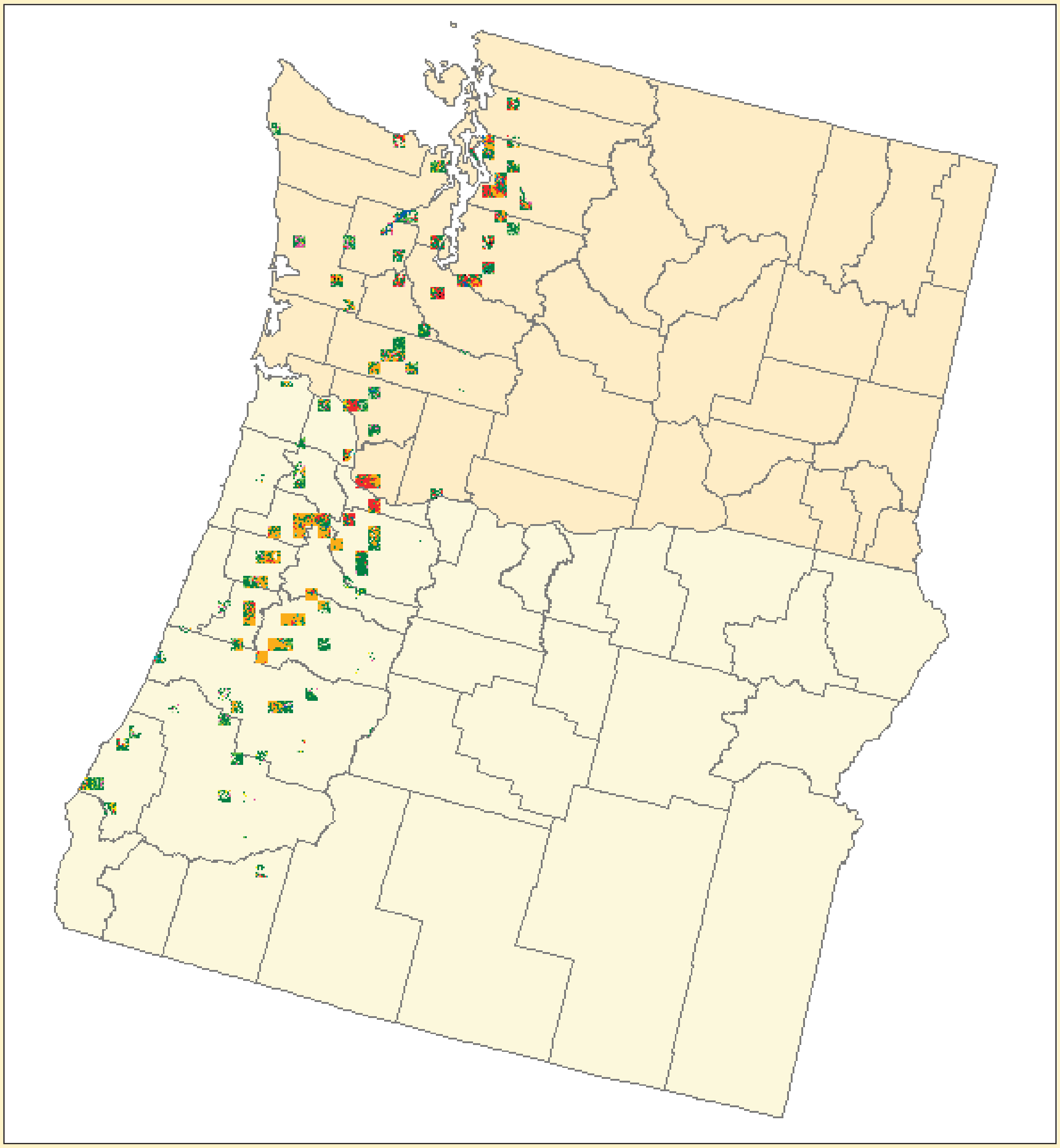
USDA
United States Department of Agriculture

A Spatial Econometric Analysis of Land-Use Change With Land Cover Trends Data: An Application to the Pacific Northwest

David J. Lewis and Ralph J. Alig



Forest Service | Pacific Northwest Research Station | Research Paper PNW-RP-600 | September 2014

Authors

David J. Lewis is an associate professor, Department of Applied Economics, Oregon State University, 200A Ballard Ext. Hall, Corvallis, OR 97331; **Ralph J. Alig** is a research forester emeritus, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331.

Abstract

Lewis, David J.; Alig, Ralph J. 2014. A spatial econometric analysis of land-use change with land cover trends data: an application to the Pacific Northwest. Res. Pap. PNW-RP-600 Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 44 p.

This paper develops a plot-level spatial econometric land-use model and estimates it with U.S. Geological Survey Land Cover Trends (LCT) geographic information system panel data for the western halves of the states of Oregon and Washington. The discrete-choice framework we use models plot-scale choices of the three dominant land uses in this region: forest, agriculture, and urban development. The results provide a technical foundation for developing larger scale models from the LCT database. In particular, we develop a random-effects estimation method for dealing with the spatially clustered sample design underlying the LCT. We also exploit the increased spatial information content available in the LCT by exploring the estimation of a fully spatial multinomial discrete-choice land-use model by including measures of land-use agglomeration economies as independent variables in estimation. Estimation of the spatial econometric model includes a novel combination of panel-data random parameters logit estimation with instrumental variables implemented within the recently developed control function approach. The estimated econometric model is used to project landscape change in the presence of alternative assumptions regarding future urban returns. Our results indicate that variation in urban returns on the order of what was experienced in the housing boom and bust of the 2000s generates a wide range of predicted future land-use shares in developed uses. The Puget Lowland ecoregion has by far the most sensitive landscape projections in response to wide swings in urban returns.

Keywords: Land use, spatial modeling, econometric, resource economics, land development.

Summary

Empirical models of land-use change have long been used in environmental and resource economics for policy analysis of the effects of land-use change on the forest land base, including analyses of urban sprawl and ecosystem services. This paper estimates a new regional-level econometric land-use model at the plot scale from the U.S. Geological Survey's Land Cover Trends (LCT) database, a spatially detailed panel dataset covering the contiguous 48 states. The model is estimated for the western halves of the states of Oregon and Washington from 1980 to 2000 and focuses on transitions among the three major land uses in the region: forest, agriculture, and urban development. The empirical data used to condition plot-level probabilities of land-use transition include county-level measures of net returns to land and plot-level measures of soil quality, distance to cities and roads, and measures of land-use agglomeration economies. The inclusion of detailed spatial independent variables provides an advance over prior large-scale models estimated from land-use surveys that do not disclose the exact location of surveyed land plots. We provide several novel additions to the empirical economics literature on land use, including (1) a method for dealing with the clustered sampling strategy implicit in the LCT design, (2) an approach for estimating spatial econometric discrete-choice model in a multinomial setting, and (3) providing empirically based land-use projections for the Pacific Northwest under alternative assumptions regarding the level of future returns to urban development. The model development also provides a technical foundation for developing future national-scale model econometric models from the LCT data.

Contents

Introduction

Plot-level empirical land-use models are widely used in environmental and resource economics for policy analysis of the effects of land-use change on urban sprawl, forest land loss, ecosystem services, and biological diversity. Many of these analyses can be classified according to the scale with which they are estimated and applied. The first category can be labeled small-scale models and are generally estimated at the scale of a single county or smaller (e.g., Butsic et al. 2011, Irwin and Bockstael 2002, Lewis et al. 2009, Newburn et al. 2006, Towe et al. 2008). These studies are typically estimated from parcel map data provided by local or regional planning or assessor authorities. The strength of these studies is their being extremely detailed in their spatial information content; their weakness is that their small scale limits any broader applicability. The second category can be labeled large-scale models, and includes land-use models that are estimated at regional (e.g., multistate) or national levels (e.g., Langpap and Wu 2011, Lewis and Plantinga 2007, Lubowski et al. 2006). These studies are typically estimated from land-use surveys such as the U.S. Department of Agriculture (USDA) National Resources Inventory (NRI) or Forest Inventory and Analysis (FIA) conducted by the USDA Forest Service. Large-scale models have the strength of being much more broadly applicable with the ability to analyze larger scale environmental problems and policies. The weakness of large-scale models is the lack of detail in their information content, which leads to a spatial specificity that is generally far inferior to small-scale models. This weakness is particularly important for large-scale analyses of the many environmental goods whose provision depends on fine-scale spatial patterns.

The purpose of this paper is to improve on the specification of large-scale land-use models by developing a multistate empirical land-use model from a publicly available spatial panel dataset that has far more spatial information than land-use surveys such as the NRI or FIA. The discrete-choice framework we use models plot-scale choices of the three dominant land uses in the western half of the Pacific Northwest: forest, agriculture, and urban development. The model is estimated for the western halves of the states of Oregon and Washington using micro spatial panel data from the U.S. Geological Survey (USGS) Land Cover Trends (LCT) dataset from 1980 to 2000. The LCT dataset includes four snapshots of the landscape, corresponding to three separate land-use transition periods. The study region includes two states with large urban centers (e.g., Seattle, Portland, etc.) and significant amounts of undeveloped agricultural and forest land. The empirical data used to construct independent variables include plot-level measures of soil quality, distance to cities and roads, and measures of land-use agglomeration economies. The data also include temporal variation in county-level measures of the net returns of an acre (hectare) of land to urban, agricultural, and forest uses of land.

The inclusion of spatial independent variables representing the distance of plots from cities and roads has been absent in many large-scale plot-level econometric analyses of land-use change. This absence has been driven by the fact that the national land-use surveys—such as the NRI—do not disclose the exact location of sample plots. However, the LCT provides an alternative land-use dataset to the NRI that potentially can be used as the basis for a national-scale econometric model similar to Lubowski et al.'s (2006) NRI-based model. The LCT is a nationwide geographic information system (GIS) panel database derived from manually edited satellite images, aerial photography, and topographic maps. A primary advantage of the LCT over the NRI is the spatial information content—the exact location of plots can be observed and corresponding information can be included as independent variables in econometric estimation. Further, while land use/land cover datasets derived from automated interpretation of satellite imagery have been found to underrepresent low-density development (Irwin and Bockstael 2007, Kline et al. 2009), we provide a comparison of urban development probabilities generated from the LCT and NRI that suggests minimal differences for this region. Similar to many land-use analyses, we use the LCT's Anderson level I land cover classification as a good proxy for land use.

Using the LCT for estimation brings up multiple research possibilities and challenges that do not arise in models estimated from land-use surveys. First, unlike the NRI, the LCT does not provide a random sample of plots within counties. Rather the LCT provides a time-series of GIS 10- by 10-km maps randomly sampled within U.S. ecoregions, and we sample plots within each LCT block for a computationally feasible estimation. However, as we show formally, this process generates spatial correlation in the model unobservables—plots within a block will be affected by a different set of unobservables than plots from a separate block. Second, the observation of land use within each block brings up the possibility of estimating a fully spatial model that includes variables representing spatial externalities induced by neighboring land use. For example, Irwin and Bockstael (2002) found evidence of negative spatial externalities resulting from urban land on the development probability of neighboring lands, while Lewis et al. (2011) found evidence of positive spatial externalities resulting from organic dairy farms on the decision of neighboring conventional farmers to convert to organic. These two papers are binary models with only two land-use choices. A primary challenge is how to estimate similar spatial effects in a multinomial discrete-choice model of three land-use choices when the spatial lag is endogenous.

The econometric analysis in this paper features several novel additions to the empirical economics literature on land use. First, we develop the first large-scale econometric land-use change model estimated from the LCT and explore a method for dealing with the LCT's clustered sampling properties in estimation. The method amounts to including landscape block random effects in a random parameters logit (RPL) framework that accounts for spatially correlated unobservables induced by the LCT's sampling strategy. We provide Monte Carlo evidence on the method's ability to yield consistent parameter estimates. Accounting for landscape-block unobservables with a random effect is preferred to fixed effects in our nonlinear model because fixed effects would generate inconsistent estimates owing to the incidental parameters problem.

Second, we develop a novel approach for estimating a full spatial econometric discrete-choice land-use model in a multinomial setting that is estimable with standard maximum simulated likelihood techniques. In addition to modeling spatially correlated unobservables, the spatial properties of the model include a type of spatial lag—the inclusion of the share of a plot's neighboring landscape in various land uses—as a way to explicitly model and test for spatial agglomeration economies in land-use change. Our approach accounts for the endogeneity of the spatial lag with the recently developed control function approach for including instruments in multinomial discrete-choice models (see Train 2009: chap. 13). We use exogenous land-quality indexes at the landscape-block level for instruments. To date, fully spatial multinomial discrete-choice econometric models have been little used owing to computational constraints (see LeSage and Pace 2009). However, the combination of the LCT's clustered sample of discrete landscape blocks, along with panel data and control-function RPL techniques, provides an alternative framework for incorporating both a spatial lag and spatial errors.

Third, we use the estimated land-use econometric model to provide empirically based landscape projections for the Pacific Northwest under alternative assumptions regarding the level of future urban returns. The landscape projections indicate substantial urban growth and loss of forest and farmlands in the Puget Lowland and Willamette Valley ecoregions, with minimal changes in the Coast Range, Cascades, and North Cascades regions. Our results demonstrate that landscape projections in the Puget Lowland ecoregion are extremely sensitive to the assumed level of future urban returns that was witnessed in the decade of the 2000s. In particular, the share of the Puget Lowland ecoregion projected to be developed is between 12 and 14 percentage points higher if the level of real urban returns remains at the levels observed in 2006 as compared to if real urban returns remain at the more modest levels observed in 2012. Variation in the projected development share of the Puget Lowland ecoregion between 12 and 14 percentage points presents major challenges for managing a landscape for ecosystem service provision, as 12 percent of a 1.8 million ha region is approximately 0.2 million ha of land.

Finally, this analysis serves as a pilot study for the potential creation of a national-scale econometric land-use model based on LCT data. The Forest Service extensively used Lubowski et al.'s (2006) NRI-based national model for Resource Planning Assessment (RPA) and climate change analyses. However, changes in the NRI sampling scheme after 1997 has created a need for a new data source for future national-scale models. This paper develops the technical foundation necessary to estimate a large-scale econometric land-use model with LCT data.

Land-Use Modeling by the USDA Forest Service

The Forest Service (e.g., Adams and Haynes 2007) has conducted national and regional assessments of the forest and rangeland land base for many decades. A land-base assessment (e.g., Alig et al. 2010) is a key part of such natural resource assessments, given the importance of land and land-use changes for forest and agricultural products, living space, and ecosystem services provided by forest ecosystems. With a broad range of economic, ecological, and biophysical phenomena of interest in such assessments, considerable uncertainty exists about future projections of outcomes, particularly projections that look forward 50 years with the added pressures of climate change. Past land-base assessments by the Forest Service have typically focused on one "business-as-usual" future, although there have been variations in this approach in analyzing additional scenarios. For example, varying assumptions about future population have been used to create high, medium, and low trajectories of supply and demand. Now with the growing interest in markets for carbon as an ecosystem service, analysts can compare results to those of other studies that explicitly include carbon price scenarios (e.g., Alig et al. 2010).

Land-base assessments provide information that can help shape perceptions about whether we can sustain both increasing consumption of forest products, forest resource conditions, and ecosystem services, and how to best adapt to and mitigate climate change. Related data illustrate the dynamics of our Nation's land base and how adjustments are likely to continue in the future. Land-use change projectons can also provide inputs into a larger system of models that project forest resource conditions and harvests, wildlife habitat, climate change, and other natural resource conditions. Current debates about sustainability and concerns about climate change involve both physical notions of sustainability and competing socioeconomic goals for public and private land management. The fixed land base necessitates viewing "sustainability" across the entire land base and across sectors.

Methods and data sources used in land-base assessments have changed materially over the last several decades. In general, since around 1980, land-use projections have moved from reliance on expert opinions (e.g., Wall 1981) to systematic models involving systems of land-use equations. Early land-use models relied on FIA data (e.g., Alig 1986), with initial models estimated for the Southern United States, which has a significant amount of the Nation's private timberland as well as some of the larger forestry datasets. The first cross section of NRI data was used to create urban-area models (Alig and Healy 1987) for use in land-base assessments. Panels of NRI data were later used to create subsequent land-use models used in RPAs (e.g., Alig and Plantinga 2004, Alig et al. 2010). With the availability of spatial or georeferenced data, models were developed at a regional or subregional level (e.g., Kline and Alig 2005; Kline et al. 2003) to complement larger scale aspatial land-use models, especially to help improve finer scale ecological investigations (e.g., Lewis and Plantinga 2007). In addition, other analyses by the Forest Service and others used U.S. Census Bureau data on housing densities combined with spatial land cover data (e.g., National Land Cover Database) at a watershed level to investigate pressures on U.S. private forests from housing developments, referred to as "Forests on the Edge" studies (e.g., Stein et al. 2005).

Basic Econometric Framework and Database Construction

Conceptual Model

The econometric framework begins with the assumption that landowners allocate a plot of homogeneous quality land to the use that maximizes the present discounted value of expected net returns less any costs of converting land. Expectations of net returns are assumed static and depend on current and historical net returns, thereby generating the decision rule that the landowner chooses the use generating the greatest annualized net return less conversion cost (Plantinga 1996). Although the landowner is assumed to observe the net returns to alternative uses in each decision period, this information is not perfectly available to the researcher. As such, the annualized net return to land can be specified as a function of both deterministic and random components (Lewis and Plantinga 2007, Lubowski et al. 2006). For plot i that begins period t in use j and ends in use k, the real annualized net return (R_{ikt}) net of annualized conversion costs ($r\mathrm{C}_{ijkt}$) are:

$$\mathrm{R}_{ikt} - r\mathrm{C}_{ijkt} = \alpha_{jk} + \beta_{0jk}\,\mathrm{R}_{\mathrm{C}(i)kt} + \beta_{1jk}\,\mathrm{LQ}_{ik} R_{C(i)kt} + [\omega_{ik} + \varphi_{\mathrm{B}(i)k} + \varepsilon_{ijkt}] \tag{1}$$

for all uses $k = 1,\ldots.\mathrm{K}$ and time periods $\mathrm{t} = 1,\ldots.,\mathrm{T}$, where the unobservable in the square bracket of the right hand side of (1) is an unobservable term composed of a time-invariant plot/use random effect (ω_{ik}), a random effect specific to the landscape block B that contains plot i ($\varphi_{B(i)k}$), and an IID unobservable (ε_{ijkt}). The variable $R_{C(i)kt}$ is the average net return from use k at time t in the county C that contains parcel i, and LQ_{ik} is the land quality of parcel i in use k. Land quality is represented as soil quality when k = agriculture or forest, and land quality is represented as distance to cities and roads when k = urban. The interaction of $R_{C(i)kt}$ and LQ_{ik} allows the deterministic parcel return to deviate from the county average owing to observable measures of land quality. Finally, the land-use specific constants α_{jk} are assumed to account for real annualized costs of converting from use j to use k.

If the IID unobservable ε_{ijkt} is distributed type i Extreme Value, then equation (1) is the latent equation version of an error components random parameters logit model (Train 2009: chap. 6). This flexible error components structure allows for unobserved correlation that is specific to the spatial-temporal nature of the panel data used in estimation. First, because the land-use choice for each parcel i is observed repeatedly, ω_{ik} captures time-invariant parcel unobservable determinants of i's return to use k. For example, the parcel's distance to a river does not change over time and might affect the return to developing the parcel. Second, because we observe parcels in landscape blocks B, then $\varphi_{B(i)k}$ captures block-specific determinants of i's return to use k. For example, the Pacific Northwest has many microclimates that affect the tree species that can be grown (e.g., the commercially valuable Sitka spruce in the far west Coast Range). This flexible error structure allows for temporally correlated unobservables with ω_{ik}, and spatially correlated unobservables with $\varphi_{B(i)k}$.

Land Use Dataset Used for Estimation

This project uses spatial panel data from the LCTs project funded by the USGS. The LCT is a national dataset derived from manually edited satellite images, aerial photography, and topographic maps that classifies land cover into 11 categories for five periods in time: 1973, 1980, 1986, 1992, and 2000. Owing to the lack of available data on the net returns to land prior to 1978, we exclude the 1973 to 1980 transition period in our analysis. Importantly, the LCT is not derived solely from an automated algorithm run on satellite data. Rather, accuracy is improved by combining satellite data with manual editing from aerial photographs and topographic maps. This long panel provides both spatial and temporal variation in land use, allowing the researcher to observe changes in land use over multiple periods of time. Table 1 indicates the classification system used in the LCT.

The LCT data are collected by first stratifying the continental United States into Envirnomental Protection Agency (EPA) level III ecoregions, and then randomly selecting a sample of 10- by 10-km blocks from each ecoregion. Satellite images of each block are then placed into 60- by 60-m pixels and assigned one of the land cover classes from table 1. Figure 1 shows a map of the study region, Oregon and Washington west of the Cascade crest, including the privately owned

Table 1— Land Cover Classifications in the Land Cover Trends data

Land Cover Classification	Description
Water	Areas persistently covered with water, such as streams, canals, lakes, reservoirs, bays, or oceans.
Urban	Areas of intensive use with much of the land covered with structures or anthropogenic impervious surfaces (e.g., high-density residential, commercial, industrial, roads, etc.) or less intensive uses where the land cover includes both vegetation and structures (e.g., low-density residential, recreational facilities, cemeteries, parking lots, utility corridors, etc.), including any land functionally related to urban or suburban environments (e.g., parks, golf courses, etc.)
Mechanically disturbed	Land in an altered and often unvegetated state that, owing to disturbances by mechanical means, is in transition from one cover type to another. Mechanical disturbances include forest clear-cutting, earthmoving, scraping, chaining, reservoir drawdown, and other similar human-induced changes.
Mining	Areas with extractive mining activities and directly related land uses.
Barren	Land comprised of naturally occurring soils, sand, or rocks where less than 10 percent of the area is vegetated.
Forest	Tree-covered land where the tree cover density is greater than 10 percent. Note that cleared forest land (i.e., clearcuts) is mapped according to current cover (e.g., mechanically disturbed).
Grass/shrub	Land predominately covered with grasses or shrubs. The vegetated cover must comprise at least 10 percent of the area.
Agriculture	Land in either a vegetated or an unvegetated state used for the production of food and fiber. This includes cultivated and uncultivated croplands, hay lands, pasture, orchards, vineyards, and confined livestock operations. Note that forest plantations are considered forests regardless of the use of the wood products.
Wetland	Land where water saturation is the determining factor in soil characteristics, vegetation types, and animal communities. Wetlands usually contain both water and vegetated cover.
Nonmechanically disturbed	Land in an altered and often unvegetated state that, owing to disturbances by nonmechanical means, is in transition from one cover type to another. Nonmechanical disturbances are caused by fire, wind, floods, animals, and other similar phenomena.
Ice/snow	Land where the accumulation of snow and ice does not completely melt during the summer period (e.g., alpine glaciers and snowfields).

Source: Land Cover Trends Project, http://landcovertrends.usgs.gov/main/classification.html.

portion of the sampled LCT blocks. Because the econometric model is motivated as a model of private decisionmaking, public land pixels are identified and removed. The data used for estimation include the five EPA level III ecoregions in western Oregon and Washington: Willamette Valley, Cascades, Puget Lowland, Coast Range, and North Cascades.

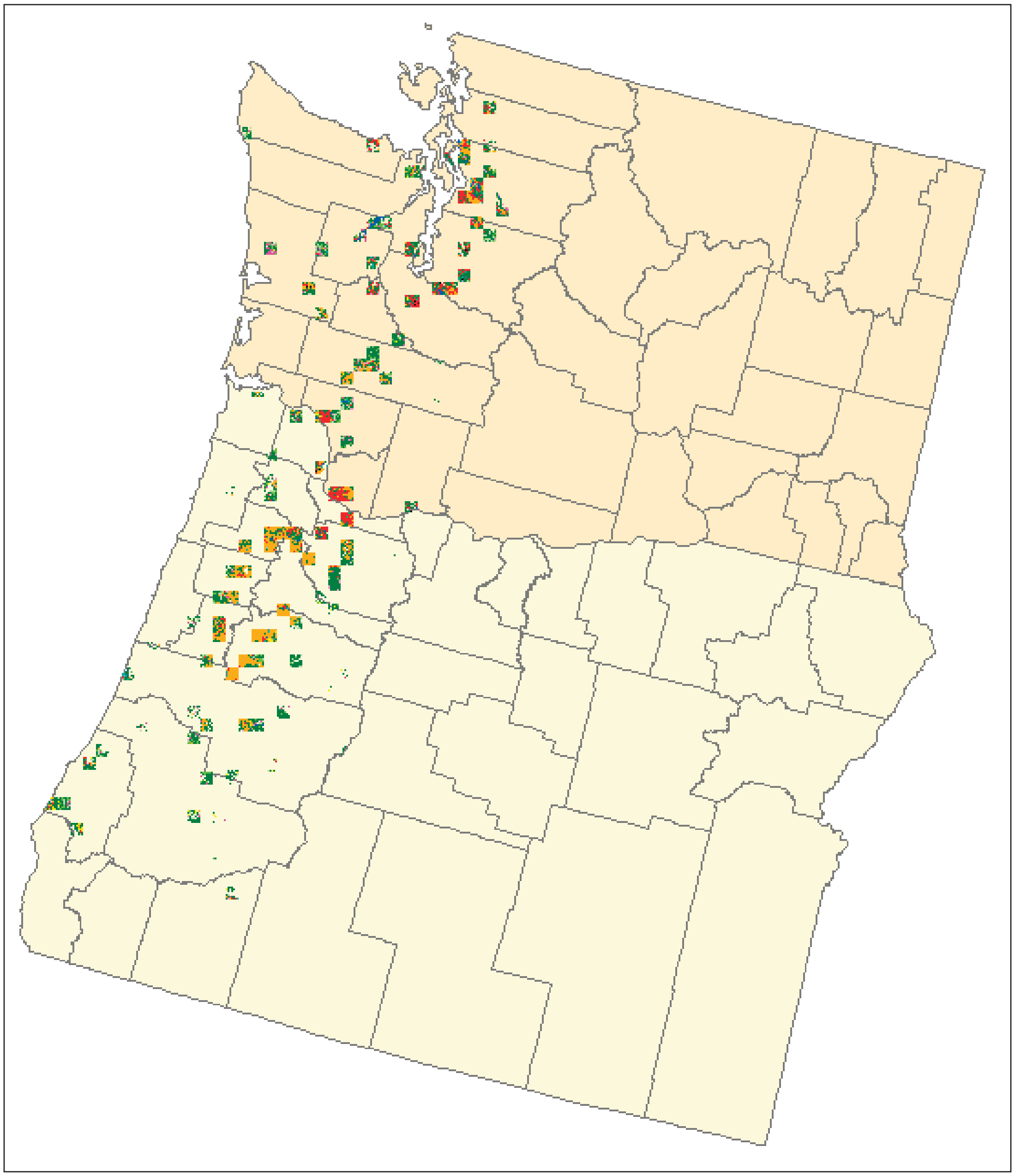


Figure 1—Map of the sample Land Cover Trends blocks used in analysis for western Oregon and Washington.

The region is dominated by the top three land uses of forest, agriculture, and urban lands. The distribution of land-use change by transition period is presented in table 2. The most common transitions between land covers are forest to mechanically disturbed, and mechanically disturbed to forest. Importantly, the forest to mechanically disturbed transition is typically clearcut forestry in this region, while mechanically disturbed to forest transitions are regrowth after clearcut (see Sleeter et al. 2012). As such, transitions between forest and mechanically disturbed are treated as changes in land cover but not as a change in land use. Therefore, the most

Table 2—Distribution of Pacific Northwest land use change by land cover trend transition period for private land

	1980–1986	1986–1992	1992–2000
		Percent	
Forest → forest	98.96	98.55	98.16
Forest → agriculture	0.28	0.36	0.15
Forest → urban	0.76	1.09	1.68
Agriculture → agriculture	98.14	97.98	98.07
Agriculture → forest	0.27	0.56	0.39
Agriculture → urban	1.59	1.46	1.54

common land-use transitions in this region involve the development of forest and agricultural lands into an urban use (see Sleeter et al. 2012). As shown in table 2, for private lands, the development of forested and agricultural lands is much more common than conversions between agriculture and forest.

Sampling of Plots for Estimation

The full LCT dataset in this region consists of over 3 million pixels, a number that is far too much for the maximum simulated likelihood techniques that will be used in econometric estimation. Therefore, we sample 5,000 unique pixels and observe their land-use classifications over all LCT change periods. To ensure that we capture enough variation in land-use change, we oversample change pixels. Following the sampling analysis of LCT data conducted by Chambers (2010), we divide the data into two separate strata: "change pixels" and "nonchange pixels." The "change pixels" are pixels that change land use (transitions between forest and mechanically disturbed are not considered a change—see above) at some point during the LCT time period. The "nonchange pixels" are pixels that do not change land use during the LCT time period. The LCT population of private pixels has a probability of choosing a change pixel at 27 percent, while our oversample of change pixels has a probability of a change pixel equal to 35 percent. Standard Logit estimation off this "choice-based" sample generates consistent estimates of all parameters except the alternative-specific constants, which can be adjusted in a straightforward manner (Train 2009). We discuss the adjustment of the alternative-specific constants in the "Landscape Projections and Out-of-Sample Forecast Text" section.

Comparing Development Probabilities Across the LCT and NRI

Land-use datasets derived from automated interpretation of satellite imagery have been found to underrepresent low-density development (Irwin and Bockstael 2007, Kline et al. 2009). Irwin and Bockstael (2007) compared the widely used National Land Cover Dataset (NLCD) to local parcel maps available from local planning agencies and found widespread misclassification of exurban development. Likewise, Kline et al. (2009) compared auto-interpreted satellite data of land cover for western Washington to survey-based FIA plots and also found common misclassification of low-density development as nondevelopment. In contrast to the satellite-derived datasets considered by Irwin and Bockstael and Kline et al., the LCT does not simply use an automated interpretation from satellite imagery. Although Landsat satellite imagery is the primary source for the LCT maps, a variety of ancillary sources are used to improve classification (see Sleeter et al. 2012: app. 4). In particular, all classifications were manual and combined the base satellite imagery with aerial photographs, topographic maps, and Google Earth[1] imagery. Because we are aware of no formal tests of the LCT in determining urban development, we compare land-use change probabilities generated from the LCT with those same probabilities generated by the survey-based NRI in table 3. The LCT probabilities are calculated from private land parcels, while the NRI probabilities are calculated from nonfederal lands. Both probabilities are annualized to account for the different time period lengths in the two datasets and to provide an apples-to-apples comparison. Results suggest minimal differences for this region. In particular, results

[1] The use of firm or trade names in this publication is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.

Table 3—Annualized land-use change probabilities for the Land Cover Trends (LCT) and Natural Resources Inventory (NRI) for Oregon and Washington west of the Cascades

	Period 1 NRI—1982 to 1987 LCT—1980 to 1986	Period 2 NRI—1987 to 1992 LCT—1986 to 1992	Period 3 NRI—1992 to 1997 LCT—1992 to 2000
Forest to forest:			
NRI	0.9997	0.9994	0.9996
LCT	0.9979	0.9971	0.9974
Forest to agriculture:			
NRI	0.0002	0.0005	0.0003
LCT	0.0006	0.0007	0.0002
Forest to urban:			
NRI	0.0001	0.0001	0.0001
LCT	0.0015	0.0022	0.0024
Agriculture to forest:			
NRI	0.0052	0.0050	0.0020
LCT	0.0005	0.0011	0.0006
Agriculture to agriculture:			
NRI	0.9937	0.9935	0.9941
LCT	0.9963	0.9959	0.9972
Agriculture to urban:			
NRI	0.0011	0.0015	0.0038
LCT	0.0032	0.0029	0.0022

in table 3 provide no evidence that the LCT systematically underrepresents urban development compared to the NRI. This result is in contrast to Irwin and Bockstael (2009) and Kline et al.'s (2009) analyses of land-use datasets solely derived by automated interpretation of satellite imagery alone and provides confidence in the LCT as a data source for econometric modeling.

Data for Independent Variables Used in Econometric Estimation

County-level net return variables come from Lubowski's (2002) national-level dataset of annual per-acre net returns to crops, pasture, range, and urban land uses from 1978 to 1997. Landowners are assumed to form expectations of net returns based on average returns from the 3-year period preceding each transition period. Because the LCT bundles crop and pasture lands together as agriculture, average agricultural net returns are constructed as a weighted average of crop and pasture returns, with the weights derived from the NRI. All net returns are adjusted to 1990 dollars using the consumer price index. The net returns are linked to each sample plot by identifying the county in which each parcel resides using a GIS overlay of a county layer and the sample LCT plots. Table 4 presents average county-level returns for the three time periods used to construct the average returns used in estimation. Urban returns have generally climbed over time, agricultural returns have fallen, and timber returns were flat over the first two periods and significantly higher during the final period. The significantly higher level of forest returns in the final period was likely influenced by the large-scale reductions in federal timber harvests from public lands resulting from the spotted owl restrictions (Wear and Murray 2004).

Data on plot-level soil quality comes from the Soil Survey Geographic Database (SSURGO from the USDA National Resources Conservation Service). The SSURGO data categorize soil quality into eight nonirrigated Land Capability Classes (LCC) corresponding to land productivity (1 is highest, 8 is lowest). A GIS map of LCC is overlaid to the sample of LCT plots to determine the soil quality of each plot. Soil quality is a strong determinant of crop and timber yields, and so is used as an interaction with agricultural and forest returns in econometric estimation to account for observable plot soil characteristics. The LCC is widely used as a soil indicator in econometric land-use models.

Table 4—Average county-level real per-acre net returns for western Oregon and Washington

Years	Urban	Forest	Agriculture
		Dollars	
1978–1980	3,965	15	134
1984–1986	4,362	14	26
1990–1992	5,803	55	41

The Euclidian distance of each plot to the nearest city of greater than 10,000 people and the distance to a major road are included as interactions with urban returns in econometric estimation to account for observable location characteristics that influence plot-level development probabilities. Data on city populations come from GIS datasets provided by the Oregon Geospatial Enterprise Office (OGE) and the Washington Office of Financial Management (WA OFM). Data on roads come from the U.S. Census Bureau, and we include the distance to the nearest major road, where major roads are defined by the Census Bureau as interstate, U.S., state, or county routes. Figure 2 presents histograms of the distance of LCT sample plots to cities and roads. As expected, forest plots are generally farther from both cities and roads than agricultural plots, representing the fact that Northwest forest lands tend to be located in the mountains where few live, while agriculture tends to be in the valleys closer to cities.

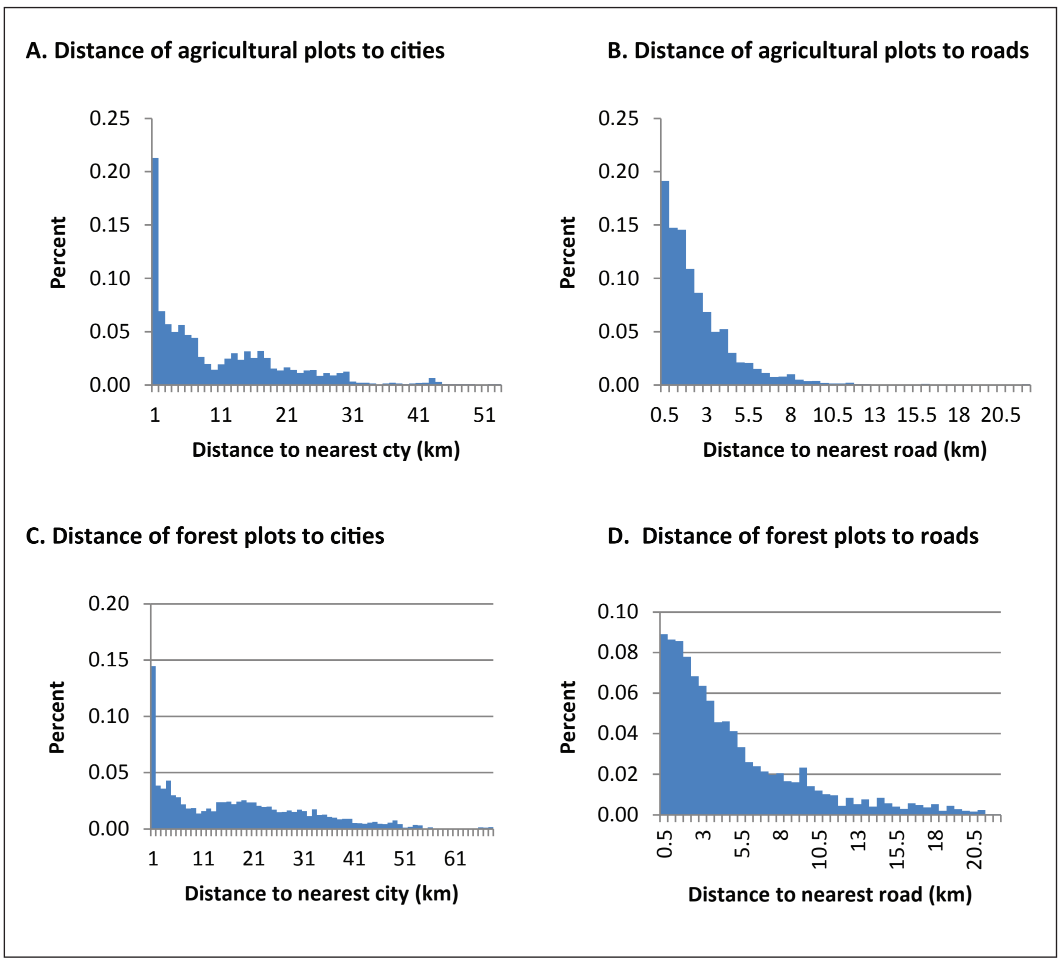


Figure 2—Distribution of the distance of sample Land Cover Trends plots to cities and roads.

Accounting for the LCT's Clustered Landscape Block Sampling Strategy

The LCT data are collected by first stratifying the continental United States into EPA level III ecoregions, and then randomly selecting a sample of 10- by 10-km blocks from each ecoregion. This brings up the issue of how to account for this clustered sampling technique in econometric estimation. We present a simple Monte Carlo exercise that illustrates properties of econometric estimation with a simple random sample of plots from counties, and compare that to a clustered sample similar to the LCT design. Results confirm that the LCT's sampling design creates a block-level unobservable that induces bias in estimation by generating spatial correlation in the error component of the econometric model. Results also confirm that consistent estimation can be achieved by including a block-level random effect in estimation. For simplicity, our results are illustrated with a binary plot-level model to develop or not.

A random sample of plots from counties—

This section describes a process for generating data used in the Monte Carlo exercise. The computer-generated data are developed to mimic properties of the LCT data. For plot *i* that begins period t in an undeveloped use u and ends period t in an undeveloped use, real annualized net value of the land is:

$$R_{iuut} = \alpha_{uu} + \beta_{uu} R_{c(i)ut} + \varepsilon_{iuut} \tag{2}$$

Real annualized net returns less conversion costs for developing parcel *i* into use d during period t are:

$$R_{iudt} - rC_{iudt} = \alpha_{ud} + \beta_{ud} R_{c(i)dt} + \varepsilon_{iudt} \tag{3}$$

For all uses k = 1,2,…..,K and time periods t = 1,….,T, where α_{jk} and β_{0jk} are parameters, $R_{C(i)kt}$ is the county average return to use k, and ε_{ijkt} is a standard normal unobservable. For simplicity, we have no land quality information for plot i, though that could be easily introduced. The plot will be developed if the net value of conversion (3) – (2) is positive:

$$(R_{iudt} - rC_{iudt}) - R_{iuut} = (\alpha_{ud} - \alpha_{uu}) + \beta_{ud} R_{c(i)dt} - \beta_{uu} R_{c(i)ut} + (\varepsilon_{iudt} - \varepsilon_{iuut}) > 0 \tag{4}$$

For simulation purposes, assume that the scale of (2) and (3) is such that both are plot-specific net returns written as deviations from the county average return. This is accomplished with $\alpha_{uu} = \alpha_{ud} = 0$, and $\beta_{uu} = \beta_{ud} = 1$:

$$(R_{iudt} - rC_{iudt}) - R_{iuut} = R_{c(i)dt} - R_{c(i)ut} + (\varepsilon_{iudt} - \varepsilon_{iuut}) \tag{5}$$

So, given a random sample of plots within each county, we can estimate a probit model of (5):

$$Prob(y_{it} = 1 | R_{c(i)dt} - R_{c(i)ut}) = Prob[(\varepsilon_{iudt} - \varepsilon_{iuut}) < \beta(R_{c(i)dt} - R_{c(i)ut})] \tag{6}$$

Consistent estimation should yield a sampling distribution of the estimate of β that is centered at one, the population parameter value of β.

Monte Carlo simulation steps for random sampling from counties—

1. Draw 100 county average net returns to conversion ($R_{c(i)dt} - R_{c(i)ut}$) from a U[0,1] distribution.[2]
2. Draw 30 plots per county, with ($\varepsilon_{iudt} - \varepsilon_{iuut}$) being drawn from a N(0,1). This establishes (5) as a probit model.
3. Create the dependent variable y_{it} to be equal to one if (5) is positive, and zero otherwise.
4. Estimate (6) as a probit model with one variable, the net county returns to conversion ($R_{c(i)dt} - R_{c(i)ut}$).
5. Repeat steps 1 through 4 many times.

Monte Carlo results—

Given the setup of the Monte Carlo simulation, the estimate of β should be equal to one when averaged over 1,000 independent simulations, where the 1,000 simulations approximate $\hat{\beta}$'s sampling distribution. Results confirm that $(1/S)\sum_{s=1}^{S} \hat{\beta}^s = 1$, where $S = 1{,}000$, indicating that a random sample of plots from each county can be used to provide consistent parameter estimates of this land-conversion model.

A random sample of plots from clustered blocks within counties—

The LCT dataset consists of a sample of 10- by 10-km blocks randomly sampled within ecoregions. Many counties have multiple blocks within them, and a block can straddle multiple counties. Importantly, sampling plots from LCT blocks no longer comprise a random sample from counties, so we explore the implications of such a sampling design for econometric estimation.

Begin by rewriting (5) as the plot-specific deviation from the block-specific average return to conversion, where b indexes blocks:

$$(R_{iudt} - rC_{iudt}) - R_{iuut} = (R_{c(i)dt} + \omega_{b(i)dt}) - (R_{c(i)ut} + \omega_{b(i)ut}) + (\varepsilon_{iudt} - \varepsilon_{iuut}) \tag{7}$$

Where ($R_{c(i)dt} + \omega_{b(i)dt}$) is the average developed return for block b, which contains plot i, and ($R_{c(i)ut} + \omega_{b(i)ut}$) is the average undeveloped return for block b, which contains plot i. Equation (7) can be arranged into the relevant latent form used for estimation:

$$(R_{iudt} - rC_{iudt}) - R_{iuut} = (R_{c(i)dt} - R_{c(i)ut}) + (\omega_{w(i)dt} - \omega_{w(i)ut}) + (\varepsilon_{iudt} - \varepsilon_{iuut}) \tag{8}$$

---

[2] A uniform distribution is standard for drawing values of independent variables in Monte Carlo analyses, though results are not contingent on this distribution.

Equation (8) highlights that the additional term $(\omega_{b(i)dt} - \omega_{b(i)ut})$ generates an unobservable that will be correlated across all plots within block *b*. All plots within block *b* share this same unobservable. Failure to account for such correlation could lead to inconsistent parameter estimates arising from a standard probit model. This econometric complication arises entirely from the fact that sampling plots from LCT blocks generates a random sample of plots within blocks, but not a random sample of plots within counties. Consistent estimation can be achieved by including a block-specific random effect to account for $(\omega_{b(i)dt} - \omega_{b(i)ut})$. A random effect can be used because $(\omega_{b(i)dt} - \omega_{b(i)ut})$ can be reasonably assumed to be independent of $(R_{c(i)dt} - R_{c(i)ut})$ since the LCT randomly samples blocks within ecoregions. Given a random sample of plots within each block, a probit model of the latent equation (8) is:

$$Prob(y_{it}=1|R_{c(i)dt}-R_{c(i)ut})=Prob[(\varepsilon_{iudt}-\varepsilon_{iuut})<\beta(R_{c(i)dt}-R_{c(i)ut})+\omega_{b(i)dt}-\omega_{b(i)ut}] \quad (9)$$

Consistent estimation should again yield a sampling distribution of $\hat{\beta}$ that is centered at one, the population parameter value of β.

Monte Carlo simulation steps for random sampling from landscape blocks—

1. Draw 100 county average net returns to conversion $(R_{c(i)dt} - R_{c(i)ut})$ from a U[0,1] distribution (see footnote 2).
2. Draw 100 block deviations from the county average net returns to conversion $(\omega_{b(i)dt} - \omega_{b(i)ut})$ from a U[-*a*, *a*] distribution, where parameter *a* will be varied over different scenarios. This step implicitly assumes that each county has one block.
3. Draw 30 plots per county, with $(\varepsilon_{iudt} - \varepsilon_{iuut})$ being drawn from a N(0,1) distribution. This establishes (8) as a probit model.
4. Create the dependent variable y_{it} to be equal to one if (8) is positive, and zero otherwise.
5. Estimate (9) as a probit model with one variable, the net county returns to conversion $(R_{c(i)dt} - R_{c(i)ut})$. This should generate inconsistent parameter estimates.
6. Estimate (9) as a probit model with one variable, the net county returns to conversion $(R_{c(i)dt} - R_{c(i)ut})$, and a block random effect. This should generate consistent parameter estimates by accounting for correlated unobservables across all plots within a block.
7. Repeat steps 1 through 6 many times.

Given the setup of the Monte Carlo simulation, the estimate of β should be equal to one when averaged over 1,000 independent simulations. Again, the 1,000 independent simulations form an approximation of $\hat{\beta}$'s sampling distribution. Results confirm that (1/S) $\sum_{s=1}^{S} \hat{\beta}^s = 1$ for the random effects model, indicating that a random sample of plots from each block can be used to provide consistent parameter estimates of this land-conversion model. However, failure to include a random effect generates inconsistent estimates, which are worse as the |a| value is larger, where 2a is the length of the interval through which the block unobservable was uniformly drawn. Intuitively, |a| is larger when block-specific mean returns to conversion are further from the county mean. Results from different values of a are found in table 5, and probit estimates without a random effect become increasingly biased away from the true value of one as the block average return to conversion becomes increasingly different from the county average return. Including block random effects generates consistent estimates. This Monte Carlo analysis supports the inclusion of LCT block random effects in the primary econometric model from equation (1).

Estimation Technique—Maximum Simulated Likelihood

There are three sets of fixed (not random) parameter vectors to be estimated in the basic econometric model of equation (1): a set of alternative specific constants for the transition from land-use j to k (α_{jk}), a set of parameters on the average returns to use k in time t for the county that contains parcel i (β_{0jk}), and a set of parameters on the interaction of county returns with observable land quality for parcel i in use k (β_{1jk}). One alternative specific constant must be normalized to zero, and so we always normalize the alternative specific constant on the starting use ($\alpha_{jj} = 0$), which gives the other constants the interpretation as a conversion cost. A fully specified model also must make distributional assumptions regarding the random parcel and block effects (ω_{ik} and $\varphi_{B(i)k}$, respectively). We assume both are normally distributed with zero mean. The parcel effects have standard deviation σ_{1jk}, and the block effects have standard deviation σ_{2jk}. Because we have three separate land uses (agriculture, forest, urban), there are three separate parcel effects and three separate block effects. Given our use of panel data, there is no need to normalize one transition-specific random effect to zero (Walker et al. 2007). Finally, assuming

Table 5—Monte Carlo simulation results of including block random effects to account for the Land Cover Trend's clustered sampling approach

Block deviation (-a to a)	Average probit estimate using block mean returns	Average probit estimate using county mean returns	Average probit estimate using county mean returns with block random effects
-0.25 to 0.25	1.001	0.989	1.002
-0.5 to 0.5	1.001	0.963	0.992
-0.75 to 0.75	1.001	0.914	1.006
-1 to 1	1.001	0.862	1.018
-1.25 to 1.25	1.001	0.809	1.001
-1.5 to 1.5	1.002	0.74	1.035

that ε_{ijkt} is an iid type I extreme value unobservable, then the probability that parcel i changes from use j to use k in time t is:

$$P_{ijkt} = \int\int L_{ijkt}\,(\alpha_{jk}, \beta_{0jk}\, \beta_{1jk}, \sigma_{1jk}, \sigma_{2jk}) f(\theta_{ik}) f(\mu_{B(i)k}) d\theta_{ik}\, d\mu_{B(i)k} \tag{10}$$

where

$$L_{ijkt}(\alpha_{jk}, \beta_{0jk}\, \beta_{1jk}, \sigma_{1jk}, \sigma_{2jk}) = \left[\frac{e^{\alpha_{jk} + \beta_{0jk} R_{C(i)kt} + \beta_{1jk} LQ_{ik} R_{C(i)kt} + \sigma_{1jk}\theta_{ik} + \sigma_{2jk}\mu_{B(i)k}}}{\sum k e^{\alpha_{jk} + \beta_{0jk} R_{C(i)kt} + \beta_{1jk} LQ_{ik} R_{C(i)kt} + \sigma_{1jk}\theta_{ik} + \sigma_{2jk}\mu_{B(i)k}}} \right] \tag{11}$$

And where θ_{ik} and $\mu_{B(i)k}$ are standard normal random variables. The probability in (10) is known as a mixed logit, or random parameters logit probability (Train 2009). The integrals in (10) do not have a closed form solution, and so parameters must be estimated via simulation of the log likelihood function. Such maximum simulated likelihood estimation employs the assumption that the parcel (ω_{ik}) and block ($\varphi_{B(i)k}$) random effects are independent and normally distributed, which allows us to solve (10) by repeatedly drawing from standard normal distributions for θ_{ik} and $\mu_{B(i)k}$, calculating (11), and averaging. Following this logic, and denoting D_B as the full set of observed land-use decisions in block B, the probability of the observed land-use decisions on parcel i at time t conditional on draws from θ_{ik}
and $\mu_{B(i)k}$ is simply (11); the probability of the **sequence of land-use choices** in D_B is thus:

$$\Pr(D_B) = \prod i{\in}B \prod t \prod k \left[\frac{e^{\alpha_{jk} + \beta_{0jk} R_{C(i)kt} + \beta_{1jk} LQ_{ik} R_{C(i)kt} + \sigma_{1jk}\theta_{ik} + \sigma_{2jk}\mu_{B(i)k}}}{\sum k e^{\alpha_{jk} + \beta_{0jk} R_{C(i)kt} + \beta_{1jk} LQ_{ik} R_{C(i)kt} + \sigma_{1jk}\theta_{ik} + \sigma_{2jk}\mu_{B(i)k}}} \right]^{y_{ijkt}} \tag{12}$$

where y_{ijkt} equals one if plot i converts from use j to use k during time t, and zero otherwise. The likelihood of the observed land-use behavior in block B can be simulated by drawing randomly from the independent standard normal distributions of θ and μ. Taking R sets of draws, with each set r denoting a single draw for each observation generates the approximated likelihood function:

$$Pr^{Sim}(D_B) = \tfrac{1}{R} \sum_{r=1}^{R} Pr(D_B^r) \tag{13}$$

The full simulated log-likelihood function over all B^T blocks in the study region is defined as

$$\sum_{b=1}^{B^t} log[Pr^{Sim}(D_B)] \tag{14}$$

Maximizing the function in (14) over the parameter set ($\alpha_{jk}, \beta_{0jk}\, \beta_{1jk}, \sigma_{1jk}, \sigma_{2jk}$) generates the estimated parameters for the land-use model. Two hundred independent Halton draws are used to draw from the random effect distributions. All estimation is done with original code written in Matlab.

Alternative Specifications and Parameter Estimates

Model 1: Base Random Parameters Logit (RPL)

We estimate the model for plots starting in agriculture separately from those plots starting in a forested use. Plots starting in urban do not change, so we do not estimate a model for those plots beginning in urban. For the agriculture-to-agriculture transition, the land quality variable is the LCC measure of the plot, and LCC categories are grouped as three dummies representing classes 1 and 2, classes 3 and 4, and classes 5 and above. The dummy representing classes 1 and 2 is omitted as the base category. For the agriculture-to-forest transition, LCC categories 5 through 8 are the base category, with a dummy for classes 1 through 4 included. For the forest-to-agriculture transition, the LCC categories 1 through 4 are the base category, and a dummy for LCC categories 5 through 8 is included. For the forest-to-forest transition, LCC categories 5 through 8 are the base category, with separate dummies for LCC 1 and 2 and LCC 3 and 4 included. This specification with grouped LCC categories reflects the lack of variation in transitions within these grouped categories. All LCC dummies are interacted with their respective county-average forest or agricultural returns. For both the agriculture-to-urban and the forest-to-urban transitions, land quality is included as two variables representing distance of the plot to the nearest city with 10,000+ people, and distance of the plot to the nearest road. Both distances are interacted with county-average urban returns and measured in thousands of kilometers.[3] One other specification note is that all block random effects are interacted with two time-dummies representing the 1986–1992 transition period and the 1992–2000 transition period, allowing us to test for temporal variation in the standard deviation of the block random effects.

The dataset includes 3,789 transition opportunities for plots that begin in agriculture and 7,489 transition opportunities for plots that begin in forest. Plots that never leave agriculture between 1980 and 2000 are included three times in the dataset for each of the three transition periods. Those that transition at some point before the final transition period are dropped in subsequent periods. Plots that convert to agriculture or forest after 1980 can also have fewer than three observations in the data. The same principles hold for plots that begin in forest. Of those observations that begin in agriculture, there are 405 observed urban development conversions and 102 observed forest conversions. Of those observations that begin in forest, there are 716 observed urban conversions and 170 agricultural conversions.

---

[3] Kline et al. (2003) used another approach that weights the distance from cities with population, creating a gravity index. Our approach has the advantage of explicitly weighting distance by the economic determinant of urban development (net returns), though a fruitful future research approach could explore combining our approach with a gravity index.

All parameter estimates are presented in table 6. The parameters generally conform to expectations, with positive signs on most of the parameters on land-use returns. Further, the marginal effect on forest returns generally falls for land that is of different quality than LCC 5 through 8 while the marginal effect on agricultural returns falls for agricultural parcels of lower land quality. We see the effects of location on urban development probabilities as urban returns will have smaller marginal effects for plots farther from cities and roads. Conclusions about statistical significance for individual variables should not be drawn by parameter estimates in highly nonlinear models such as this one. A more complete analysis of statistical significance will be accomplished in the "Marginal Effects" section where we

Table 6—Parameter estimates for model 1

	Starting in agriculture				Starting in forest		
	B	SE	t		B	SE	t
Agriculture choice:				Agriculture choice:			
Ag returns	3.59	2.53	1.42	Constant	-5.15	0.27	-19.18
Ag returns * LCC 3,4	1.00	1.40	0.72	Ag returns	2.82	1.58	1.79
Ag returns * LCC 5-8	5.18	2.31	2.24	Ag returns * LCC 3,4	-1.02	2.04	-0.50
Forest choice:				Forest choice:			
Constant	-4.51	0.32	-14.31	Forest returns	22.87	3.66	6.25
Forest returns	8.31	6.25	1.33	Forest Ret*LCC 3,4	-17.01	3.06	-5.56
Forest ret * LCC 1-4	-15.28	5.92	-2.58	Forest Ret*LCC 1,1	-20.79	4.10	-5.06
Urban choice:				Urban choice:			
Constant	-3.01	0.34	-8.85	Constant	-3.98	0.20	-20.40
Urban returns	0.15	0.03	4.47	Urban returns	0.26	0.02	12.84
Urb ret*city dist	-21.76	2.72	-7.99	Urb Ret*City Dist	-9.37	1.24	-7.55
Urb ret*road dist	-0.24	0.64	-3.75	Urb Ret*Road Dist	-0.06	0.03	-2.15
Random effects standard deviations:				Random effects standard deviations:			
Ag parcel	0.41	0.10	4.28	Ag parcel	0.33	0.13	2.52
Forest parcel	0.01	0.23	0.03	Forest parcel	0.15	0.09	1.62
Urban parcel	0.44	0.09	4.71	Urban parcel	0.39	0.07	5.39
Ag block	2.30	0.29	7.94	Ag block	2.22	0.20	10.97
For block	1.57	0.29	5.40	For block	0.68	0.12	5.52
Urban block	1.57	0.23	6.81	Urban block	0.72	0.12	5.84
Ag block * d8692	1.05	0.37	2.88	Ag block * d8692	0.48	0.19	2.58
For block * d8692	1.15	0.32	3.62	For block * d8692	0.38	0.13	2.89
Urb block * d8692	0.89	0.31	2.86	Urb block * d8692	0.60	0.16	3.87
Ag block * d9200	0.70	0.31	2.27	Ag block * d9200	0.91	0.44	2.07
For block * d9200	0.55	0.50	1.10	For block * d9200	1.05	0.16	6.68
Urb block * d9200	0.31	0.34	0.92	Urb block * d9200	0.88	0.16	5.44
Log likelihood	-1293.38			Log likelihood	-2391.4		
N	3,789			N	7,489		

Notes: All returns are measured in thousands of dollars. All parameters with t-stats above 1.96 are significantly different from zero at the 5 percent level.

account for the nonlinear structure of the model and evaluate marginal effects at each plot. The parameters also reveal substantial unobserved heterogeneity in that most of the standard deviations for the various random effects are significantly different from zero at the 5 percent level. There is also evidence that the block random effects change over the various transition periods.

The model includes an estimated urban development gradient by interacting distance to nearest city with county urban returns. However, there are two major regulatory policy changes during the time period of this study that might be expected to influence the urban development gradient. First, Oregon's well-known urban growth boundaries were first instituted by about 1980, with an explicit goal of concentrating development in nearby cities. Second, Washington's Growth Management Act of 1990 also specified the use of urban growth boundaries, and was implemented in most cities during the mid-1990s. As such, one might expect the urban development gradient to differ across the two states, at least prior to the 1992–2000 transition period. This possibility is explored by including an interaction between the urban development gradient and a dummy representing whether the parcel is in Oregon. We then introduce interactions between the urban development gradient and two separate dummies representing the 1986–1992 and 1992–2000 transition periods. We also introduce three-way interactions between the Oregon dummy, the urban development gradient, and the same two separate dummies representing the 1986–1992 and 1992–2000 transition periods. As such, this flexible specification allows the urban development gradient to vary across states and across transition periods. Likelihood ratio tests of these five additional parameters fail to reject the null hypothesis that they are jointly zero (5 percent level) for both the agriculture and forest models. As such, there is no evidence that Oregon and Washington had significantly different development gradients over the various transition periods in the data.

Model 2: RPL With Agglomeration Economies

The second model we estimate adds variables representing an explicit spatial dependency:

$$R_{C(i)Bkt} - rC_{ijkt} = \alpha_{jk} + \beta_{0jk} R_{C(i)kt} + \beta_{1jk} LQ_{ik} R_{C(i)kt} + \beta_{2jk} LU_{B(i)kt} + [\omega_{ik} + \varphi_{B(i)k} + \varepsilon_{ijkt}] \tag{15}$$

where $LU_{B(i)kt}$ represents the share of landscape block B in use k at time t – a measure of neighboring land use. The idea behind adding $LU_{B(i)kt}$ as an explanatory variable in equation (15) is to capture potential agglomeration economies induced by spatial externalities across land uses. The net revenue of converting to parcel i to use k in time t ($R_{C(i)Bkt} - rC_{ijkt}$) is hypothesized to be higher if there are more neighbors also in use k. Conversion of land to an alternative use is similar to the decision to adopt a new technology in that there is likely a fixed cost of learning the new technique (Lewis et al. 2011). The fixed costs of converting to agriculture or forest will be lower if there are more close neighbors who have demonstrated what crops, tree species, or management techniques work in the many microclimates around the Pacific Northwest. For example, wine and hop production work best with certain microclimates and the prevailing direction of slope, and the presence of neighboring wine and hop production provides an avenue for learning how to manage the new land use. Inclusion of $LU_{B(i)kt}$ in (15) for k = agriculture or forest can thus be thought of as capturing the fixed costs of learning the new use.

Measures of neighboring urban land used as explanatory variables in parcel-scale econometric models dates back to Irwin and Bockstael (2002). In their model of development, Irwin and Bockstael (2002) postulated that urban development creates a negative externality that lowers the likelihood of development on neighboring parcels, and they found supporting evidence in an exurban region of Maryland. However, it is also possible that having more neighboring developed parcels induces positive externalities that increase the likelihood of development on a particular parcel. For example, some may be reluctant to live "in the woods" with no close neighbors. As another example, the presence of more developed neighbors likely lowers the cost of extending public utilities such as sewer to a new housing development as construction of such utilities tends to be associated with high fixed costs. In any event, inclusion of $LU_{B(i)kt}$ in equation (15) allows for an empirical test of the type of spatial externalities associated with agglomeration economies. Although the correct neighborhood size for specifying agglomeration economies is unclear, and not clearly testable, the LCT blocks themselves provide reasonable approximations of the neighborhood in which agglomeration economies operate. Alternative sizes (e.g., 1000-m radius, 100-m radius, etc.) are not implementable for plots near the edge of LCT windows because land use is not observed outside the windows. Below, we discuss a random-parameters specification of the parameter β_{2jk} to account for the fact that each plot within block B is in a different location.

Identification Strategy With Agglomeration Economies

Building off Manski's (1993) analysis of general social interactions, Irwin and Bockstael (2002) argue that measures of neighboring land use are necessarily endogenous in an econometric land-use model. In our case, endogeneity bias would arise because the neighboring land-use variable $LU_{B(i)kt}$ is necessarily correlated with the unobservable block-level random effect $\varphi_{B(i)k}$. For example, when k = urban, $\varphi_{B(i)k}$ captures unobserved urban amenities and disamenities such as protected open-space, school quality, scenery, and other local public goods. However, when $\varphi_{B(i)k}$ is high, signaling a landscape block with many local public goods, there will likely be more development, and so $LU_{B(i)kt}$ is necessarily correlated with $\varphi_{B(i)k}$ when k = urban. Likewise, when k = forest or agriculture, $\varphi_{B(i)k}$ picks up the micro-climate effects of the Pacific Northwest that influence crop and timber types, and crop and timber yields. When $\varphi_{B(i)k}$ is high, signaling an attractive microclimate for use k, there will be more neighboring land in that use, and so $LU_{B(i)kt}$ is necessarily correlated with $\varphi_{B(i)k}$ when k = agriculture or forest. The urban economics literature on identifying agglomeration economies makes essentially this same argument: geographic concentration of an industry does not necessarily signal the presence of agglomeration economies because certain geographic regions have inherent natural advantages for that industry (Ellison and Glaeser 1997).

We adopt the recently developed control function strategy for identifying the effects of neighboring land-use ($LU_{B(i)kt}$) on land-use change. We follow Train (2009: chap. 13) in outlining the control function approach for our model. The control function works by specifying the endogenous land-use share variable $LU_{B(i)kt}$ as a function of observed instruments and unobserved factors, where the notation with parcel i is suppressed for simplicity:

$$LU_{Bkt} = \mathrm{W}(z_{Bk},\gamma) + \boldsymbol{\vartheta}_{Bkt} \tag{16}$$

where the block random effect φ_{Bk} and $\boldsymbol{\vartheta}_{Bkt}$ are assumed to be uncorrelated with the vector of instruments z_{Bk}, while there is assumed to be correlation between φ_{Bk} and $\boldsymbol{\vartheta}_{Bkt}$. It is this correlation between φ_{Bk} and $\boldsymbol{\vartheta}_{Bkt}$ that induces the concern that block-level land-use shares LU_{Bkt} are endogenous in (15). The control function approach works when the distribution of φ_{Bk} conditional on $\boldsymbol{\vartheta}_{Bkt}$ takes a convenient form. Following Train (2009: chap. 13), we decompose φ_{Bk} into its mean conditional on $\boldsymbol{\vartheta}_{Bkt}$ and deviations around this mean:

$$\varphi_{Bk} = E(\varphi_{Bk} \,|\, \boldsymbol{\vartheta}_{Bkt}) + \widetilde{\varphi_{Bk}} = \lambda\boldsymbol{\vartheta}_{Bkt} + \widetilde{\varphi_{Bk}} \tag{17}$$

Deviations $\widetilde{\varphi_{Bk}}$ are not correlated with $\boldsymbol{\vartheta}_{Bkt}$ by construction, and so therefore not correlated with the block-level land-use share variable LU_{Bkt}. The term $E(\varphi_{Bk} \,|\, \boldsymbol{\vartheta}_{Bkt})$ is known as the control function, and we follow Train in setting up the simplest specification of the control function $E(\varphi_{Bk} \,|\, \boldsymbol{\vartheta}_{Bkt}) = \lambda\boldsymbol{\vartheta}_{Bkt}$. Substituting (17) into (15) yields:

$$\alpha_{jk}+\beta_{0jk}\, R_{C(i)kt}+\beta_{1jk}\, LQ_{ik}\, R_{C(i)kt}+\beta_{2jk}\, LU_{B(i)kt}+[\omega_{ik}+\lambda\boldsymbol{\vartheta}_{B(i)kt}+\widetilde{\varphi_{B(i)k}}+\varepsilon_{ijkt}] \tag{18}$$

where the control function is simply an additional independent variable, and $\widetilde{\varphi_{B(i)k}}$ is the new block-level random effect, which is uncorrelated with the block level land-use share $LU_{B(i)kt}$ by construction. It is this transformation of the original block-level random effect (φ_{Bk}) into an observable $\lambda\boldsymbol{\vartheta}_{B(i)kt}$ and an unobservable $\widetilde{\varphi_{B(i)k}}$ that corrects for the endogeneity of block-level land-use shares.

Implementation of this model requires two steps. First, equation (16) is estimated by ordinary least-squares with the following block-level instruments: the shares of the block in the various soil quality classes (LCC), the distance of the centroid of block B to the nearest city greater than 10,000+ people, and the distance of the centroid of block B to the nearest major road. These variables are arguably exogenous in that they should only influence land-use decisions on parcel i through their effects on block-level land-use shares. Second, residuals from the first stage are calculated as estimates of $\vartheta_{B(i)kt}$ and included as separate regressors into maximum simulated likelihood estimation of (18). The remaining block-level random effect $\widetilde{\varphi_{B(i)k}}$ is then uncorrelated with $LU_{B(i)kt}$, and the endogeneity problem is solved. As one final specification note, we specify the parameter β_{2jk} on the land-use share variables $LU_{B(i)kt}$ as random parameters at the parcel level—formally denoted β_{i2jk}—to account for the fact that different plots are at different geographical locations within each block.

Evaluating the Instruments in the Control Function Approach

Similar to standard instrumental variables estimation in linear models, two features make a good instrument in the control function approach to nonlinear models. First, the instruments must be correlated with the endogenous variables representing block-level land-use shares. The first stage of the control function approach is to regress observed land-use shares at the block level at four points in time (1973, 1980, 1986, 1992) on the set of instruments representing exogenous land quality at the block level. Table 7 presents ordinary least squares (OLS) parameter estimates for three different models, one for each land-use category. Results generally conform to expectation, as landscape blocks farther from cities and roads have smaller shares in urban. Likewise, soil quality is particularly important for forest and agriculture shares, as blocks with higher shares of low-quality soil have more forest and less agriculture. The respective R^2 statistics indicate that the block-level land quality measures explain more variation in block-level forest and agricultural lands than they do block-level urban. In total, we would reject the null that all parameters are jointly zero in each model (5 percent level), indicating that these instruments are correlated with observed block-level land-use shares.

The second characteristic of good instruments in this model is that they must be uncorrelated with the main block-level random-effect from the base model in equation (15) – φ_{Bk}. This characteristic is more difficult to evaluate with data as the random effect is an unobservable. Therefore, we state clearly what our primary identifying assumptions are for the control function to generate consistent parameter estimates. The maintained identifying assumption is that it is parcel i's

Table 7—Parameter estimates from regressing block-level land-use shares on instruments

	B	Se	t
Dependent variable = block urban share:			
Constant	0.13	0.05	2.66
LCC 3 or 4 share	0.08	0.07	1.15
LCC 5 or 6 share	-0.07	0.05	-1.37
LCC 7 or 8 share	-0.09	0.04	-2.10
City dist centroid	-0.74	0.20	-3.73
Road dist centroid	-4.26	0.79	-5.39
R^2	0.28		
Dependent variable = block forest share:			
Constant	-0.01	0.04	-0.38
LCC 3 or 4 share	0.59	0.05	11.61
LCC 5 or 6 share	0.98	0.06	16.75
LCC 7 or 8 share	0.95	0.07	13.99
City dist centroid	-2.17	0.89	-2.45
Road dist centroid	2.44	2.06	1.19
R^2	0.50		
Dependent variable = block Ag share:			
Constant	0.81	0.06	13.92
LCC 3 or 4 share	-0.62	0.07	-8.66
LCC 5 or 6 share	-0.75	0.06	-11.79
LCC 7 or 8 share	-0.85	0.06	-14.14
City dist centroid	-0.85	0.39	-2.15
Road dist centroid	0.20	1.29	0.15
R^2	0.50		
N	508		

LCC rating that affects the land-use decision on parcel *i*, and not the LCC ratings of other parcels in the landscape block in which *i* resides. Likewise, it is parcel *i*'s distance from cities and roads that affects the decision to develop parcel *i*, and not the distance from roads and cities of the centroid of the landscape block in which parcel *i* resides. In general, a good (excluded) instrument affects the dependent variable indirectly through the endogenous variable, and should not itself be a direct independent variable in the primary estimation equation. This description fits our particular application, as neighboring land quality should influence a parcel's land-use decision (the dependent variable) only indirectly through its effect on neighboring land-use shares (the endogenous variable).

Model 2 Results

All parameter estimates for model 2 with agglomeration economies are presented in table 8. Given that we include independent variables that are estimated residuals from the first-stage model, standard errors must be adjusted to account for this extra source of variation. We use the bootstrap approach developed by Petrin and Train (2002) for use with control function estimation. In particular, we repeatedly estimate the primary discrete-choice land-use model with bootstrapped samples of the first-stage residuals. The variance in parameter estimates over the bootstrapped samples is added to the traditional variances, and the total adjusted standard errors are presented in table 8. This approach of repeatedly estimating a maximum simulated likelihood model is extremely computationally intensive.

The parameters generally conform to expectations, with positive signs on most of the parameters on land-use returns. Further, the probability of converting to forest generally falls for land that is of different quality than LCC 5 through 8 while the probability of converting to agriculture is slightly higher for agricultural parcels of lower land quality. We again see the effects of location on urban development probabilities as urban returns will have smaller marginal effects for plots farther from cities and roads. A likelihood ratio test rejects the null hypothesis that the additional variables added from model 1 to model 2 are jointly zero at the 5 percent level, providing some evidence that neighboring land-use shares affect land-use transition probabilities. Again, conclusions about statistical significance for individual variables should not be drawn by parameter estimates in highly nonlinear models such as this one. A more complete analysis of statistical significance will be accomplished in the "Marginal Effects" section where we account for the nonlinear structure of the model and evaluate marginal effects at each plot. The parameters also reveal substantial unobserved heterogeneity in that most of the estimated standard deviations for the various random effects are significantly different from zero at the 5 percent level. There is also evidence that the block random effects change over the various transition periods, and that the parameters on the land-use share variables are random and not fixed.

Marginal Effects

In this section, we examine the statistical significance of some of the individual variables by calculating marginal effects, along with their standard errors. All marginal effects are calculated for a discrete change in variable x by calculating the probability of a land-use transition with the change, x + c, less the probability of the land-use transition without the change, where c is the change. The probability of the land-use transition—equation (10)—is calculated by simulation. Standard errors for the marginal effects are calculated using the Krinsky and Robb (1986)

Table 8–Parameter estimates for model 2 (with agglomeration economies)

	Starting in agriculture				Starting in forest		
	B	Adj SE	Adj t		B	Adj SE	Adj t
Agriculture choice:				Agriculture choice:			
Ag returns	4.98	2.63	1.89	Constant	-6.43	0.92	-7.02
Ag returns * LCC 3,4	0.97	1.54	0.63	Ag returns	0.69	1.71	0.41
Ag returns * LCC 5-8	4.36	2.40	1.81	Ag returns * LCC 3,4	0.38	2.21	0.17
Forest choice:				Forest choice:			
Constant							
Forest returns							
Forest ret*LCC 1-4	-6.10	1.19	-5.13	Forest returns	16.12	4.58	3.52
	6.32	7.33	0.86	Forest ret*LCC 3,4	-15.81	3.39	-4.67
	-13.08	6.38	-2.05	Forest ret*LCC 1,1	-18.54	4.80	-3.86
Urban choice				Urban choice			
Urban returns	-4.99	1.01	-4.94	Constant	-5.89	0.89	-6.65
Urb ret*cty dist	0.12	0.04	2.85	Urban returns	0.28	0.03	10.28
Urb ret*road dist	-9.75	3.41	-2.86	Urb ret*city dist	-9.95	1.86	-5.36
Urban returns	-0.26	0.07	-3.87	Urb ret*road dist	-0.07	0.04	-1.82
Land use shares:				Land use shares:			
Ag share	0.25	1.08	0.23	Ag share	5.32	1.55	3.43
Forest share	3.38	1.87	1.81	Forest share	0.84	1.17	0.72
Urban share	13.87	6.36	2.18	Urban share	5.31	4.33	1.23
Ag residuals	2.11	1.17	1.80	Ag residuals	1.29	1.43	0.91
Forest residuals	-5.66	2.03	-2.78	Forest residuals	0.90	0.98	0.92
Urban residuals	-13.14	6.78	-1.94	Urban residuals	-1.43	5.03	-0.29
Random effects standard deviations:				Random effects standard deviations:			
AgP arcel	0.08	0.09	0.83	Ag parcel	0.37	0.13	2.83
Forest parcel							
Urban parcel							
Ag block	0.31	0.20	1.58	Forest parcel	0.11	0.07	1.58
	0.08	0.10	0.78	Urban parcel	0.01	0.08	0.07
	2.75	0.33	8.42	Ag block	1.85	0.30	6.24
For block	1.45	0.33	4.35	For block	1.57	0.16	9.56
Urban block	0.69	0.29	2.36	Urban block	0.96	0.27	3.53
Ag block * d8692	0.97	0.36	2.69	Ag block * d8692	0.85	0.26	3.28
For block * d8692	0.71	0.29	2.41	For block * d8692	0.55	0.19	2.93
Urb block * d8692	1.30	0.34	3.86	Urb block * d8692	0.55	0.19	2.94
Ag block * d9200	0.32	0.33	0.97	Ag block * d9200	0.12	0.47	0.26
For block * d9200	1.40	0.54	2.62	For block * d9200	0.58	0.17	3.40
Urb block * d9200	0.72	0.43	1.68	Urb block * d9200	1.57	0.31	5.08
Ag share	1.45	0.40	3.65	Ag share	1.44	0.56	2.60
Forrest share	0.06	0.57	0.11	Forrest share	3.79	0.19	19.65
Urban share	1.22	0.54	2.27	Urban share	2.31	0.56	4.11
Log likelihood	-1253.7			Log likelihood	-2034.23		
N	3,789			N	7,489		

Notes: All returns are measured in thousands of dollars. All parameters with t-stats above 1.96 are significantly different from zero at the 5 percent level. Standard errors are bootstrapped.

simulation method. In particular, a parameter vector is drawn from the estimated distribution to calculate the estimated land-use transition probabilities in equation (10). The simulated parameter vector is equal to $\beta_s = \hat{\beta} + C'x_k$, where $\hat{\beta}$ is the estimated parameter vector for all parameters (including random effects), C is the $k \times k$ Cholesky decomposition of the estimated variance-covariance matrix, and x_k is a K-dimensional vector of draws from a standard normal distribution. Performing this simulation 1,000 times allows computation of the standard deviation of the 1,000 simulated marginal-effect estimates to generate a standard error for marginal effects.

Figures 3 through 5 present marginal effects for a $100 increase in each of the land-use returns for model 1 (no agglomeration economies), **calculated at the value of independent variables of each sample point**. The figures are scatter plots, where each point on the graph indicates an estimated marginal effect (and corresponding z-statistic) for a sample point. The effect of the model's nonlinearity comes through when examining the various patterns of these marginal effects. In general, higher forest returns increase the probability of remaining in forest while

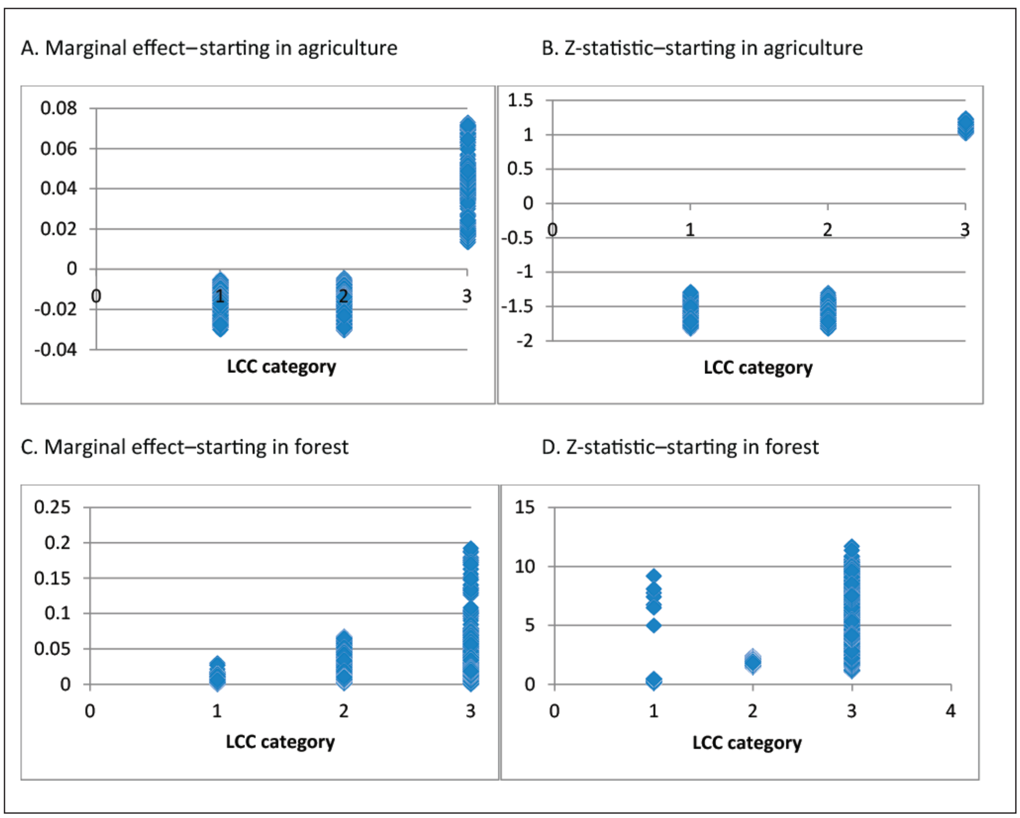


Figure 3—Marginal effects of a $100 increase in forest returns for model 1 on choosing forest. Land Classification Categories (LCC) are as follows: 1 = LCC 1 or 2; 2 = LCC 3 or 4; and 3 = LCC 5+.

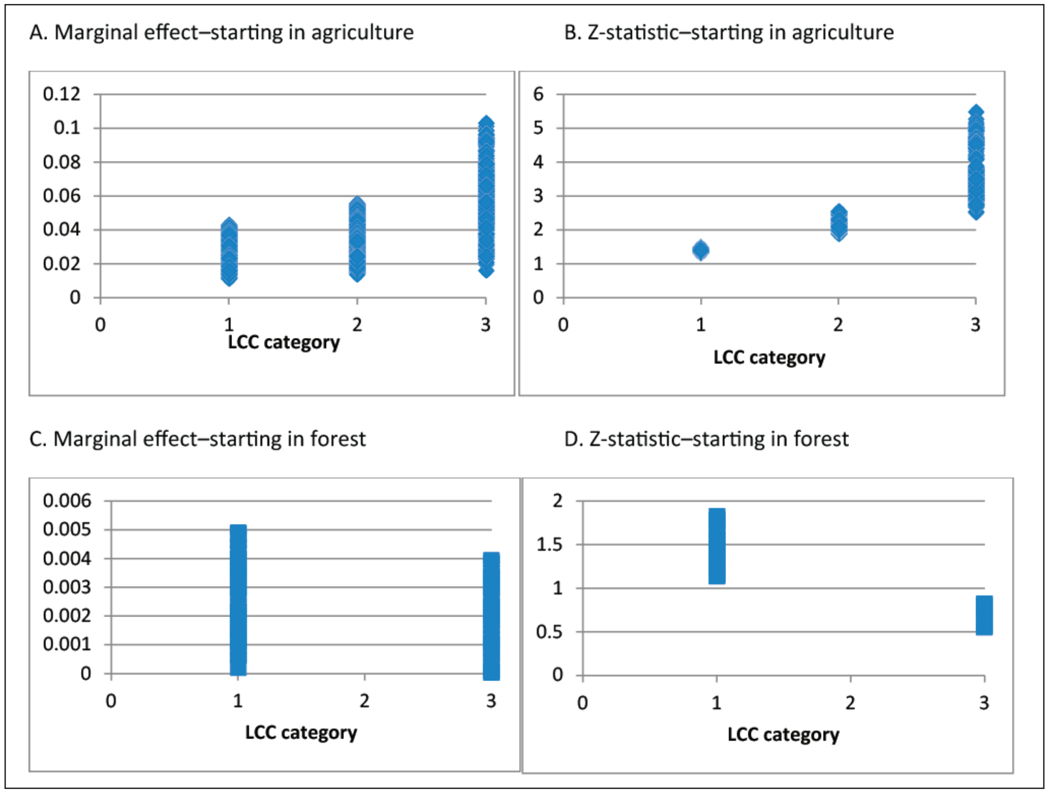


Figure 4—Marginal effects of a $100 increase in agricultural returns for model 1 on choosing agriculture. Categories (LCC) are as follows: 1 = LCC 1 or 2; 2 = LCC 3 or 4; and 3 = LCC 5+.

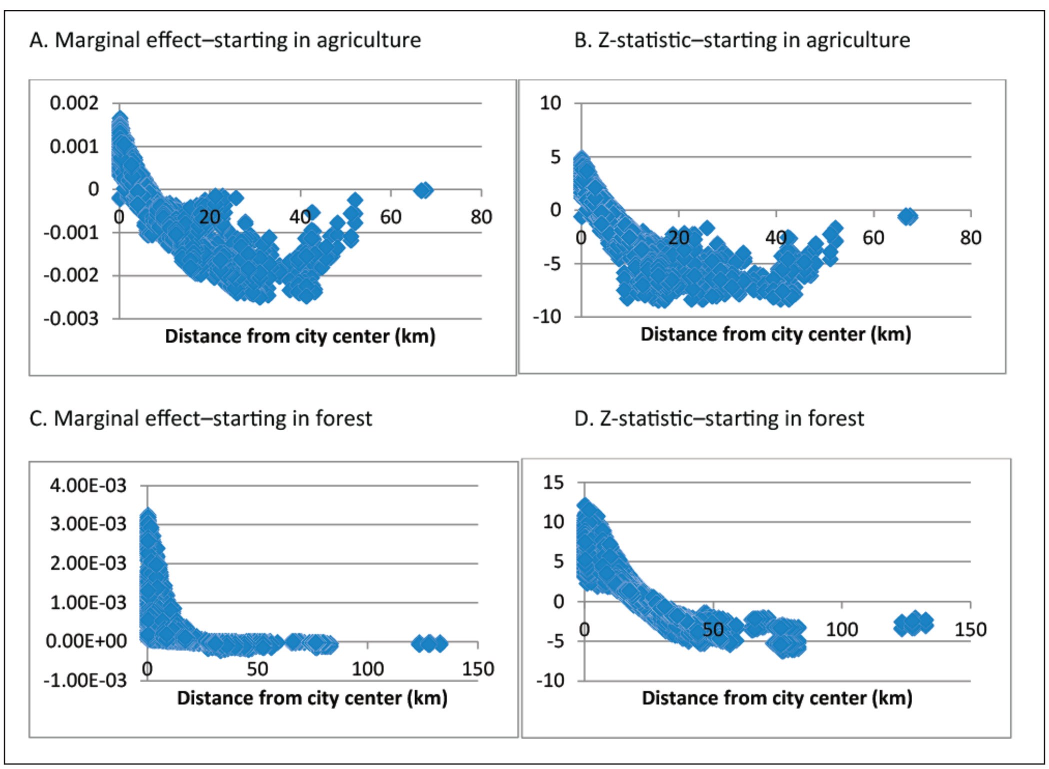


Figure 5—Marginal effects of a $100 increase in urban returns for model 1 on choosing urban.

higher agricultural returns increase the probability of remaining in agriculture in statistically significant manners. The magnitude of the marginal effects is generally higher for lower quality LCC classes, although there is significant heterogeneity in these results. Higher forest returns have no significant effect on converting from agriculture to forest, whereas higher agricultural returns generally have no significant effect on converting from forest to agriculture. This lack of significance likely comes from the fact that very few parcels are observed to transition between forest and agriculture in the Pacific Northwest, so there is likely not enough variation to estimate these effects with precision. A clear urban development gradient emerges from this model as higher urban returns typically have a statistically significant positive effect on the probability of urban development for those plots near cities, with no significant effect or even negative effects for those plots far from cities.

Figures 6 through 8 present the same marginal effects for a $100 increase in each of the land-use returns for model 2 (with agglomeration economies), again **calculated at the value of independent variables of each sample point**. The pattern that emerges is very similar to model 1 without agglomeration economies, suggesting that inclusion of agglomeration economies has minimal implications for the

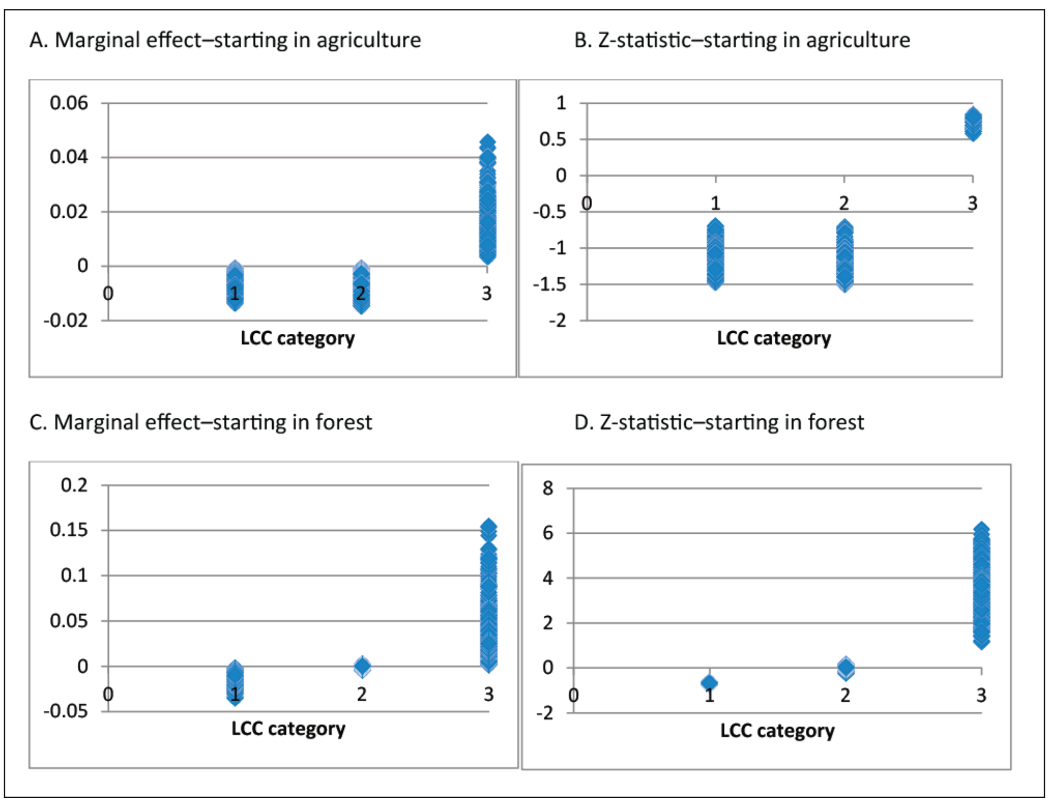


Figure 6—Marginal effects of a $100 increase in forest returns for model 2 on choosing forest. Categories (LCC) are as follows: 1 = LCC 1 or 2; 2 = LCC 3 or 4; and 3 = LCC5+.

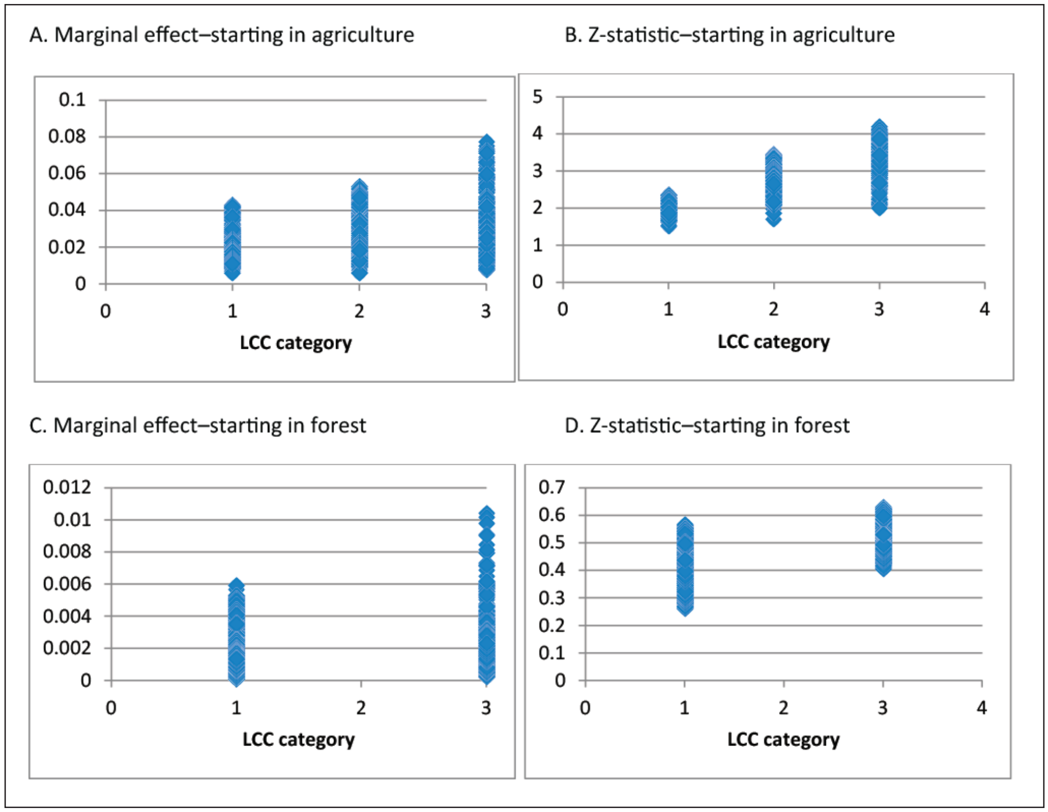


Figure 7—Marginal effects of a $100 increase in agriculture returns for model 2 on choosing agriculture.

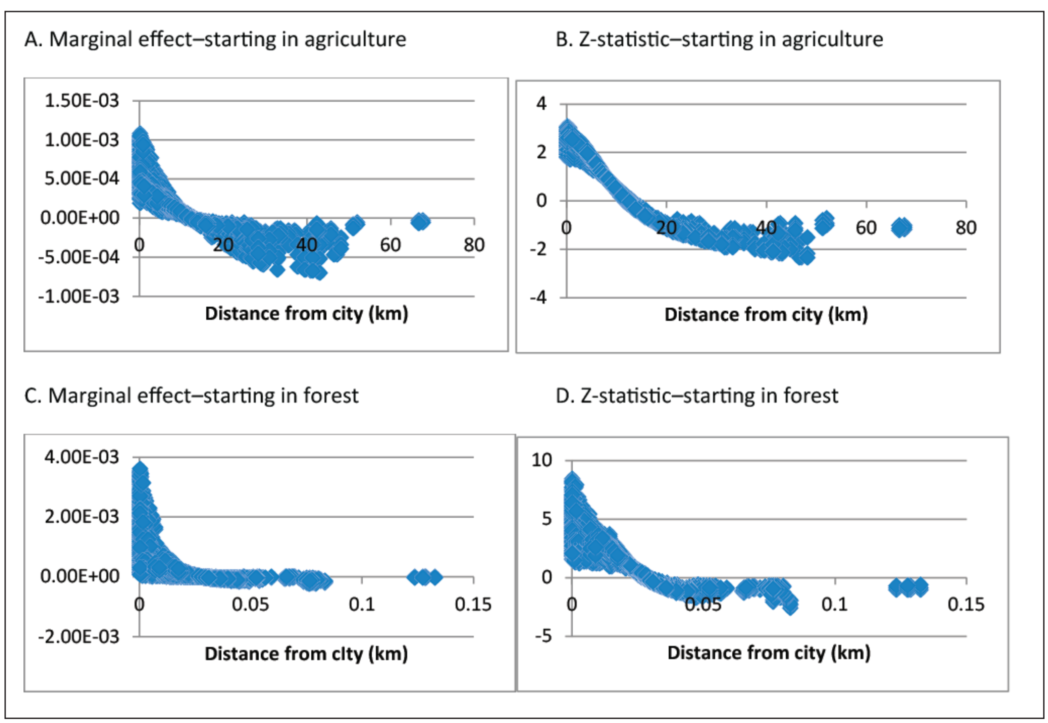


Figure 8—Marginal effects of a $100 increase in urban returns for model 2 on choosing urban.

marginal effects of the net returns variables. One must be cautious with this interpretation; however, as the generally positive signs on the land-use shares parameters implies a degree of path-dependence in land-use change that could cause long-term divergence in land-use patterns generated across the two models. We explore this in the "Landscape Projections and Out-of-Sample Forecast Test" section with full landscape forecasts over a 40-year time horizon.

Figure 9 examines the marginal effects of a 10 percentage point increase in the landscape block's land-use share in specific uses for lands starting in agriculture. Figure 9A examines the effects of more agricultural neighbors on the probability of agricultural land remaining in agriculture. No statistically significant effect is found for any observation. Figure 9C examines the effects of more forest neighbors on the probability of agricultural land converting to forest. There is some evidence for statistically significant agglomeration economies here, though it is weak. Figure 9E examines the effects of more urban neighbors on the probability of agricultural land converting to urban. Here there are generally consistent positive and statistically significant effects, which tend to increase in magnitude as the initial share of the block in urban development increases. Therefore, there appears to be strong evidence of urban agglomeration economies in the development of agricultural land, even when controlling for the average level of urban rents.

Figure 10 examines the marginal effects of a 10 percentage point increase in the landscape block's land-use share in specific uses for lands starting in forest. Figure 10A examines the effects of more agricultural neighbors on the probability of forest land converting to agriculture. We see strong evidence of positive agglomeration economies when converting forest land to agriculture. Figure 10C examines the effects of more forest neighbors on the probability of forest land remaining in forest. There is some evidence for statistically significant disagglomeration economies here, particularly for blocks that have sizable amounts of initial forest. These significant effects appear to be driven by the large estimated standard deviation of the random parameter on the forest share variable. Figure 10E examines the effects of more urban neighbors on the probability of forest land converting to urban. Here there are generally positive but statistically insignificant effects, which tend to increase in magnitude as the initial share of the block in urban development increases. One could use a one-tailed test to interpret some weak positive agglomeration economies for urban development on existing forest land.

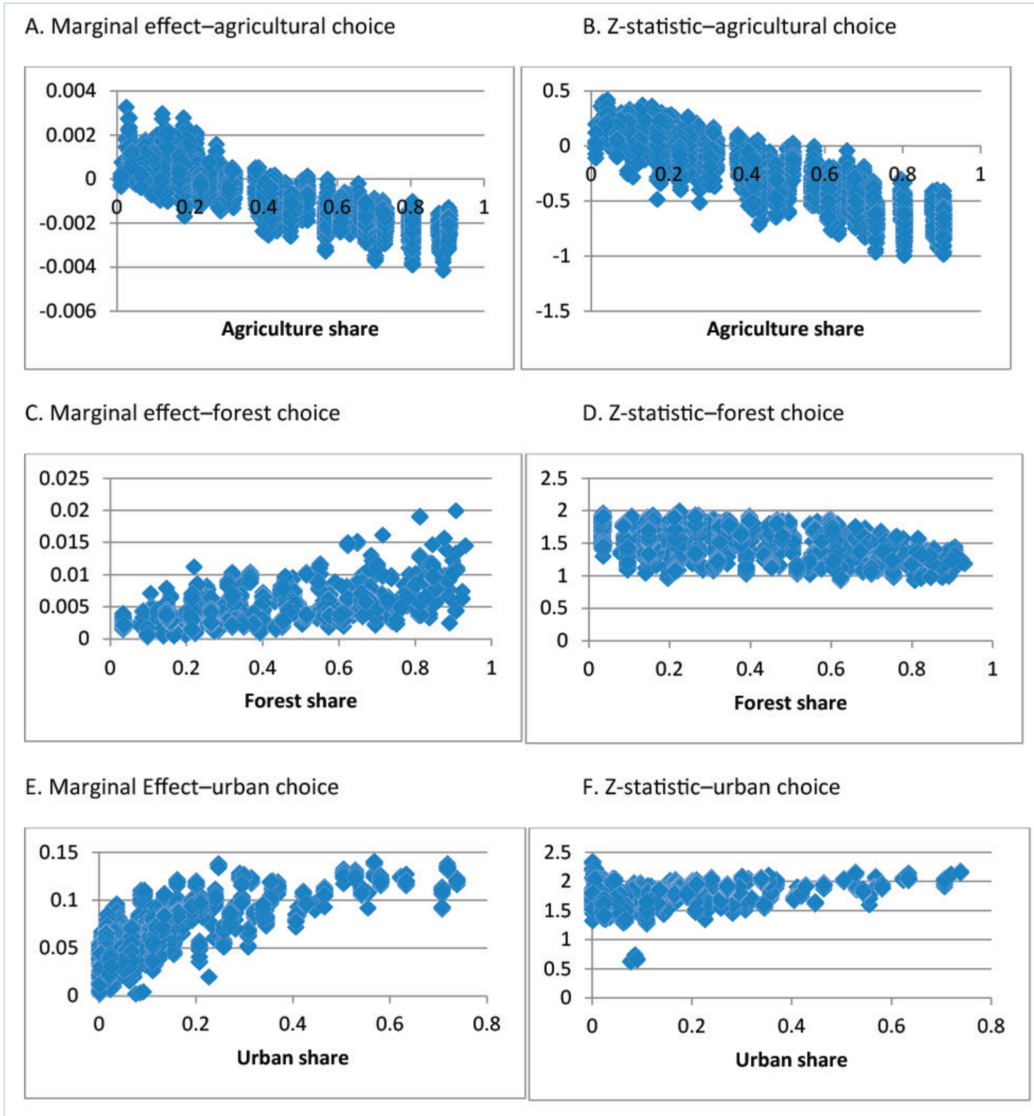


Figure 9—Marginal effects of a 10 percentage point increase in landscape block land-use share of the own use for model 2, lands starting in agriculture. Note: Figure 9A calculates the effect of an additional 10 percentage points in the block's share of agricultural land on the probability of remaining in agriculture, plotted against the initial block share of agriculture. Figure 9C calculates the effect of an additional 10 percentage points in the block's share of forest land on the probability of converting to forest, plotted against the initial block share of forest land. Figure 9E calculates the effect of an additional 10 percentage points in the block's share of urban land on the probability of converting to urban, plotted against the initial block share of urban.

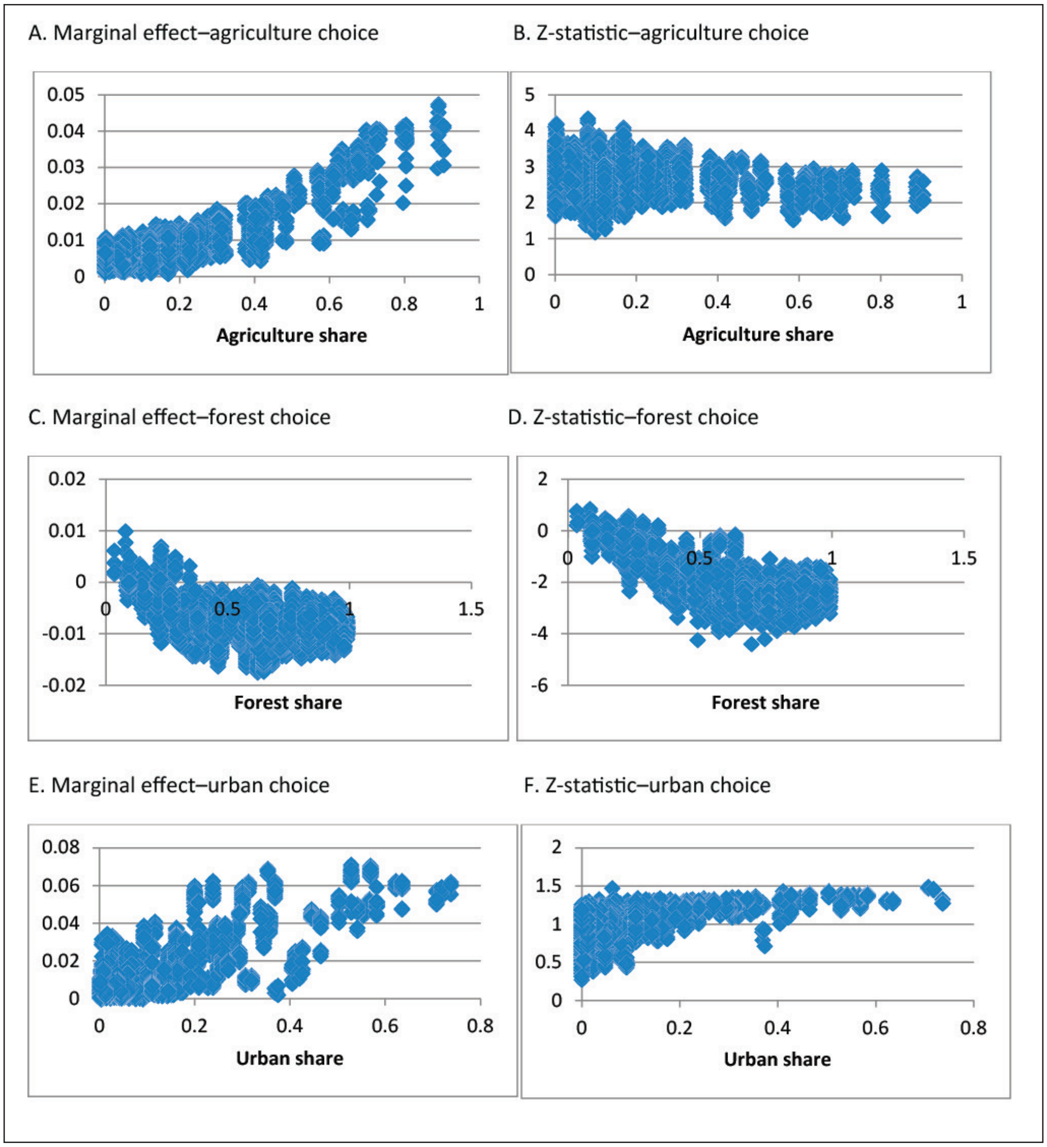


Figure 10—Marginal effects of a 10 percentage point increase in landscape block land-use share for model 2, lands starting in forest. Note: Figure 10A calculates the effect of an additional 10 percentage points in the block's share of agricultural land on the probability of converting agriculture, plotted against the initial block share of agriculture. Figure 10C calculates the effect of an additional 10 percentage points in the block's share of forest land on the probability of remaining in forest, plotted against the initial block share of forest land. Figure 10E calculates the effect of an additional 10 percentage points in the block's share of urban land on the probability of converting to urban, plotted against the initial block share of urban.

Landscape Projections and Out-of-Sample Forecast Test

With the estimated parameters for both models 1 and 2 in hand, we can project landscape-scale land-use change for the western halves of Oregon and Washington. Before doing so, we must adjust the estimated alternative-specific constants to correct for our oversampling of "change" pixels in the estimation sample. For standard logit estimation (rather than random parameters logit), each alternative-specific constant is adjusted by adding the log of the ratio of A_j to S_j, where A_j is the share of observations in the population who choose land-use *j,* and S_j is the share of the sample who choose land-use *j* (Train 2009: chap. 3). It is unclear if this method works for random parameters logit models with simulated probabilities, so we instead use the more general recalibration method for adjusting alternative-specific constants (see Train 2009: chap. 2.8 for this method). This method is used to adjust alternative-specific constants to match a specific forecast area or time period by using observed changes to recalibrate the constants.[4] This approach captures average unobserved effects for that time period or area into the alternative-specific constants by iterating the constants until the predicted land-use shares equal the actual land-use shares. We calibrate our forecasts to the 1992–2000 transition period in the LCT—the final observed period. The process works as follows. First, let S_{jk} denote the share of plots that transition from *j* to *k* during this period, and let α_{jk}^{0} be the original alternative-specific constant. Second, using the estimated discrete-choice model, predict the share of plots that transition from *j* to *k* and label them $\widehat{S_{jk}^{0}}$. Third, compare the predicted and actual shares and adjust the alternative-specific constant if needed. The adjustment $\alpha_{jk}^{1} = \alpha_{jk}^{0} + \ln(S_{jk}/\widehat{S_{jk}^{0}})$ guarantees that adjustment will always move the actual and predicted shares closer. All adjustment factors and corresponding predicted and actual shares are presented in table 9.

To examine model validity in forecasting, we perform an out-of-sample forecast test by comparing forecasted transition probabilities to actual transition probabilities on a separate sample of 16,000 LCT plots that began the 1992–2000 period in either agriculture or forest. The econometric model was used to forecast transition probabilities between 1992 and 2000. This approach follows Kline et al. (2003) in reserving data not used in estimation as a validation dataset. Table 10 presents average actual and estimated transition probabilities grouped by various land quality classes—soil quality and distance to the nearest city. Results show that the estimated and actual transition probabilities are reasonably close in values, thus providing some confidence as to the general validity of the estimated econometric models for projections.

---

[4] Alternatively, one can also adjust for such oversampling by weighting the likelihood function (see Lewis and Plantinga 2007).

Table 9—Adjustments of alternative specific constants to correct for oversampling "change" pixels—calibrated to 1992–2000 Land Cover Trend transition period

	Adjustment	Actual share (1992–2000)	Predicted share (1992–2000)
Model 1:			
Forest to agriculture	-2.9753	0.0015	0.0016
Forest to urban	-4.1146	0.0168	0.0173
Agriculture to forest	-4.4665	0.0039	0.0039
Agriculture to urban	-5.7132	0.0154	0.0154
Model 2:			
Forest to agriculture	-6.8140	0.0015	0.0016
Forest to urban	-4.4286	0.0168	0.0173
Agriculture to forest	-4.2617	0.0039	0.0039
Agriculture to urban	-5.4088	0.0154	0.0154

Table 10—Out-of-sample forecast test for 1992–2000 period

	Sample transition probabilities			Model 1 transition probabilities			Model 2 transition probabilities		
	Ag	Forest	Urban	Ag	Forest	Urban	Ag	Forest	Urban
Land starting in agriculture:									
Plots with LCC 1 or 2	0.9823	0.0024	0.0153	0.9824	0.0032	0.0144	0.9842	0.0023	0.0135
Plots with LCC 3 or 4	0.9803	0.0046	0.0151	0.9819	0.0041	0.0140	0.9823	0.0036	0.0141
Plots with LCC 5 to 8	0.9722	0.0040	0.0238	0.9788	0.0093	0.0119	0.9803	0.0088	0.0110
Plots ≤ 5 km from a city	0.9583	0.0053	0.0364	0.9724	0.0029	0.0247	0.9733	0.0037	0.0231
Plots > 5 km from a city	0.9931	0.0028	0.0041	0.9872	0.0052	0.0076	0.9884	0.0037	0.0079
Land starting in forest:									
Plots with LCC 1 or 2	0.0000	0.9708	0.0292	0.0023	0.9804	0.0172	0.0030	0.9775	0.0194
Plots with LCC 3 or 4	0.0043	0.9622	0.0335	0.0030	0.9756	0.0215	0.0019	0.9761	0.0220
Plots with LCC 5 to 8	0.0002	0.9955	0.0044	0.0013	0.9932	0.0055	0.0010	0.9927	0.0063
Plots ≤ 5 km from a city	0.0031	0.9373	0.0596	0.0026	0.9580	0.0394	0.0014	0.9537	0.0449
Plots > 5 km from a city	0.0016	0.9897	0.0088	0.0019	0.9909	0.0071	0.0015	0.9916	0.0069

LCC = Land Classification Class.

For future landscape projections, we now draw a **random** sample of 20,000 pixels from the population of private pixels in this region that begin in the year 2000 in either agriculture or forest to project future land-use change by ecoregion. This larger draw is desirable because we do not have the computational restrictions that we do under the estimation process, which takes days to run with the bootstrapping procedure. The larger draw also ensures we have good coverage of the distribution of land quality within each LCT landscape block. For each plot, the transition probabilities between forest, agriculture, and urban development are calculated. Next, average block-level transition probabilities are computed and multiplied with the relevant land-use shares for each LCT block to generate block-level land-use projections for an 8-year period. Transition probabilities are then updated to reflect the changed landscape and another 8-year period is projected. This process continues for a 40-year time horizon up to 2040.

A primary benefit of the econometric modeling framework is the ability to simulate changes in the net returns to land, where these changes could be driven by policy or other factors. We use the recent housing market boom and bust to project three different levels of urban returns. First, we project land use to 2040 and keep the real levels of all net returns at 1990s levels—specifically the average of 1990–1992. The second scenario uses the peak urban returns from the recent housing market boom, but does not alter agriculture or forest returns from 1990s levels. Because the Lubowski dataset does not include any returns estimates past the 1990s, we consult the widely used Case-Schiller housing market index, which had a peak value in 2006 at 184 percent of the 1992 level in real terms. Therefore, we multiply all urban returns by 1.84 in the second scenario. The third scenario uses a more modest urban return estimate by considering the 2012 Case-Schiller housing market index, which is 114 percent of the 1992 level in real terms. Therefore, we multiply all urban returns by 1.14 in the third scenario. This exercise allows us to explore land-use change scenarios within the full range of observed urban returns over the past 20 years.

Ecoregion urban land-use projections are presented in table 11. The ecoregion is a convenient reporting scale given that LCT blocks are random samples within ecoregions. Our projection is that there are very modest increases under all scenarios in urban land expected for the Coast Range, Cascades, and North Cascades regions. Little urban development exists in these mountainous regions, which are generally far from roads and cities. The Puget Lowland and Willamette Valley ecoregions are projected to see much more substantial increases in urban development, with much more variation across the scenarios. Projections for the Puget Lowland range from a 37 percent increase in urban development up to a

Table 11—Ecoregion urban land-use projections for 2040 under alternative urban returns scenarios (private land)

		Model 1		Model 2 (with agglomeration economies)	
	Urban share (2000)	Urban share (2040)	Change	Urban share (2040)	Change
Coast Range:	4.24		*Percent*		*Percent*
1992 urban return level		4.95	16.78	5.25	23.95
2012 urban return level		4.97	17.44	5.30	25.14
Peak (2006) urban return level		5.25	23.98	5.68	34.20
Puget Lowland:	22.27				
1992 urban return level		30.59	37.36	32.88	47.62
2012 urban return level		32.75	47.06	35.83	60.88
Peak (2006) urban return level		44.31	98.97	49.49	122.22
Willamette Valley:	13.18				
1992 urban return level		17.55	33.13	18.56	40.80
2012 urban return level		17.95	36.12	19.23	45.89
Peak (2006) urban return level		21.00	59.28	23.67	79.51
Cascades:	2.81				
1992 urban return level		3.45	22.66	4.09	45.48
2012 urban return level		3.48	23.75	4.12	46.58
Peak (2006) urban return level		3.77	34.10	4.40	56.72
North Cascades:	2.68				
1992 urban return level		4.24	58.53	4.50	68.16
2012 urban return level		4.43	65.39	4.66	73.94
Peak (2006) urban return level		6.01	124.57	5.90	120.19

122 percent increase under the peak 2006 urban return level. Likewise, projections for the Willamette Valley range from a 33 percent increase up to a 79 percent increase. Clearly, the real urban return level going forward has major implications for urban development in the Puget Lowland and Willamette Valley. Likewise, there is some variation across model 1 and model 2, with the path dependence feature of agglomeration economies in model 2 inducing slightly higher levels of urban development. However, the assumed urban return level has a much larger effect on urban projections than does the assumption about agglomeration economies.

Ecoregion forest and agricultural land-use projections are presented in tables 12 and 13. Similar to urban land use, very modest changes in forest and agricultural lands are expected for privately owned portions of the Coast Range, Cascades, and North Cascades. The Puget Lowland is expected to lose the most forest, and these projections are sensitive to the assumed urban return level. The most extreme loss would occur if urban returns jump back to peak 2006 levels. Projected forest losses

Table 12—Ecoregion forest land-use projections for 2040 under alternative urban returns scenarios (private land)

		Model 1		Model 2 (with agglomeration economies)	
	Forest share (2000)	Forest share (2040)	Change	Forest share (2040)	Change
Coast Range:	75.36		*Percent*		*Percent*
1992 urban return level		74.78	-0.77	74.54	-1.09
2012 urban return level		74.75	-0.82	74.49	-1.16
Peak (2006) urban return level		74.44	-1.22	74.12	-1.66
Puget Lowland:	52.52				
1992 urban return level		45.05	-14.22	44.53	-15.21
2012 urban return level		43.02	-18.08	42.12	-19.80
Peak (2006) urban return level		32.28	-38.53	31.26	-40.47
Willamette Valley:	35.49				
1992 urban return level		34.75	-2.08	33.96	-4.30
2012 urban return level		34.44	-2.96	33.58	-5.38
Peak (2006) urban return level		32.02	-9.76	31.22	-12.02
Cascades:	85.12				
1992 urban return level		84.29	-0.97	83.56	-1.82
2012 urban return level		84.25	-1.01	83.53	-1.87
Peak (2006) urban return level		83.95	-1.37	83.23	-2.22
North Cascades:	63.70				
1992 urban return level		61.61	-3.28	61.75	-3.07
2012 urban return level		61.42	-3.58	61.59	-3.32
Peak (2006) urban return level		59.82	-6.09	60.34	-5.29

for the Willamette Valley are much more modest. Agricultural losses as percentage points are highest for the more agricultural Willamette Valley, and are somewhat sensitive to the assumed urban return level. Because models 1 and 2 generate similar projections within an assumed level of urban returns, the assumption of urban return level has a larger effect on forest and agricultural projections than any assumptions about the presence of agglomeration economies.

Conclusion and Suggestions for Future Research

This paper develops a plot-level spatial econometric land-use model and estimates it with USGS LCT GIS panel-data from 1980 to 2000 for the western halves of the states of Oregon and Washington. The discrete-choice framework we use models plot-scale choices of the three dominant land uses in this region: forest, agriculture, and urban development. The land-use choice is a function of county average net returns to alternative land uses, and plot level measures of land quality, including

Table 13—Ecoregion agricultural land-use projections for 2040 under alternative urban returns scenarios (private land)

		Model 1		Model 2 (with agglomeration economies)	
	Ag share (2000)	Ag share (2040)	Change	Ag share (2040)	Change
Coast Range:	9.89		*Percent*		*Percent*
1992 urban return level		9.74	-1.50	9.67	-2.24
2012 urban return level		9.75	-1.40	9.67	-2.24
Peak (2006) urban return level		9.78	-1.13	9.66	-2.35
Puget Lowland:	12.53				
1992 urban return level		11.58	-7.54	9.88	-21.12
2012 urban return level		11.45	-8.58	9.34	-25.43
Peak (2006) urban return level		10.63	-15.11	6.54	-47.77
Willamette Valley:	47.26				
1992 urban return level		43.63	-7.68	43.41	-8.16
2012 urban return level		43.55	-7.86	43.12	-8.77
Peak (2006) urban return level		42.91	-9.21	41.05	-13.15
Cascades:	5.35				
1992 urban return level		5.38	0.55	5.36	0.21
2012 urban return level		5.38	0.66	5.36	0.32
Peak (2006) urban return level		5.39	0.85	5.38	0.65
North Cascades:	6.07				
1992 urban return level		6.26	3.19	6.02	-0.74
2012 urban return level		6.27	3.38	6.03	-0.58
Peak (2006) urban return level		6.30	3.79	6.07	-0.02

soil classes and Euclidian distances to cities and roads. The estimation includes measures of spatial land-use agglomeration economies in a framework that allows for both spatial and temporal correlation in the unobservables. We specify a random parameters logit RPL model, which is estimated by maximum simulated likelihood. The estimation uses discrete-choice panel data techniques and instruments for the spatial agglomeration variables with landscape-level land quality variables in an application of the recently developed control function. The combination of panel data RPL with the control function is a novel approach to estimating a fully spatial multinomial discrete-choice land-use model with GIS landscape data.

The estimated econometric model is used to project landscape change in the presence of alternative assumptions regarding future urban returns. The level of future returns to urban uses is uncertain, especially given the recent housing boom and bust, which has seen wide swings in urban returns. The landscape projections indicate substantial urban growth and loss of forest and farmlands in the Puget Lowland and Willamette Valley ecoregions, with minimal changes in the Coast Range, Cascades, and North Cascades regions. Although these results are not surprising, our analysis of the sensitivity of projections to alternative real urban returns levels is perhaps less well understood, and yet extremely relevant for environmental planning. The Puget Lowland ecoregion is the most sensitive portion of the Northwest to swings in the returns to urban uses of land. In particular, the share of the Puget Lowland ecoregion projected to be developed is between 12 and 14 percentage points higher if the level of real urban returns remains at the levels observed in 2006 as compared to whether real urban returns were at the more modest levels observed in 2012. Variation in the projected development share of the Puget Lowland ecoregion between 12 and 14 percentage points presents major challenges for environmental management, as 12 percent of a 1.8 million ha region is approximately 0.2 million ha. The four other Northwest ecoregions have much less variation in response to alternative levels of future urban returns.

The analysis presented here provides a technical foundation for developing larger scale models from the LCT database, which includes a random sample of 10- by 10-km landscape blocks within each ecoregion in the contiguous United States for 1973, 1980, 1986, 1992, and 2000. Additional data for 2006 are currently being processed by USGS, and the LCT data are expected to be updated in the future. The combination of a repeated spatial-panel database and continued updating makes the LCT a candidate database for future large-scale econometric land-use models. Many past national and regional-scale models have been estimated from the NRI (e.g., Langpap and Wu 2011, Lewis and Plantinga 2007, Lubowski et al. 2006). For example, the Forest Service has extensively used the NRI-based Lubowski et al. (2006) model for RPA and climate change assessments. However, because the NRI has no plans to release post-1997 plot-level data with a similar sampling methodology to 1982–1997 NRI data, a new data source and new model will need to be developed to account for post-1997 land-use decisions. Our analysis shows that the LCT provides a plausible alternative database to the NRI for future national-scale land-use models. We provide evidence in the "Basic Econometric Framework and Database Consruction" section that the LCT does not systematically underrepresent urban development as compared to the NRI. This contrasts with what others have found with different land-cover data sources derived solely from automated interpretation of satellite imagery (Irwin and Bockstael 2007, Kline et al. 2009). Although the LCT uses satellite imagery as a primary data source, the USGS manually edits and combines the satellite imagery with aerial photos, topographic maps, and Google Earth to improve classification accuracy.

There are many potential avenues for future econometric modeling from LCT data. First, scaling the regional analysis here up to the national level would serve multiple purposes of interest for the Forest Service, including national-scale resource assessments, analysis of national climate change policies such as carbon sequestration payments, and analysis of large-scale open-space conservation programs and ecosystem services. An open question is whether a single national model would be preferable to a set of multiple regional-scale models such as we have here. A single national-scale model would have an advantage of simplicity (one set of parameters), and would be particularly applicable for simulations of national-scale land-use programs (e.g., carbon sequestration payments) that require some approach to endogenizing net returns measures as a function of land-use changes (e.g., Lubowski et al. 2006). On the other hand, estimating multiple regional-scale models would surely provide more accurate local- and regional-scale projections by estimating regional parameters. One useful research project that would yield insights here would involve testing whether estimated parameters are structurally different across regions.

Second, unlike land-use survey-based datasets such as the NRI, the LCT allows researchers to observe the spatial pattern of land-use change, opening up the possibilities for more complete spatial econometric modeling. Our analysis with LCT block-level unobservables and block-level land-use patterns provides one possibility for specifying spatial econometric land-use models, but the availability of LCT data enables researchers to explore alternative spatial specifications of land-use models. We view this as a fruitful future path for improving the spatial properties of empirical models and corresponding landscape simulations. Another useful improvement over the present analysis would be to include finer scale data on independent variables that drive land-use decisions, such as parcel-level land-use regulations. Alternatively, a fruitful approach to improving spatial specificity could adapt Bockstael's (1996) two-step approach of including estimated parcel-level land-use returns directly into the econometric land-use model, where the returns are estimated in a first-step hedonic price model that explains property sales or rental prices. However, adding such spatial detail is extremely labor intensive, and so there will certainly be tradeoffs between increasing spatial specificity and the scale with which a model can be estimated and applied in a simulation or projection.

Finally, on a more regional note, the most common land cover change in the Pacific Northwest LCT data arises from timber harvesting, which can be tracked with the LCT methodology by following transitions between covers classified as forest, mechanically disturbed, and grass/shrub. Timber harvesting is not a change in land use as modeled in this paper, but the LCT data provide a source for econometrically modeling the spatial properties of the timber harvest decision and its relevant impacts on environmental resources.

References

Adams, D.; Haynes, R., eds. 2007. Resource and market projections for forest policy development: twenty-five years of experience with the U.S. Resources Planning Act (RPA) Timber Assessment. Dordrecht, The Netherlands: Springer. 589 p.

Alig, R. 1986. Econometric analysis of forest acreage trends in the Southeast. Forest Science. 32(1): 119–134.

Alig, R.; Healy, R. 1987. Urban and built-up land area changes in the United States: an empirical investigation of determinants. Land Economics. 63(3): 215–226.

Alig, R.; Plantinga, A. 2004. Future forestland area: impacts from population growth and other factors that affect land values. Journal of Forestry. 102(8): 19–24.

Alig, R.; Plantinga, A.; Haim, D.; Todd, M. 2010. Area changes in U.S. forests and other major land uses, 1982–2002, with projections to 2062. Gen. Tech. Rep PNW-GTR-815. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 92 p.

Butsic, V.; Lewis, D.J.; Ludwig, L. 2011. An econometric analysis of land development with endogenous zoning. Land Economics. 87(3): 412-432.

Chambers, S.W. 2010. An evaluation of Oregon's land-use planning system: evidence from remotely sensed land cover data. Madison, WI: University of Wisconsin-Madison. MS thesis.

Ellison, G.; Glaeser, E.L. 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. Journal of Political Economy. 105(5): 889–927.

Irwin, E.; Bockstael, N.E. 2002. Interacting agents, spatial externalities and the evolution of residential land use patterns. Journal of Economic Geography. 2: 331–354.

Irwin, E.G.; Bockstael, N.E. 2007. The evolution of urban sprawl: evidence of spatial heterogeneity and increasing land fragmentation. Proceedings of the National Academy of Sciences of the United States of America. 104(52): 20672–20677.

Kline, J.D.; Azuma, D.L.; Moses A. 2003. Modeling the spatially dynamic distribution of humans in the Oregon (USA) Coast Range. Landscape Ecology. 18(4): 347–361.

Kline, J.; Alig, R. 2005. Forestland development and private forestry with examples from Oregon. Forest Policy and Economics. 7: 709–720.

Kline, J.D.; Moses, A.; Azuma, D.; Gray, A. 2009. Evaluating satellite imagery-based land use data for describing forestland development in western Washington. Western Journal of Applied Forestry. 24(4): 214–222.

Krinsky, I.; Robb, A. 1986. On approximating the statistical properties of elasticities. The Review of Economics and Statistics. 86: 715–719.

Langpap, C.; Wu, J. 2011. Potential environmental impacts of increased reliance on corn-based bioenergy. Environmental and Resource Economics. 49: 147–171.

LeSage, J.; Pace, R.K. 2009. Introduction to spatial econometrics. Boca Raton, FL: CRC Press. 374 p.

Lewis, D.J.; Plantinga, A.J. 2007. Policies for habitat fragmentation: combining econometrics with GIS-based landscape simulations. Land Economics. 83(2): 109–127.

Lewis, D.J.; Provencher, B.; Butsic, V. 2009. The dynamic effects of open-space conservation policies on residential development density. Journal of Environmental Economics and Management. 57: 239–252.

Lewis, D.J.; Barham, B.L.; Robinson, B. 2011. Are there spatial spillovers in the adoption of clean technology? The case of organic dairy farming. Land Economics. 87(2): 250–267.

Lubowski, R.N. 2002. Determinants of land-use transitions in the United States: econometric analysis of changes among the major land-use categories. Cambridge, MA: Harvard University. Ph.D. dissertation.

Lubowski, R.N.; Plantinga, A.J.; Stavins, R.N. 2006. Land-use change and carbon sinks: econometric estimation of the carbon sequestration supply function. Journal of Environmental Economics and Management. 51(2): 135–152.

Manski, C.F. 1993. Identification of endogenous social effects: the reflection problem. The Review of Economic Studies. 60(3): 531–542.

Newburn, D.A.; Berck, P.; Merenlender, A.M. 2006. Habitat and open space at risk of land-use conversion: targeting strategies for land conservation. American Journal of Agricultural Economics. 88(1): 28–42.

Petrin, A.; Train, K.E. 2002. Omitted product attributes in discrete-choice models. Working Paper. Berkeley, CA: University of California-Berkeley: Department of Economics. 22 p.

Plantinga, A.J. 1996. The effect of agricultural policies on land use and environmental quality. American Journal of Agricultural Economics. 78: 1082–1091.

Sleeter, B.M.; Wilson, T.S.; Acevedo, W., eds. 2012. Status and trends of land change in the Western United States—1973 to 2000: U.S. Geological Survey Professional Paper 1794–A. Reston VA: U. S. Department of the Interior, Geological Survey. 324 p. http://pubs.usgs.gov/pp/1794/a/. (December 18, 2013).

Stein, S.M.; McRoberts, R.E.; Alig, R.J.; Nelson, M.D.; Theobald, D.M.; Eley, M.; Dechter, M.; Carr, M. 2005. Forest on the edge: housing development on America's private forests. Gen. Tech. Rep. PNW-GTR-636. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 16 p.

Towe, C.A.; Nickerson, C.J.; Bockstael, N. 2008. An empirical examination of the timing of land conversions in the presence of farmland preservation programs. American Journal of Agricultural Economics. 90(3): 613–626.

Train, K.E. 2009. Discrete-choice methods with simulation. 2nd ed. Cambridge, United Kingdom: Cambridge University Press. 408 p.

U.S. Department of Agriculture, Forest Service. 1989. RPA assessment of the forest and rangeland situation in the United States, 1989. For. Resour. Rep. 26. Washington, DC. 72 p.

Walker, J.L.; Ben-Akiva, M.; Bolduc, D. 2007. Identification of parameters in normal error component logit-mixture (NECLM) models. Journal of Applied Econometrics. 22: 1095–1125.

Wall, B. 1981. Trends in commercial timberland area in the United States by state and ownership, 1952–77, with projections to 2030. Gen. Tech. Rep. WO-GTR-31. Washington, DC: U.S. Department of Agriculture, Forest Service. 26 p.

Wear, D.N.; Murray, B.C. 2004. Federal timber restrictions, international spillovers, and the impact on U.S. softwood markets. Journal of Environmental Economics and Management. 47(2): 307–330.

Pacific Northwest Research Station

Web site	http://www.fs.fed.us/pnw/
Telephone	(503) 808–2592
Publication requests	(503) 808–2138
FAX	(503) 808–2130
E-mail	pnw_pnwpubs@fs.fed.us
Mailing address	Publications Distribution Pacific Northwest Research Station P.O. Box 3890 Portland, OR 97208–3890

U.S. Department of Agriculture
Pacific Northwest Research Station
1220 SW 3rd Ave
P.O. Box 3890
Portland, OR 97208–3890

Official Business
Penalty for Private Use, $300