Staff Paper

Capturing Household-Level Spatial Influence in Agricultural Management Using Random Effects Regression

Scott M. Swinton

Staff Paper 2002-15 **July 2002**

Department of Agricultural Economics
MICHIGAN STATE UNIVERSITY
East Lansing, Michigan 48824

Capturing Household-level Spatial Influence in Agricultural Management Using Random Effects Regression

Scott M. Swinton
swintons@pilot.msu.edu

ABSTRACT:

Data on agricultural and natural resource management typically have spatial patterns related to the landscapes from which they came. Consequently, econometric models designed to explain the determinants of humans' natural resource management practices or their outcomes often have spatial structure that can bring bias or inefficiency to parameter estimates.

Although econometric tools are available to correct for spatial structure, such tools are largely lacking for use with discrete dependent variable models. While one obvious solution would be to develop the necessary tools, an alternative is to identify conditions under which spatial dependency can be managed effectively without formal spatial autoregressive models.

This study examines conditions under which spatial structure corresponds closely to defined agro-ecological zones, making it possible to model spatial effects by random effects regression. Using household survey data sampled along agro-ecological zone strata, this article develops two models of links between farmer assets and agricultural natural resource degradation in southern Peru. The first stage model looks at determinants of crop yield loss over time (an index of soil productivity), while the second stage model looks at determinants of the extent of fallow cycles in crop rotation, a key agricultural practice reducing crop yield loss.

Diagnostic statistics for spatial dependency reveal spatial structure, particularly in the fallow model. This spatial dependency is eliminated in the ordinary least squares (OLS) models by inclusion of the agro-ecological zone random effects. In the spatially dependent fallow model, comparison of coefficient estimates between OLS and the spatial autoregressive maximum likelihood models showed OLS with random effects to give virtually identical results to the spatial autoregressive models, making the latter unnecessary.

These results show that spatial structure in natural resource management models can sometimes be captured by zonal variables. When this occurs, random effects regression can largely eliminate spatial dependency. A necessary precondition for this approach with household survey data is prior sample stratification according to landscape characteristics. Where random effects models can effectively capture spatial structure, they may also offer analysts greater flexibility in analyzing models with limited dependent variables.

JEL codes: Q12, R15

Keywords: random effects models, spatial autoregressive models, spatial lag, spatial error, land use, agricultural technology adoption, natural resource management

20 pages

CAPTURING HOUSEHOLD-LEVEL SPATIAL INFLUENCE IN AGRICULTURAL MANAGEMENT USING RANDOM EFFECTS REGRESSION

by Scott M. Swinton[1]

Michigan State University

Department of Agricultural Economics

Staff Paper No. 2002-15

Dr. Scott M. Swinton
Department of Agricultural Economics
Michigan State University
East Lansing, MI 48824-1039
Tel: 1-517-353-7218; Fax: 1-517-432-1800;
E-mail: swintons@msu.edu

[1] Scott M. Swinton is associate professor at Michigan State University, East Lansing, MI, U.S.A.

The author wishes to acknowledge Roberto Quiroz, head of the Department of Production Systems and Natural Resource Management at the International Potato Center (CIP) in Lima, Peru, for support and collaboration during a fruitful 1999 sabbatical at CIP. For help in designing and implementing both data collection and management, I also thank Jorge Reinoso, Roberto Valdivia, Sebastian Paredes, Silvia Agüero, Mariana Cruz, Yonny Barriga, Pedro Cabrera, José Rivera, and Andrés Valladolid. For helpful advice and comments on methods, I thank Luc Anselin, Bill Greene, Jerry Nelson, Jeff Wooldridge, and two anonymous reviewers. Remaining errors are, of course, my own.

Capturing Household-level Spatial Influence in Agricultural Management Using Random Effects Regression

Location is relevant to many classes of economic problems. Von Thünen's early marginalist insights were illustrated by the effect of transportation costs on the location and technology of agricultural production (Von Thunen, 1966). As economists sought ways to incorporate location into applied analyses, linear programming tools made possible the transportation cost minimization model. Until recently, however, the complexity of spatial effects has tended to defy explicit incorporation of spatial attributes beyond distances between pairs of points and selected landscape characteristics.

New tools for gathering and analyzing spatial data have begun a wave of econometric exploration into how location affects economic phenomena. Recent papers have explored the spatial patterns in demand for food and real estate (Case, 1991, Pinkse and Slade, 1998), agricultural technology diffusion (Case, 1992, Foster and Rosenzweig, 1995, Torero, 1992), and land use transformation (Bockstael, 1996, Nelson, et al., 2001, Nelson and Hellerstein, 1997)

The proliferation of geographic information database management systems (GIS) and the accessibility of locational technologies such as the global positioning system (GPS) have dramatically reduced the cost of acquiring and managing spatial data. Despite notable advances in spatial econometrics, led by Anselin (Anselin, 1988), the statistical tools available to analyze spatial data are lagging behind the ability to collect it. Anselin has developed both concepts and software for the diagnosis and correction of spatial autocorrelation (Anselin, 1999, Anselin, 1988) . At present, however, methods to correct for spatial structure in limited dependent variable models are restricted to binary dependent variables (Kelejian and Prucha, 1999, Pinkse and Slade, 1998). Yet many microeconomic phenomena, particularly those involving choices among discrete alternatives, call for analysis with other types of discrete variables based on multinomial, rank or count data.

This article develops and illustrates conditions under which a random effects model may be suitable to correct for spatial structure. Since random effects modeling tools are widely available for limited dependent variable analysis (e.g., LIMDEP, STATA), this substitution can potentially improve the efficiency and freedom from bias of parameter estimates from spatially structured models.

The approach developed here is particularly suited to spatially-referenced survey data. Household surveys that focus on agricultural or natural resource management often employ a stratified sampling design based on landscape units such as eco-regions. Rather than sample randomly from an area frame[1], limited budgets for household surveys often dictate cluster sampling (e.g., by village), followed by randomized list sampling within the cluster (Deaton, 1997). Such sampling permits more effective use of enumerator time, at the cost of household data that are spatially random.

Much of the literature on spatial data about natural resource management focuses on spatially dense remote sensing data (e.g., (Bockstael, 1996, Nelson, et al., 2001, Nelson and Hellerstein, 1997). Such data are spatially comprehensive, although their precision depends upon the sensing equipment (Aronoff, 1998). Because these data are gathered automatically, they come at high fixed but low marginal cost. For natural resource management purposes, however, they are have two limitations. First, the electromagnetic wavelengths sensed do not directly correspond with agricultural and landscape attributes of interest to humans, although indexes developed from these wavelengths may correlate with variables of scientific, management or policy interest (Swinton and

Jones, 1998). Second, remotely sensed data *per se* are not linked to the human decision makers who may play a big role in the fate of the land. This latter limitation has sparked interest in household surveys to add the missing human behavioral link (Bockstael, 1996).

Because managing spatial dependence in statistical analysis of natural resources management can pose difficulties (especially with limited dependent variable models), this paper examines conditions under which natural resource-based strata can adequately capture spatial structure in a random-effects regression. After developing a conceptual econometric model that incorporates both spatial effects and random effects, the paper tests that model with data from a spatially referenced household survey of links between agricultural practices and land degradation in Peru's Andean high plain (Altiplano).

Econometric model

The general spatial regression model takes two parts (Anselin, 1988):

$$y = \rho Wy + X\beta + \varepsilon \qquad (1)$$
$$\varepsilon = \lambda W\varepsilon + \mu$$

The dependent variable is a lagged function of spatial neighbors with an error term that is spatially autocorrelated with neighbors. Interpreted for a household survey with N households, W is a (N x N) spatial weights matrix, where ρ is a (N x 1) spatial lag parameter, X is an (N x K) matrix of explanatory variables and β is a (K x 1) parameter vector. The spatial error term, ε, is a lagged function of nearby errors, using the same spatial weights matrix, W, the (N x 1) lag parameter, λ, and a (N x 1) i.i.d. normal disturbance vector, μ.

The spatial weights matrix characterizes the spatial relationships among observations. Typically, spatial effects decay with distance. Of interest for our purposes is whether they decay gradually or abruptly. If the decay is gradual, then a weights matrix is appropriate to describe it. If the decay is abrupt but error or lag patterns are relatively homogeneous before a fairly discrete boundary, then it may be adequate to know if an observation *i,j* is in or out of a local area of spatial influence. Under such circumstances, a random effects (RE) model (where sample observations are available) or a fixed effects model (where data from an entire population are available) may well be adequate without the need for explicit modeling of spatial structure. In the extreme, when W is block diagonal with blocks having homogeneous spatial influence within each block (i.e., within a block n_1 x n_1, $w_{ij} = 1/n_1$ for all i, j), then spatial effects are perfectly captured by a dummy variable that identifies observations in n_1.

Case (1991) first proposed fixed and random effects models to proxy for a spatial regression and illustrated her case using district-level models of spatial rice demand in Indonesia. This paper extends her method to zones defined to describe natural resources conditions.

A simple, intercept-only RE model can be developed by inserting into Equation (1) what Bryk and Raudenbusch (1992) refer to as a Level-2 district-specific RE disturbance term, φ, yielding the hybrid model (Case, 1991):

$$y = \rho Wy + X\beta + \varepsilon \qquad (2)$$
$$\varepsilon = \lambda W\varepsilon + \varphi + \mu$$

Equation (2) can be simplified to separate the components into three econometrically distinct parts,

$$y = X\beta + \mu + [\rho Wy + \lambda W\varepsilon + \varphi] \qquad (3)$$

In Equation (3), the first two terms represent unbiased parameter estimates and random error, while the bracketed group captures spatial effects, both those related to the spatial weights matrix, W (as lagged dependent variable or disturbances), and those that are binary random effects, φ, that correspond to defined geographic zones.

The first step empirically is to test the hypothesis that the bracketed term in Equation (3) has null parameters,

$$H1: \rho = \lambda = \varphi = 0.$$

If that hypothesis cannot be rejected, then the model does not display pronounced spatial effects, and non-spatial regression models can be estimated appropriately.

If spatial structure is detected, the next step is to test whether those effects are chiefly embodied in the zonal dummy variables,

$$H2: \rho = \lambda = 0 \text{ (but allow } \varphi \neq 0).$$

This test may be applied to a dependent variable in the data set that is continuous (hence suited to OLS or GLS estimation) in order to characterize spatial structure among right-hand side variables being considered for a discrete dependent variable analysis. If H2 cannot be rejected, then spatial effects are chiefly captured by the zonal dummy variables, so random effects regression can provide consistent parameter estimates without further spatial correction. If H2 is rejected, then there exists either a spatial lag or a spatial error effect, and spatial correction with use of a spatial weights matrix will be necessary to obtain consistent parameter estimates.

Data and sampling method

These hypotheses will be tested using survey data gathered from single-visit interviews in 170 households in southern Puno department, Peru, during April-June, 1999. The purpose of the survey was to characterize natural resource degradation problems and their relationship to household poverty. The approach to understanding poverty-environment links in the region follows two stages.

First, specific natural resource outcomes are regressed on a range of explanatory variables in order to discern whether specific agricultural practices affect the status of the natural resource base.

The case examined here involves crop yield decline over a 20-year period (as an index of soil degradation). The general structure of the model is as follows:

$$\Delta Y = f(P, X, Z) \tag{4}$$

where ΔY is the change in proportionate "typical" crop yield over a twenty-year period ($[Y_0 - Y_{20}]/Y_{20}$), P is a vector of relative price variables, X is a vector of management variables hypothesized to affect crop yield, and Z is a vector of agrophysical, institutional and household conditioning factors. Ideally, such a model should be a panel with all variables expressed as differences over the 20-year period. In the absence of such information, current practices and conditions are assumed to serve as adequate proxies for unobserved patterns over time.

Second, individual variables describing influential agricultural management practices are regressed on a set of explanatory variables in order to understand whether and how poverty may influence the choice of agricultural practices. Some of these variables are the same as in Equation (4), but this group of variables also includes ones that describe various asset categories (physical, infrastructural, financial, human and social capital). In the case examined here, the proportion of fields in fallow (X_f) turns out to be a key determinant of yield decline, so its determinants are modeled as follows:

$$X_f = f(P, X_{(f)}, A, Z) \tag{5}$$

In Equation (5), $X_{(f)}$ is a vector of agricultural practices other than fallowing, A is a vector of asset variables, and P and Z are again prices and conditioning variables.

Definitions and descriptive statistics for all variables included in the two empirical models are presented in Table 1. Of special interest are the two dependent variables for the way in which their properties might affect error terms. Both dependent variables are continuous, but occur in bounded ranges. Both the yield loss over 20 years variable (ΔY in Equation (4)) and the proportion of fields in fallow (X_f in Equation (5)) range from 0 to 75 percent. Although truncation at zero would normally imply the need for a tobit model, this was not feasible for diagnosis of spatial structure using the SpaceStat 1.90 software (Anselin, 1999) and the use of ordinary least squares (OLS) regression was deemed suitable for illustrative purposes.

The survey employed a clustered, stratified sampling design in the Ilave-Huenque river watershed of the Lake Titicaca basin (Figure 1). Strata were defined based on three agro-ecological zones that vary with distance from Lake Titicaca (Tapia, 1996). The flat, Lakeside zone has a frost-free cropping season of 5-6 months and 700-750 mm. of rainfall annually. Farming in the Lakeside zone is characterized by intensive potato-based crop rotations that include rising shares of forage crops to supplement the lake reeds traditionally used for livestock feed. Official measures of poverty are lowest in the Lakeside zone. Moving up and away from the lake, the Suni zone is next, with a frost-free season ranging from three to five months and slightly less rainfall than the Lakeside zone. Frost risk depends upon landscape position, leading to a distinction between the Suni A and Suni B zones, where the former has more "lake effect" and is less prone to night-time frosts. Potatoes can be grown in the Suni A zone, whereas they are a very risky crop in the poorer Suni B zone. Consequently, households in the Suni B zone rely more heavily on rangefed livestock production and less on crops than their counterparts in the Suni A zone. Ex post statistical comparisons supported the existence of distinct agro-ecological zones as defined for the sampling strata (Swinton, et al., 1999).

Within each of the three agro-ecological zone strata, two to three villages were selected as primary sampling units (household clusters). Based on the advice of regional government officials, these village pairs or triples were chosen to include one relatively less and one relatively more poor than the norm for the zone. Within villages, an attempt was made to stratify households by apparent wealth level, in consultation with village leaders. The sampling stratification scheme was designed to ensure a broad range of asset levels across the agro-ecological zones in order to test the research hypotheses about poverty-environment links.

The eight villages surveyed in the three agro-ecological zones included a total of 197 households. The location of those households is illustrated in Figure 2. Of these, 170 households provided complete records that were usable for the analyses in the following section. Key among those records are readings of farmstead location (latitude, longitude and altitude) taken with hand-held Global Positioning System (GPS) units without differential correction. Details on the study and empirical results can be found in Swinton and Quiroz (2000).

An empirical test of random effects in a spatial regression model

The spatial nature of agro-ecological zones offers reason to expect that the zones could capture part of any existing spatial structure in the household data. We test the hypotheses set forth above using the two-stage analysis in Equations (4) and (5), which focus on determinants of crop yield loss and the related agricultural practice of using fallow in crop rotation.

A distance matrix was developed from latitude-longitude data of the farmstead locations. Based on Bell and Bockstael's evidence that statistical results can be sensitive to the spatial weights methodology when using household-level point data (Bell and Bockstael, 2000), five different methods were pursued to obtain row-standardized spatial weights matrices:

1. Inverse distance (with upper limit 200 times minimum distance),
2. Inverse distance squared (with upper limit 200 times minimum distance),
3. Four nearest neighbors,
4. Eight nearest neighbors,
5. Sparse distance weights (minimum distance set to ensure at least one neighbor to least-connected point).

Hypothesis H1 is first tested by evaluating Equations (4) and (5) for spatial structure using diagnostic statistics for spatial error and spatial lag. Spatially correlated errors are diagnosed by Moran's I statistic and the spatial error Lagrange multiplier test. Spatial lag structure is tested for using the spatial lag Lagrange multiplier test (Anselin, 1988, Bell and Bockstael, 2000, Case, 1991).

If H1 is rejected and spatial structure is evident, hypothesis H2 is evaluated in two ways. First, it is evaluated at the level of the same diagnostic statistics as in H1. Second, a closer examination is made of how choice of the wrong model might affect coefficient estimates (Havlicek and Seagraves, 1962). Model estimation results from OLS are compared with those from spatial autoregressive maximum likelihood (SAR-ML) models with and without the agro-ecological zone random effects. The three criteria examined are a) numbers of coefficient estimates that are significant at the 95% level, b) magnitude of differences in those coefficient estimates, and c) overall goodness of fit. For ease of comparing coefficient estimates, all non-binary variables were standardized according to the formula $(x_i - x_\mu)/\sigma$, where x_μ is the sample mean and σ is the sample standard deviation of x.

Given similarities between the pairs of inverse distance and nearest neighbors weights matrices, we focus only on spatial weights matrices developed from the simple inverse distance and the nearest four neighbors. Because the sparse distance weights admit only analysis by Kelejian and Prucha's generalized moments estimator (Kelejian and Prucha, 1999), which does not provide a standard error of the λ spatial error coefficient in Equation (1), the sparse distance weights are omitted. Both spatial error and spatial lag models are estimated, due to evidence that both forms of spatial structure may exist (Table 2). Although the SAR-ML models do not provide identical results, they are sufficiently similar to offer a joint benchmark for comparison of the OLS results.

Results of the test

As shown in Table 2, the evidence of spatial structure for the yield loss base model (Equation (4)) was negligible, with only the Moran's I statistic under the sparse distance weights suggesting the slightest evidence of spatial autocorrelation of errors. By contrast, evidence of spatial dependency was strong by all measures for the fallow practice base model. So hypothesis H1 could not be rejected for the yield loss model, but it was rejected for the fallow model.

Introduction of the agro-ecological zone dummy variables completely removed evidence of spatial dependency in the random effects models (Table 2). Hypothesis H2 could not be rejected for either the yield loss or the fallow model. It appears that the spatial structure was strongly correlated with the agro-ecological zones. Evidently, the correction was strong enough to be robust to differences in spatial weight matrices in all cases.

Although the hypothesis of random effects (H2) could not be rejected at the level of diagnosing spatial structure, it remains of interest to evaluate the OLS-RE model against the spatially corrected regression models. In doing this, it makes sense to focus on the fallow practice model (the one that exhibited clear spatial structure). Due to evidence in Table 2 of both spatial error and spatial lag effects in the fallow model, correction for both is examined.

Focusing first on spatial error correction, the results in Table 3 indicate that the OLS-RE model offers equivalent explanatory power to the SAR-ML models, but the base OLS model does not. The OLS-RE model eliminates the effect of spatial autocorrelation, so that the λ coefficient is insignificant in the SAR-ML-RE models. By contrast, the base SAR-ML models (without RE) show significant λ coefficient estimates (and associated likelihood ratio tests for spatial error dependence), reinforcing the evidence that the base OLS model fails to address underlying spatial structure.

Of eleven significant coefficient estimates, the base OLS model includes two that are not significant in either corresponding SAR-ML model (Home equipment and Fertilizer rate). Perhaps more troubling, the "significant"coefficient estimates from the base OLS model deviate from the comparable SAR-ML estimates by over 40 percent in some cases (e.g., Natural resources project).

By contrast, the OLS-RE model appears to generate conservative estimates, with two fewer significant coefficient estimates than the SAR-ML-RE models. It is worth noting that these two were both significant at a 6 percent probability of mistakenly rejecting the null hypothesis that β_j=0 (Unmet basic needs and Association memberships). In magnitude, the estimates of the "significant" coefficients are within 2 percent of one another across the three RE models. The agro-ecological zone dummies are significant in all the RE models. Finally, the log likelihood function values for the OLS-RE model is virtually equal to the SAR-ML-RE models and significantly higher than the OLS base model.

Turning to the spatial lag effect, again OLS with RE adequately captured the spatial structure in these data, but the base OLS model did not. The SAR-ML base models both displayed a significant spatial lag coefficient estimate (ρ) and both failed the likelihood-ratio test for spatial lag dependence. By contrast, none of the SAR-ML-RE models did so. Although the number of significant coefficient estimates in the OLS base model did not vary from its SAR-ML counterparts, the magnitude of parameter estimates did vary. The SAR-ML coefficients were up to 50 percent smaller than their OLS counterparts (e.g., Natural resources project). The RE models again showed a near-perfect correspondence in significant coefficients except for the "Unmet basic needs" variable which, as noted above, was within one percent of the 5 percent confidence threshold. The OLS-RE significant coefficients deviated from their SAR-ML-RE counterparts by no more than 6 percent. Finally, on goodness-of-fit, the OLS-RE model shares nearly identical log likelihood function values with the SAR-ML-RE models, significantly higher than that of the OLS base model.

Conclusion

The general conclusion is that careful household sampling stratification on the basis of key landscape features can capture important elements of spatial structure. Where spatial structure is strongly associated with non-overlapping landscape zones, incorporating these zones in a random effects regression model may obviate the need for any more formal spatial regression modeling. The results presented here for household data on agricultural management and change in typical crop yields over time in the Peruvian Altiplano indicate that the inclusion of random effects corresponding to predefined agro-ecological zone sampling strata eliminated all meaningful evidence of spatial structure from the data. Diagnostic statistics show that the random effects OLS models of yield and fallow proportion were free of spatial dependence using five different methods of spatial weighting.

In the spatially dependent fallow model, OLS-RE estimation proved equally robust as its spatially corrected counterpart models in identifying significant coefficient estimates of similar magnitudes, eliminating spatial effects, and fitting the data well.

The evidence indicates that with proper natural resource sampling stratification, random effects can capture spatial dependency effects in natural resource management models, as Case found could be done with political boundaries and food consumption data (Case, 1991). In both cases, an important condition is that the spatial strata be defined as bounded zones that do not overlap. If significant spatial structure were confined to the village scale, individual villages might equally well serve as the basis for RE models that capture spatial effects. As in the case illustrated here, that could be tested.

Where random effects models can effectively capture spatial structure, they offer two important benefits. First, they permit analysis with limited dependent variables that is infeasible with current methods for modeling spatial structure. Research related to that presented here that includes probit, tobit and Poisson models that would be difficult or impossible with correction for spatial structure. Second, RE models permit analysis with statistical software that is more fully featured and accessible than the current standard in spatial econometric analysis.

A potential limitation of the investigation presented here is that the sample households in each agro-ecological zonal stratum were located in two or three villages that were not distant from one another. This clustering within strata may have reduced the diversity of conditions found in each agro-ecological zone. However, a countervailing effect is that villages were paired to reflect different wealth levels, so the physical homogeneity should have been offset by socio-economic heterogeneity (that would not have been captured by the agro-ecological zones). Future research would do well to examine the effectiveness of random effects models using data from more randomized, area-frame sampling.

Endnotes

1. Spatially random samples can be drawn from an area sampling frame, as in the the USDA National Resources Inventory ((USDA), 2000)). When the land manager can be identified, this sampling approach offers a valuable means to connect the condition of the land to the humans who manage it. Because spatially random area sampling is quite an expensive way to conduct household survey research, it is not common.

Table 1: Descriptive statistics on variables included in the two models, 170 households, Puno, Peru, 1999.

Variable type	Definition	Median	Mean	St. Dev.	Min.	Max.
Dependent variables						
Yield loss	Proportion of yield lost comp. w/ 20 years ago	0.38	0.35	0.18	0.00	0.75
Fallow fields	Prop'n of fields	0.25	0.25	0.21	0.00	0.75
Location & Natural factors						
Zone: Suni A	Binary	0.00	0.37	0.48	0.00	1.00
Zone: Suni B	Binary	0.00	0.34	0.47	0.00	1.00
Sandy soil proportion	Prop'n of fields	0.40	0.42	0.27	0.00	1.00
Foot-slope location	Prop'n of fields	0.03	0.13	0.19	0.00	0.82
Hillside location	Prop'n of fields	0.00	0.12	0.18	0.00	0.75
Management factors						
Small grain fields	Prop'n planted area	0.45	0.46	0.22	0.00	0.88
Contour furrows	Prop'n of fields	0.00	0.03	0.12	0.00	0.82
Vertical furrows	Prop'n of fields	0.77	0.70	0.32	0.00	1.00
Fertilizer rate	Kg/ha	16.09	48.98	107.04	0.00	1149.13
Pesticides rate	Kg/ha	0.00	2.33	7.18	0.00	71.42
Labor value	Peru new soles	680.00	989.07	905.61	33.94	6592.50
Price and location factors						
Price of potato	Peru new soles/kg	0.43	0.41	0.12	0.17	0.87
Distance to paved road	Hours travel by car	0.42	0.40	0.26	0.01	1.00
Asset variables						
Unmet basic needs	Sum	1.00	0.78	0.62	0.00	3.00
Cropped area	Hectares	0.62	1.00	1.03	0.01	6.13
Pasture area	Hectares	0.63	10.64	35.81	0.00	286.50
Vehicles owned	Units	0.00	0.31	0.79	0.00	6.00
Store/warehouse	Units	0.00	0.28	0.56	0.00	3.00
Well equipment	Units	0.00	0.19	0.41	0.00	2.00
Other ag. equipment	Units	3.00	3.23	2.52	0.00	14.00
Home equipment	Units	2.00	2.49	1.93	0.00	9.00
Total SEVU's	Sheep value units	57.00	75.31	63.78	0.20	433.20
Nonfarm income	Peru soles	2381.00	3966.00	4728.00	-1607.00	30221.00
Family ag. labor supply	Person-years	2.00	2.38	1.57	-0.40	8.00
Credit	Peru soles	0.00	196.47	753.82	0.00	5000.00
Education of HH head	Years	1.00	1.91	1.55	0.00	6.00
Adults w/ high school	Units	1.00	0.92	1.20	0.00	6.00
Position of HH head	Binary	1.00	0.71	0.45	0.00	1.00
Assn. memberships	Units	1.00	1.18	0.87	0.00	4.00
Aynoca area	Hectares	148.00	121.32	104.15	0.00	278.00

Table 2: Spatial dependence test statistics for five distance weight matrices in two models with and without random effects, 170 households, Puno, Peru, 1999.

Model:	**Yield loss (Eq. 4)**		**Fallow practice (Eq. 5)**	
Variable set:	**Base**	**Random effects**	**Base**	**Random effects**
Inverse distance weights				
Moran's I (error)	0.350	-0.013	3.458	1.063
	(0.726)	(0.730)	(0.001)	(0.288)
LM (error)	0.042	0.145	4.271	0.026
	(0.837)	(0.704)	(0.039)	(0.872)
LM (lag)	0.198	0.548	29.335	0.276
	(0.657)	(0.460)	(0.000)	(0.559)
Inverse distance squared				
Moran's I (error)	-0.202	-0.083	2.592	1.273
	(0.840)	(0.934)	(0.010)	(0.202)
LM (error)	0.280	0.271	3.687	0.290
	(0.600)	(0.603)	(0.055)	(0.590)
LM (lag)	0.700	0.881	18.480	0.971
	(0.403)	(0.348)	(0.000)	(0.325)
4 nearest neighbors				
Moran's I (error)	-0.641	-0.881	2.684	0.871
	(0.521)	(0.379)	(0.008)	(0.384)
LM (error)	0.896	1.630	3.410	0.000
	(0.344)	(0.202)	(0.065)	(0.987)
LM (lag)	0.600	0.987	21.224	1.030
	(0.440)	(0.320)	(0.000)	(0.310)
8 nearest neighbors				
Moran's I (error)	-0.328	-0.651	2.817	1.424
	(0.743)	(0.515)	(0.004)	(0.154)
LM (error)	0.577	1.425	2.275	0.015
	(0.447)	(0.233)	(0.131)	(0.901)
LM (lag)	0.334	0.826	29.033	2.176
	(0.563)	(0.363)	(0.000)	(0.140)
Sparse distance weights				
Moran's I (error)	1.720	1.117	4.284	0.158
	(0.085)	(0.264)	(0.000)	(0.875)
LM (error)	0.792	0.031	4.592	1.125
	(0.373)	(0.861)	(0.032)	(0.289)
LM (lag)	0.840	0.083	29.16	1.198
	(0.359)	(0.773)	(0.000)	(0.274)

Notes: "Random effects" refers to models that include binary variables for agro-ecological zones Suni A and Suni B (making the constant capture the effect of the Lakeside zone). LM denotes Lagrange Multiplier tests for spatial error and spatial lag. Numbers in parentheses are probabilities of failing to reject the null hypothesis of no spatial structure.

Table 3: Regression coefficient estimates for OLS, OLS with random effects, and spatial error-corrected models using two distance matrices, 170 households, Puno, Peru, 1999.

Regression method:	**OLS**		**SAR-ML (error)**			
Spatial weights matrix:	**No spatial weights**		**Inverse distance**		**4 nearest neighbors**	
Model:	**Base**	**R.E.**	**Base**	**R.E.**	**Base**	**R.E.**
Constant	**-1.408**	**-1.317**	**-1.249**	**-1.317**	**-1.354**	**-1.317**
Price of potato	0.109	0.003	0.080	-0.001	0.091	0.003
Unmet basic needs	**0.148**	0.107	0.090	**0.109**	**0.121**	**0.107**
Cropped area	-0.036	0.095	0.022	0.097	-0.019	0.095
Pasture area	0.114	0.078	**0.110**	0.076	**0.123**	0.078
Vehicles owned	0.051	0.040	0.045	0.039	0.040	0.040
Store/warehouse	**0.151**	**0.116**	**0.128**	**0.116**	**0.146**	**0.116**
Well equipment	**0.137**	**0.139**	**0.112**	**0.141**	0.103	**0.138**
Other ag. equipment	0.059	-0.054	-0.019	-0.055	0.028	-0.054
Home equipment	**0.151**	0.107	0.099	0.108	0.117	0.107
Total SEVU's	0.041	-0.039	-0.016	-0.038	0.002	-0.039
Nonfarm income	**-0.236**	**-0.184**	**-0.165**	**-0.186**	**-0.183**	**-0.183**
Family ag. labor supply	-0.151	-0.123	-0.112	-0.125	-0.123	-0.123
Credit	-0.002	-0.030	-0.011	-0.031	-0.003	-0.030
Distance to paved road	-0.026	-0.008	-0.020	-0.010	-0.053	-0.008
Education of HH head	-0.107	-0.084	-0.080	-0.084	-0.089	-0.083
Adults with high school	**0.254**	**0.185**	**0.184**	**0.187**	**0.198**	**0.185**
Position of HH head	0.110	0.047	0.065	0.046	0.115	0.047
Association memberships	0.072	0.105	0.062	**0.108**	0.061	**0.105**
Aynoca area	**0.417**	**0.233**	**0.236**	**0.236**	**0.327**	**0.232**
Foot-slope location	**0.185**	0.074	**0.119**	0.072	**0.169**	0.074
Sandy soil proportion	0.058	-0.006	-0.002	-0.005	0.039	-0.006
Fertilizer rate	**-0.120**	-0.060	-0.083	-0.059	-0.104	-0.060
Pesticides rate	-0.081	-0.060	**-0.102**	-0.056	-0.080	-0.060
Natural resources project	**1.549**	**0.574**	**1.370**	**0.576**	**1.482**	**0.574**
Zone: Suni A		**0.840**		**0.839**		**0.841**
Zone: Suni B		**1.429**		**1.430**		**1.429**
Spatial error (lambda)			**0.647**	-0.049	**0.326**	0.003
Log Likelihood (neg)	-160.3	-140.7	-154.6	-140.6	-155.4	-139.7

Note: Boldface type denotes probability of Type II error ≤ 5 percent in rejecting H^0: β=0.

Table 4: Regression coefficient estimates for OLS, OLS with random effects, and spatial lag-corrected models using two distance matrices, 170 households, Puno, Peru, 1999

Regression method:	**OLS**		**SAR-ML (lag)**			
Spatial weights matrix:	**No spatial weights**		**Inverse distance**		**4 nearest neighbors**	
Model:	**Base**	**R.E.**	**Base**	**R.E.**	**Base**	**R.E.**
Spatial lag-Y (rho)			**0.505**	0.084	**0.375**	0.106
Constant	**-1.408**	**-1.317**	**-0.722**	**-1.215**	**-0.894**	**-1.184**
Price of potato	0.109	0.003	0.073	0.008	0.084	0.009
Unmet basic needs	**0.148**	0.107	0.102	**0.104**	**0.119**	**0.104**
Cropped area	-0.036	0.095	0.015	0.090	-0.006	0.087
Pasture area	0.114	0.078	**0.112**	0.081	**0.119**	0.083
Vehicles owned	0.051	0.040	0.038	0.039	0.038	0.037
Store/warehouse	**0.151**	**0.116**	**0.128**	**0.115**	**0.139**	**0.117**
Well equipment	**0.137**	**0.139**	**0.117**	**0.135**	**0.108**	**0.130**
Other ag. equipment	0.059	-0.054	-0.009	-0.055	0.011	-0.054
Home equipment	**0.151**	0.107	**0.117**	0.106	**0.122**	0.105
Total SEVU's	0.041	-0.039	-0.004	-0.038	-0.004	-0.041
Nonfarm income	**-0.236**	**-0.184**	**-0.185**	**-0.181**	**-0.185**	**-0.176**
Family agric. labor supply	-0.151	-0.123	-0.127	-0.122	-0.121	-0.118
Credit	-0.002	-0.030	-0.011	-0.029	0.000	-0.026
Distance to paved road	-0.026	-0.008	0.016	-0.001	-0.015	-0.007
Education of HH head	-0.107	-0.084	-0.075	-0.081	-0.080	-0.079
Adults with high school	**0.254**	**0.185**	**0.195**	**0.183**	**0.192**	**0.176**
Position of HH head	0.110	0.047	0.085	0.049	0.113	0.055
Association memberships	0.072	0.105	0.085	0.104	0.074	**0.101**
Aynoca area	**0.417**	**0.233**	**0.251**	**0.225**	**0.261**	**0.212**
Foot-slope location	**0.185**	0.074	**0.123**	0.075	**0.156**	0.079
Sandy soil proportion	0.058	-0.006	0.010	-0.007	0.028	-0.006
Fertilizer rate	**-0.120**	-0.060	-0.086	-0.059	**-0.098**	-0.061
Pesticides rate	-0.081	-0.060	**-0.096**	-0.065	-0.081	-0.063
Natural resources project	**1.549**	**0.574**	**0.762**	**0.547**	**0.953**	**0.529**
Zone: Suni A		**0.840**		**0.750**		**0.735**
Zone: Suni B		**1.429**		**1.285**		**1.251**
Log Likelihood (neg)	-160.3	-140.7	-147.8	-140.5	-150.3	-140.2

Note: Boldface type denotes probability of Type II error ≤ 5 percent in rejecting H^0: β=0.

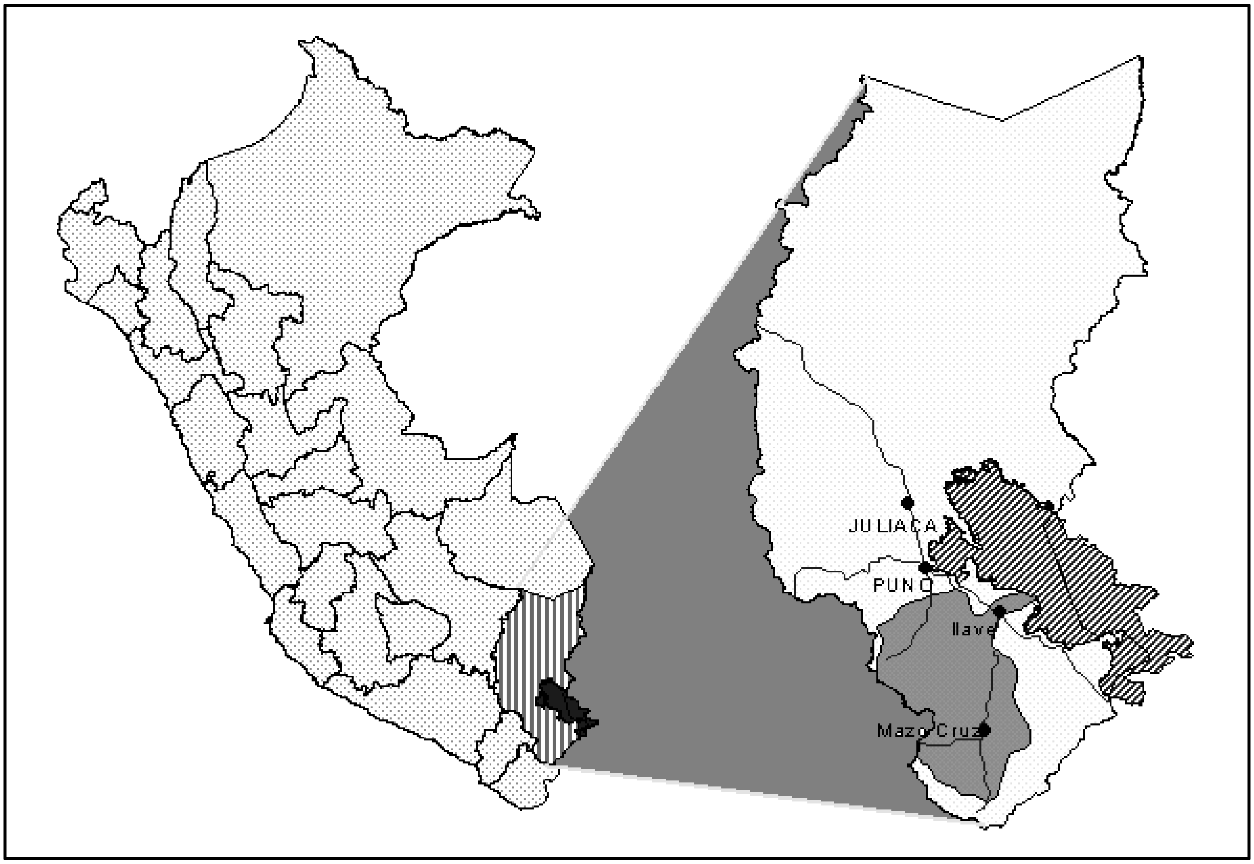

Figure 3: Location of Ilave-Huenque watershed in southern Puno Department, Peru.

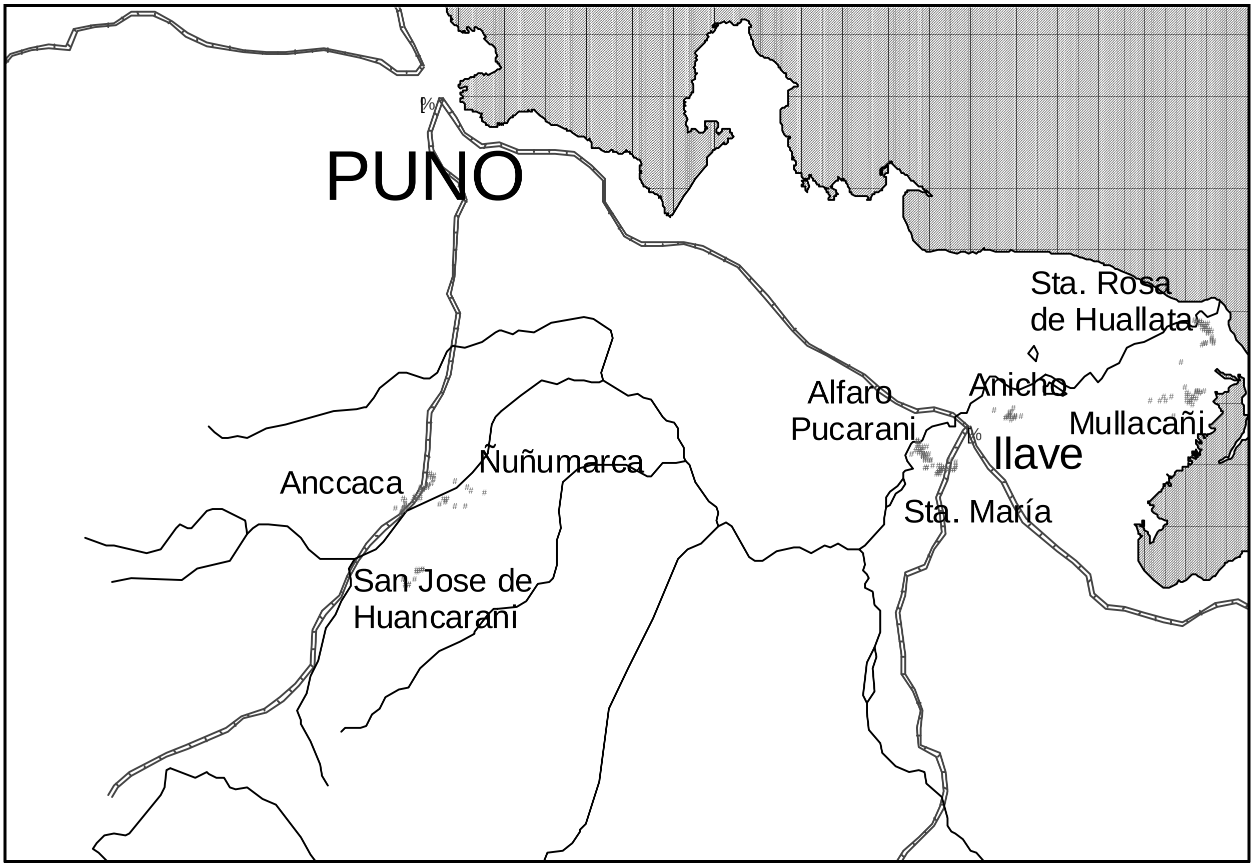

Figure 4: Location of sample households in 3 agro-ecological zones of northern Ilave-Huenque watershed (dark line), Puno, Peru.

References

Anselin, L. *SpaceStat, A Software Program for the Analysis of Spatial Data, Version 1.90*. Ann Arbor, MI: BioMedware, 1999.

Anselin, L. *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic Publishers, 1988.

Aronoff, S. *GIS: A Management Perspective*. Ottawa: WDL Publications, 1998.

Bell, K. P., and N. E. Bockstael. "Applying the generalized-moments estimation approach to spatial problems involving microlevel data." *Review of Economics and Statistics* 82(2000):72-82.

Bockstael, N. E. "Modeling economics and ecology: The importance of a spatial perspective." *American Journal of Agricultural Economics* 78(1996):1168-1180.

Bryk, A. S., and S. W. Raudenbush. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Advanced Qualitative Techniques in the Social Sciences. Edited by J. de Leeuw. Newbury Park, CA: Sage Publications, 1992.

Case, A. "Neighborhood Influence and Technological Change." *Regional Science and Urban Economics* 22(1992):491-508.

Case, A. C. "Spatial Patterns in Household Demand." *Econometrica* 59(1991):953-965.

Deaton, A. *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore: Johns Hopkins, 1997.

Foster, A. D., and M. R. Rosenzweig. "Learning by Doing and Learning from Others - Human-Capital and Technical Change in Agriculture." *Journal of Political Economy* 103(1995):1176-1209.

Havlicek, J., Jr., and J. A. Seagraves. "The 'Cost of the Wrong Decision' as a Guide in Production Research." *Journal of Farm Economics* 44(1962):157-168.

Kelejian, H. H., and I. R. Prucha. "A generalized moments estimator for the autoregressive parameter in a spatial model." *International Economic Review* 40(May, 1999):509-533.

Nelson, G. C., V. Harriss, and S. C. Stone. "Deforestation, Land Use, and Property Rights: Emprirical Evidence from Darién, Panama." *Land Economics* 77(May, 2001):185-205.

Nelson, G. C., and D. Hellerstein. "Do Roads Cause Deforestation? Using Satellite Images in Econometric Analysis of Land Use." *American Journal of Agricultural Economics* 79(February, 1997):80-88.

Pinkse, J., and M. E. Slade. "Contracting in space: An application of spatial statistics to discrete-choice models." *Journal of Econometrics* 85(1998):125-154.

Swinton, S. M., and K. Q. Jones (1998) "From Data to Information: The Value of Sampling vs. Sensing Soil Data." In P. C. Robert, et al., ed. *Precision Agriculture: Proceedings of the 4th International Conference.* St. Paul, MN: Agri-Publications, Inc, pp. 1681-1692

Swinton, S. M., and R. Quiroz (2000) "Are Poor Farmers to Blame for Environmental Degradation? Results from the Peruvian Altiplano." *16o Simposio de la Asociación Internacional de Sistemas de Producción.* (CD-ROM), Santiago, Chile: RIMISP (http://www.rimisp.cl/php/simposio/documentos/325.pdf).

Swinton, S. M., R. A. Quiroz, S. Paredes, J. R. Reinoso, and R. Valdivia (1999) "Using Farm Data to Validate Agro-ecological zones in the Lake Titicaca Basin, Peru." *SAAD-III: Third International Symposium on Systems Approaches for Agricultural Development.* Lima, Peru: International Potato Center (CD-ROM published 2001)

Tapia, M. E. *Ecodesarrollo en los Andes Altos*. Lima: Fundacion Friedrich Ebert, 1996.

Torero, M. (1992) "La adopción de la innovación tecnológica en la agricultura tradicional del Perú: La asociación geográfica como una alternativa para la difusión". In C. I. Degregori, J.

Escobal, and B. Marticorena, ed., *Peru: El problema agrario en debate/SEPIA IV*. Lima: SEPIA, pp. 371-403.

U. S. Department of Agriculture (USDA) (2000). "Summary Report: 1997 National Resources Inventory.". Natural Resources Conservation Service, USDA, Washington, DC. http://www.nhq.nrcs.usda.gov/NRI/1997/summary_report/original/contents.html.

Von Thunen, J. H. *Von Thünen's Isolated State: An English Edition of Der Isolierte Staat*. Translated by C. M. Wartenberg. Edited by P. Hall. Oxford: Pergamon, 1966.