204	0.09	0.28	0.00	1.00
Distance - dairy	42204	63.34	37.59	0.00	214.81
Distance - sheep/beef	42204	44.80	26.61	0.50	194.77
Distance - forestry	42204	25.50	14.96	0.00	132.59
Distance - market	42204	21.52	12.43	0.50	82.92
Mean profit - dairy	42204	2.76	0.24	1.93	3.12
Mean profit - sheep/beef	42204	0.38	0.21	0.00	0.76
Mean revenue - forest	42204	0.21	0.22	-0.38	0.70

As noted above, observations represent 25-hectare pixels of privately owned land (whether of unrestricted or Maori status). The estimation sample consists of 415,743 observations on the North and South Islands of New Zealand. This corresponds to approximately 10.4 million hectares, or close to 40 percent of the total land area of the country. The omitted observations are either public land (317,879), private land in exogenous uses (258,861) or private land in one of the modelled uses with incomplete data (75,790).

Using pixels as the units of decision making has advantages and disadvantages as well. On the one hand, pixels do not accurately match the geographic scale at which land use decisions are made, and this may lead to measurement error and biased estimates of any spatial influences (Bell and Irwin, 2002). On the other hand, land use decisions are often not made at the farm level either, and many farms are in multiple land uses simultaneously. Large farms are often diverse with respect to their geophysical and other attributes as well, raising the potential yet again for measurement errors. Moreover, unlike pixels, farms are not of the same size. Estimating a farm-based land use model therefore requires the modeller to make – somewhat arbitrary – decisions about weighting each observation's log-likelihood contribution. Historically, many dairy farms in New Zealand were established on parts of already existing large sheep-beef farms. Farm-level modelling would not enable the researcher to make predictions about the likely locations of such transitions.

All raster layers used in the study have an original resolution higher than the 25 hectare resolution used in the model. Likewise, the mapping scale of all polygon data layers is larger (more detailed) than the scale corresponding to a 25-ha resolution. This helps avoid potential problems stemming from spatial correlation.

3. Results

If the error term is independent and identically distributed type I extreme value, the specification leads to the multinomial logit, one of the most widely used discrete choice models. The model owes its popularity in part to the convenient, closed form expression its choice probabilities take, and in part to its transparency. Estimating the parameters of a multinomial logit does not require simulation methods, and is therefore computationally fairly undemanding even for large datasets.

I first estimate a model that relies on geophysical land attributes only as explanatory variables to determine what influence various dimensions of land quality have on private land use decisions. The estimated relationships are conditional on the price (and other) expectations prevailing at the time the land use decision was made. Results appear in table 2, with t-statistics shown in parentheses.

Table 2

Variable	Dairy	Sheep-beef	Forestry
Slope	-0.2349	-0.0668	-0.0551
	(-141.40)	(-76.57)	(-49.08)
LUC class	-0.7294	-0.4556	-0.1195
	(-111.70)	(-81.19)	(-17.51)
Pastoral productivity	1.9422	-1.2082	0.1396
	(85.94)	(-73.30)	(6.73)
Constant	3.2637	6.5948	1.2842
	(75.93)	(178.47)	(28.72)
Log-likelihood value		-352180	

The base outcome is scrub (here and in estimations to follow as well): the scrub coefficients of all variables are normalized to zero, and not shown in the table. The normalization is needed because discrete choice models operate on utility differences. A consequence of this is that the only parameters that can be estimated are those that capture differences across the choice alternatives (Train, 2009). Stated informally, the model is based on which alternative wins, and not by how much.

The specification in table 2 includes the pastoral productivity variable, but not the exotic forest productivity variable. In theory, pasture productivity should not affect the utility derived from forestry land use. The reason for presenting this specification, nonetheless, is more pragmatic: the almost perfect correlation between pasture productivity and exotic forest productivity. Although the two variables have somewhat different absolute scales, with pastoral productivity exhibiting a greater range, this difference is irrelevant from the perspective of estimation: as long as high values of one are associated with high values of the other, the utility-maximizing choice of land use will not change. The alternative, theoretically correct, specification in which both variables are included with cross-coefficients constrained to zero yields a virtually identical fit (log-likelihood value = -352179). Because doing so does not substantially change any results, I present the simpler specification.

There is an additional issue about the specification that merits some discussion. LUC class appears as a cardinal variable rather than several categorical indicators. The effects of this misspecification seem negligible. A model with separate indicators for each LUC class achieves a slightly higher log-likelihood value (-351004), but almost triples the number of explanatory variables. Model fit with four LUC class indicators (each representing two classes) is worse than the fit achieved by the 'misspecified' model. Figure 2, showing how land use shares vary within LUC classes, depicts a nearly monotonic and fairly smooth relationship for most uses, revealing the reason for the strong explanatory power of the cardinal LUC class variable. (There is a slight break in the pattern at LUC class 5 for both dairy and forestry – note however, that there are very few observations in this class, as only about 1.3% of the land falls into this category.)

Most parameters have the expected sign, and are statistically significant at the 1% level. Slope decreases the utility of all three productive uses relative to scrub utility. Additionally, the more intensive the use, the more negative the effect. The LUC class coefficients are also negative indicating that land with more limitations to use is less likely to be devoted to production. Similarly to slope, the magnitude of the LUC class coefficient is largest for dairy and smallest for forestry – again confirming prior expectations about the relationship of land use intensity and land quality.

Pastoral productivity is highly and positively significant for dairying, as it should be. It enters with a negative coefficient in the sheep-beef equation, however, implying that being able to grow more grass biomass matters more for scrub than for sheep or beef farming.[6] This certainly should not lead to the conclusion that additional productivity is undesirable for sheep and beef farms. The finding may be a manifestation of measurement error in the land use data. Sheep farms occupy vast areas of low productivity land in southern regions of the South Island; the estimated negative relationship is a result of this geographic distribution and the strong influence of climate (i.e. geographic latitude) on the productivity variable. It is possible, however, that some of this sheep land is in reality scrub. Abandoned pasture in some regions of the South Island does not grow woody biomass due to a lack of nearby seed sources. Therefore, scrubland may appear as pasture in satellite imagery in these regions, potentially leading to the perverse pasture productivity outcome.[7]

In the forestry equation, pasture productivity (recall, it essentially reflects exotic forest productivity here) has the expected sign and is statistically significant. However, statistical significance is not necessarily a good guide in a sample of over 400,000 observations. The coefficient is an order of magnitude smaller than for the other land uses, indicating, at best, a

[6] The unintuitive finding, of course, just reflects the sample distribution of land uses by pastoral productivity: sheep and beef farms are, on average, found on less productive land than scrub in the dataset.
[7] It is true, however, that extensively managed pasture land has very limited requirements with regard to productivity.

small effect. In light of the weak relationship between remote sensing measurements of NPP and the productivity index for exotic forestry (Basiden, 2006b), this is not entirely unexpected. Overall, the results for pasture productivity are somewhat mixed, though reasonable explanations can be found for the unanticipated outcomes. It is encouraging that the variable appears an extremely significant factor in the choice between the two pastoral uses, dairy and sheep-beef farming (the two carrying significant coefficients of the opposite sign relative to the baseline). This alone would warrant its inclusion in the utility function. Omitting either LUC class or pastoral productivity substantially reduces the log-likelihood value, without affecting the other variable's estimated coefficient.

Lastly, the estimated choice-specific constants merely ensure that modelled aggregate land use shares equal their observed sample counterparts. To the extent that omitted factors systematically affect the desirability of the land use alternatives, their effects are reflected in the estimated constants, but they have no inherent, easily interpretable meaning.

In the next model I estimate, the socio-economic land status indicator and variables controlling for distances to processing facilities and population centres are added. Results are shown in table 3. The cross-coefficients of the facility-distance variables are constrained to zero on theoretical grounds: for a plantation forestry plot, its distance from the nearest dairy processing facility should not matter. A similar argument can be made for the other cross-coefficients.

Table 3

Variable	Dairy	Sheep-beef	Forestry
Slope	-0.2096	-0.0661	-0.0494
	(-123.66)	(-75.10)	(-43.40)
LUC class	-0.6450	-0.4382	-0.0856
	(-96.85)	(-77.45)	(-12.41)
Pastoral productivity	1.6943	-1.1561	-0.0887
	(72.29)	(-68.29)	(-4.14)
Maori status	-1.7661	-1.1248	-0.1074
	(-37.04)	(-46.59)	(-4.06)
Distance – dairy	-0.0237	0	0
	(-102.31)		
Distance - sheep/beef	0	-0.0030	0
		(-17.51)	
Distance - forestry	0	0	-0.0309
			(-53.77)
Distance - market	-0.0153	0.0003	0.0102
	(-18.16)	(0.56)	(13.50)
Constant	4.3499	6.5932	1.7967

	(96.02)	(173.99)	(39.29)
Log-likelihood value		-341014	

Results from the previous estimation are robust to model specification. The only qualitative change is that the pasture productivity coefficient for forestry changes from a small positive value to a small negative value – the accompanying effect on utility and choice probabilities is negligible.

Land in Maori status tends to be associated with uses of lesser intensity because of legal restrictions affecting Maori land. This relationship manifests in the increasingly more negative coefficients estimated for higher intensity land use types. Results for the first three distance coefficients also make sense: the farther a piece of land is located from a processing facility, the lower the attractiveness of the corresponding land use.

Intensively cultivated pieces of land are often found closer to population centres than plots that are managed more extensively. This relationship reflects intensive uses' greater reliance on access to local produce markets (leading to a potentially higher farm-gate price), factor markets, and transportation. Accordingly, distance to the nearest supermarket has a negative effect on dairy attractiveness. Sheep and beef farming and forestry do not seem as sensitive to location relative to population centres. This finding may also be explained by the definition of scrubland. It may, in some cases, include lifestyle blocks and land awaiting development for subdivision. Both of these could be classified as unproductive land by farmers (and would often be located near towns).

Table 4 shows estimation results from the fully specified model. Cross-coefficients are, as before, constrained to zero. Previous results do not change substantially, and the regional mean profit coefficients are positive for the two pastoral uses. Regional mean revenue enters with the wrong sign in the forestry equation. Mean revenues are not necessarily closely related to mean profits: harvest and transportation costs may vary across regions creating differences between revenue and profit rankings. Additionally, there is typically a 30-year lag between the land use decision and the realization of revenues. Differences in regional mean revenues from forests planted three decades ago may not represent current region-specific unobserved factors accurately. The perverse coefficient estimate for forestry revenues is thus not completely unexpected.[8]

[8] I also estimated a specification with interaction terms between regional profits (revenues for forestry) and LUC class. Such a structure allows for parcel-level variation in profits (revenues). The parameter estimates of these

Table 4

Variable	Dairy	Sheep-beef	Forestry
Slope	-0.2072	-0.0633	-0.0497
	(-122.74)	(-71.35)	(-43.86)
LUC class	-0.6386	-0.4354	-0.0717
	(-95.84)	(-76.96)	(-10.30)
Pastoral productivity	1.9360	-1.2985	0.0590
	(71.90)	(-73.32)	(2.54)
Maori status	-1.8203	-1.1395	-0.0793
	(-38.02)	(-47.11)	(-2.99)
Distance - dairy	-0.0231	0	0
	(-99.16)		
Distance - sheep/beef	0	-0.0027	0
		(-15.95)	
Distance - forestry	0	0	-0.0307
			(-53.82)
Distance - market	-0.0150	0.0008	0.0121
	(-17.87)	(1.55)	(15.91)
Mean profit - dairy	0.3596	0	0
	(17.47)		
Mean profit - sheep/beef	0	0.5271	0
		(28.29)	
Mean revenue - forestry	0	0	-0.0082
			(-17.11)
Constant	2.9889	6.4691	2.0889
	(33.99)	(169.55)	(42.95)
Log-likelihood value		-340321	

One strategy for addressing spatial effects in land use choice models is to create a smaller dataset consisting of non-neighbouring observations only. The results in table 4 are robust to such subsampling, where the maximum likelihood estimation was performed on a systematically selected subsample, drawn to maximize the distance between sample points, representing 3% of observations from the full sample. For completeness, estimation results based on this subsample are shown in table 5. Most estimated coefficients change only marginally, suggesting that spatial effects among nearby observations are not significant.

interaction terms were not statistically significant for either land use. Accordingly, these terms are omitted from the model in table 4.

Table 5

Variable	Dairy	Sheep-beef	Forestry
Slope	-0.2023	-0.0624	-0.0482
	(-21.81)	(-12.20)	(-7.44)
LUC class	-0.6554	-0.4617	-0.0944
	(-17.11)	(-13.88)	(-2.34)
Pastoral productivity	1.9020	-1.4133	-0.0056
	(12.52)	(-13.85)	(-0.04)
Maori status	-1.4671	-0.9945	0.0412
	(-5.45)	(-7.07)	(0.26)
Distance - dairy	-0.0235	0	0
	(-18.02)		
Distance - sheep/beef	0	-0.0023	0
		(-2.44)	
Distance - forestry	0	0	-0.0322
			(-9.92)
Distance - market	-0.0144	-0.0024	0.0083
	(-3.11)	(-0.82)	(1.92)
Mean profit - dairy	0.5354	0	0
	(4.67)		
Mean profit - sheep/beef	0	0.6058	0
		(5.71)	
Mean revenue - forestry	0	0	-0.0096
			(-3.56)
Constant	2.7431	6.8215	2.5439
	(5.53)	(30.13)	(8.99)
Log-likelihood value		-10604	

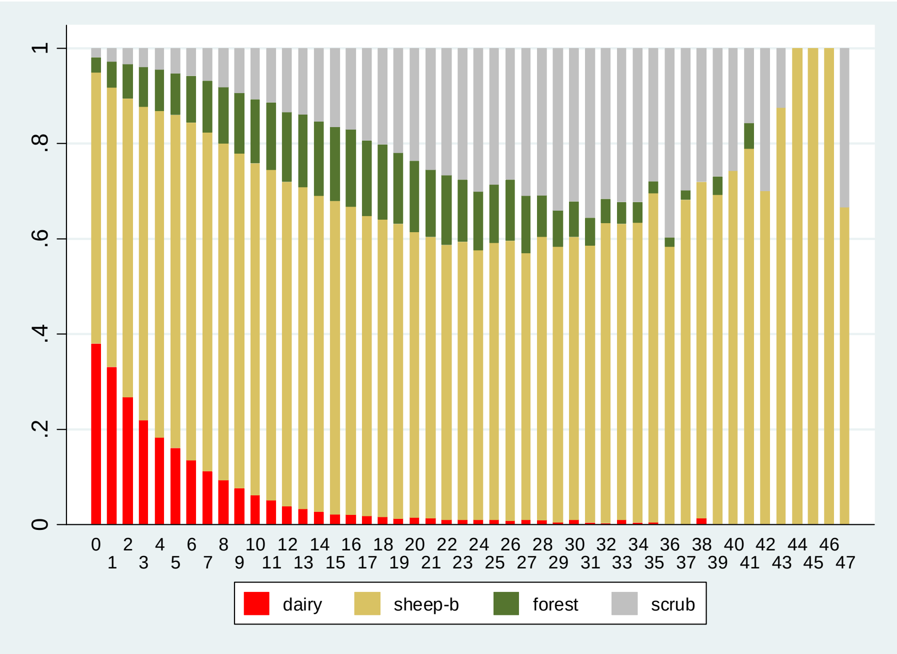
When comparing discrete choice models estimated on the same data, it is generally correct to claim that the model with the higher log-likelihood value fits the data better.[9] The log-likelihood value, however, is not helpful in determining how well the model fits the data: a measure of fit analogous in its interpretation to that of the coefficient of determination, R-squared, in regression models does not exist in discrete choice models. The figures that follow help in the subjective (and more intuitive) assessment of predictions by comparing observed and predicted land use shares across the values of some explanatory variables, across predicted probability categories, and across geographic regions. These assessments do not constitute a

[9] Log-likelihood values from the model estimated on the full sample and the model estimated on the subsample are not comparable.

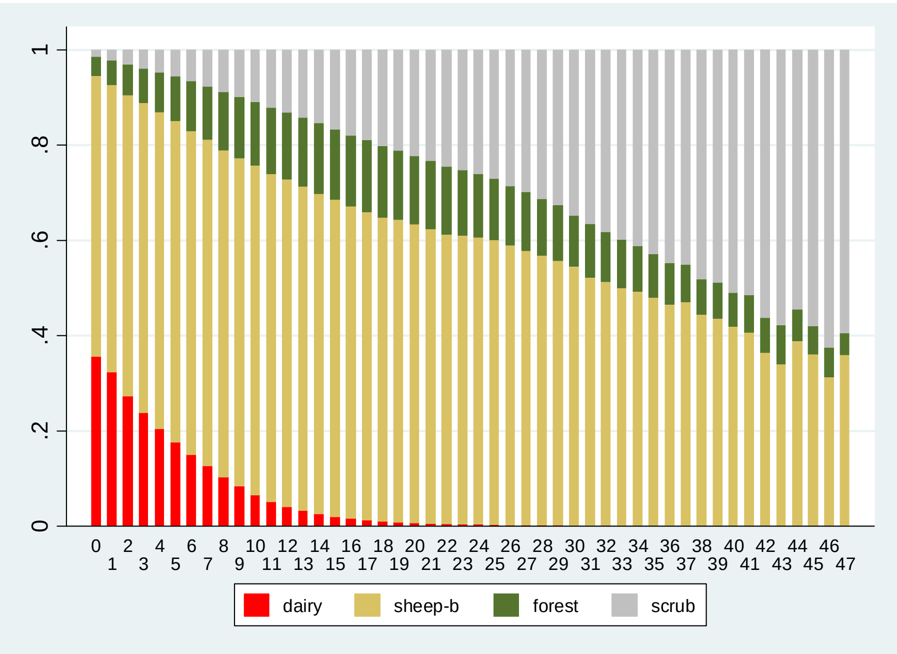
rigorous specification test, but they are informative with respect to model performance along various dimensions.

The two panels of figure 3 show the observed and predicted distributions of land uses by slope, respectively. The panels of figure 4 show these distributions by LUC class. There is a reasonable match between observations and predictions in both figures. Predictions get inaccurate in the upper half of the slope distribution, but only about 2.5% of the observations have a slope of over 25 degrees. For the remaining 97.5% of the sample, predicted shares nearly exactly equal actual shares. Similarly, the distributions over LUC class appear very similar. The difference between predictions and observations is greatest for LUC classes 1, 5 and 8, the three classes with the smallest land area in the sample.

Figure 3

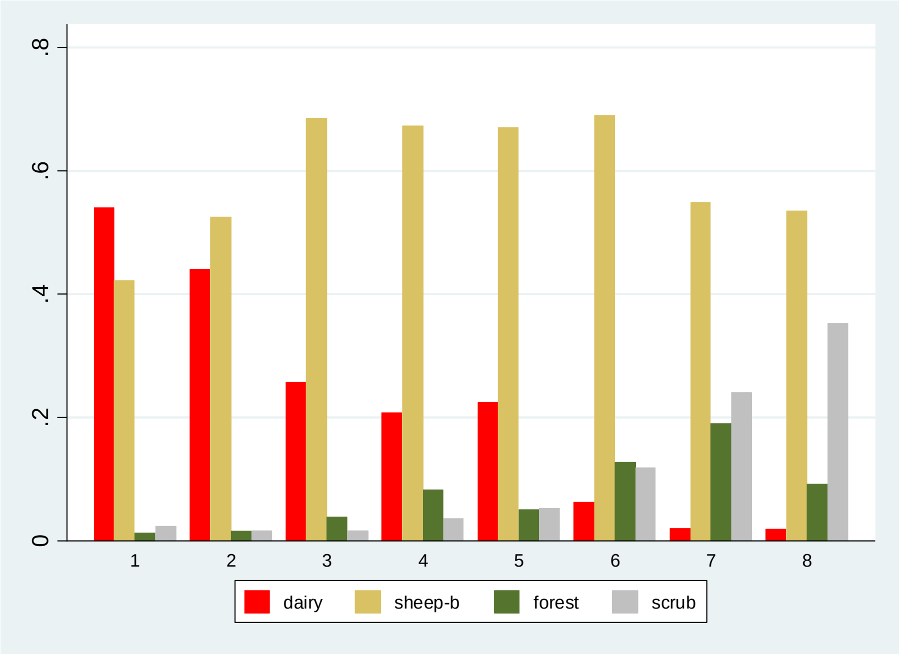


0
.2
.4
.6
.8
1
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47
dairy
sheep-b
forest
scrub

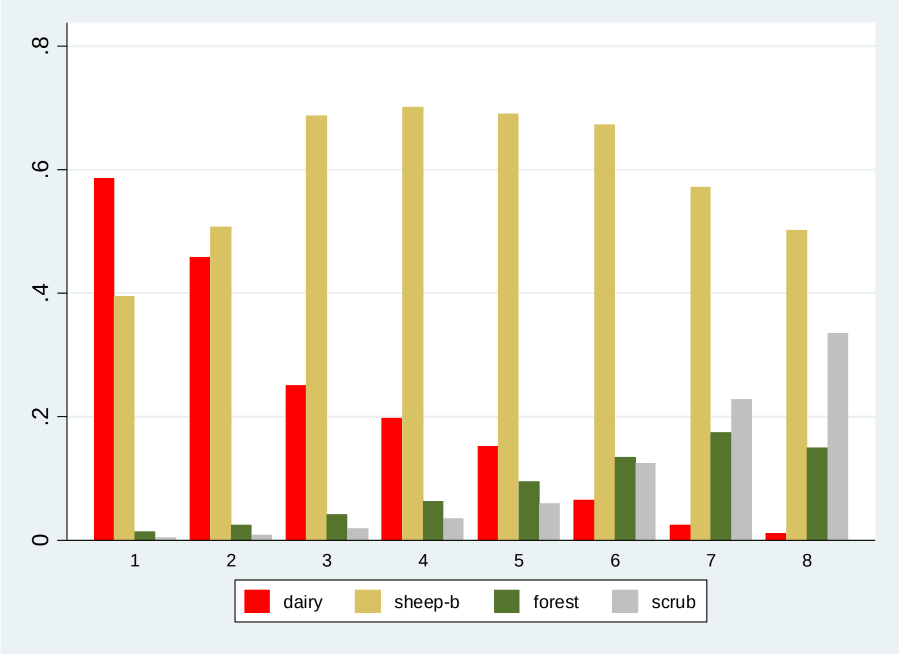


0
.2
.4
.6
.8
1
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47
dairy
sheep-b
forest
scrub

Figure 4



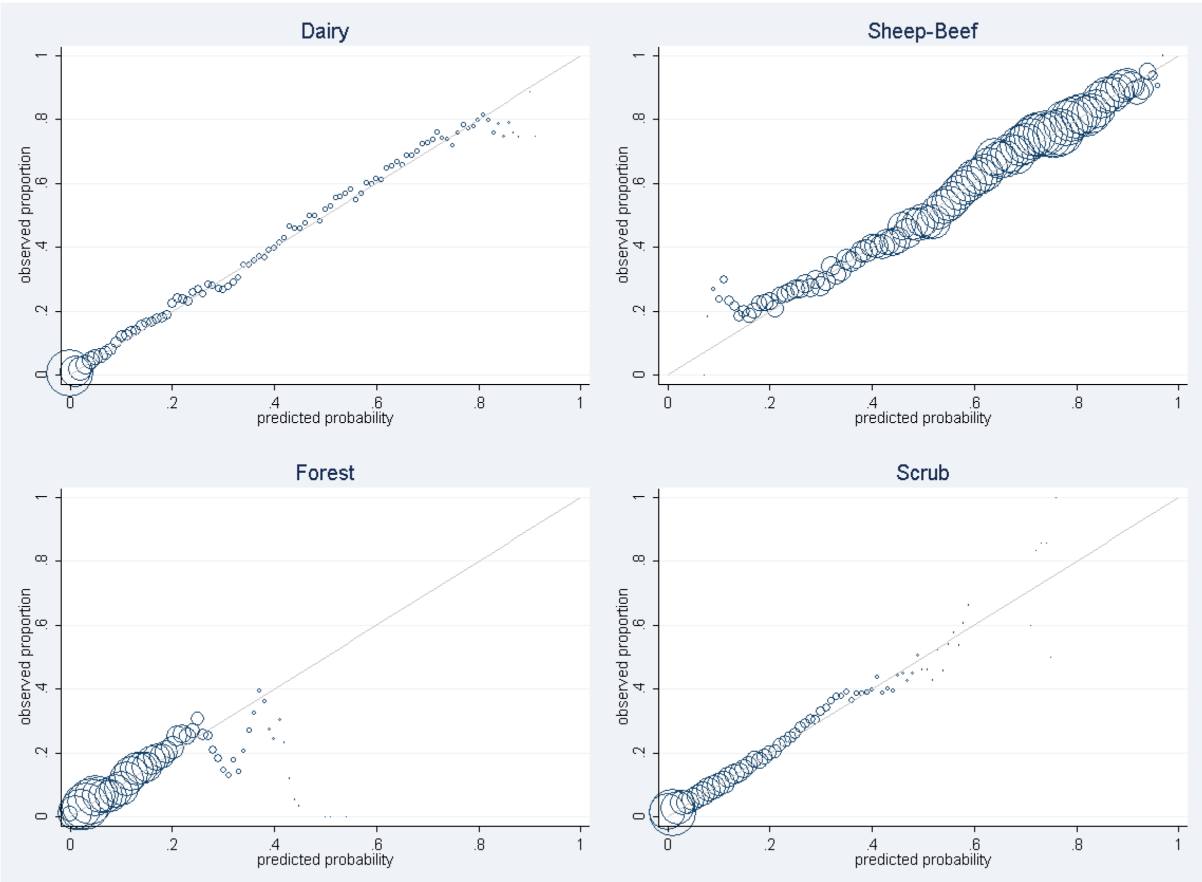
0
.2
.4
.6
.8
1
2
3
4
5
6
7
8
dairy
sheep-b
forest
scrub



0
.2
.4
.6
.8
1
2
3
4
5
6
7
8
dairy
sheep-b
forest
scrub

In each graph of figure 5, all pixels with a certain predicted probability value for a given land use have been grouped into one bin. The probability prediction is compared to the fraction of pixels within the bin that are actually in the relevant land use type, and the size of each circle is proportional to the number of observations in that bin. The graphs show that on average, modelled choice probabilities are approximately correct at the pixel level. For example, the model sees most pieces of land as unsuitable for dairying, and these are, indeed, very rarely in dairy use. Likewise, a few plots are predicted to be extremely good for dairying, and these are, indeed, most often in dairying. Being correct on average does not, of course, imply that the model can predict the choice of land use on a given parcel as any choice probability not equal to zero or one implies some level of uncertainty.

Figure 5



At the scale of Territorial Authorities (figure 6), the degree of correspondence between observed and predicted shares decreases. In these graphs, each circle represents a TA, with the area of the circle proportional to the land area of the TA. While most TA-observations lie near the 45-degree line, the relationship appears much weaker than at the pixel-level: predictions are perhaps reasonable, but far from perfect. The finding that predictions are poorest at smaller spatial scales may indicate the presence of unobserved factors at these scales, such as TA-level differences in infrastructure or legislation. In some cases, underlying historical reasons may exist. For example, the observed share of plantation forestry is more than double its predicted share in a few TA's. Some of these are located in the central North Island, where forests were planted last century due to the belief that selenium deficient soils make land unsuitable for pastoral farming in the area.

Figure 6