Spatial Autocorrelation and Non-Farm Rural Enterprises in Indonesia

John Gibson
Department of Economics,
University of Waikato
Private Bag 3105, Hamilton
jkgibson@waikato.ac.nz

Susan Olivia
University of California, Davis
olivia@primal.ucdavis.edu

Selected Paper prepared for presentation at the Australian Agricultural and Resource Economics Society Annual Conference, Queenstown, NZ, February 14-16, 2007

Spatial Autocorrelation and Non-Farm Rural Enterprises in Indonesia

John Gibson
Department of Economics,
University of Waikato
Private Bag 3105, Hamilton
jkgibson@waikato.ac.nz

Susan Olivia
University of California, Davis
olivia@primal.ucdavis.edu

Abstract

Non-farm rural enterprises (NFRE) are increasingly studied because of their role in poverty reduction. However, existing studies of the effects of infrastructure on NFRE may give incorrect inferences because they typically fail to account for spatial effects. Such effects could reflect either spatial errors due to excluded local effects or spatial lags due to excluded interactions, such as between households switching out of farm work. We use rural investment climate survey data from Indonesia that allow distances between each household to be measured so that spatial effects can be modeled to assess the bias from ignoring such effects.

Keywords: Infrastructure, Non-farm enterprises, Spatial statistics, Indonesia

1. Introduction

Non-farm rural enterprises (NFRE) are increasingly studied because of their contribution to economic growth, employment generation, livelihood diversification and poverty reduction in developing countries. The combination of rural non-farm self-employment, off-farm wage work and remittances contributes between 30 to 50 percent of rural household income in sub-Saharan Africa (Reardon, 1997) and about one-third of income in Asia (Haggblade, Hazell and Reardon, 2005). This importance is reflected in several recent studies of the determinants of non-farm rural employment and income (Lanjouw, 1999; Berdegué, Ramirez, Reardon, and Escobar, 2001; Corral and Reardon, 2001; Escobal, 2001; Lanjouw, 2001; and Zhu and Luo, 2006).

These studies often pay considerable attention to location factors. One reason is the key role that proximity to urban areas, other major markets, input sources and off-farm labour demand may play in promoting non-farm activity. Another is the location-specificity of infrastructure, since investments in roads, electricity and telecommunications are often cited as transactions costs-reducing interventions that can assist the rural non-farm economy (Zhu and Luo, 2006). A third reason is that much of the information about the importance of NFRE comes from household surveys, whose samples are usually clustered in groups of 10-20 households rather than spread randomly across space. Because households in the same cluster should all face the same locational factors, this clustered design may allow the effects of neighbourhood variables to be measured (Isgut, 2004).

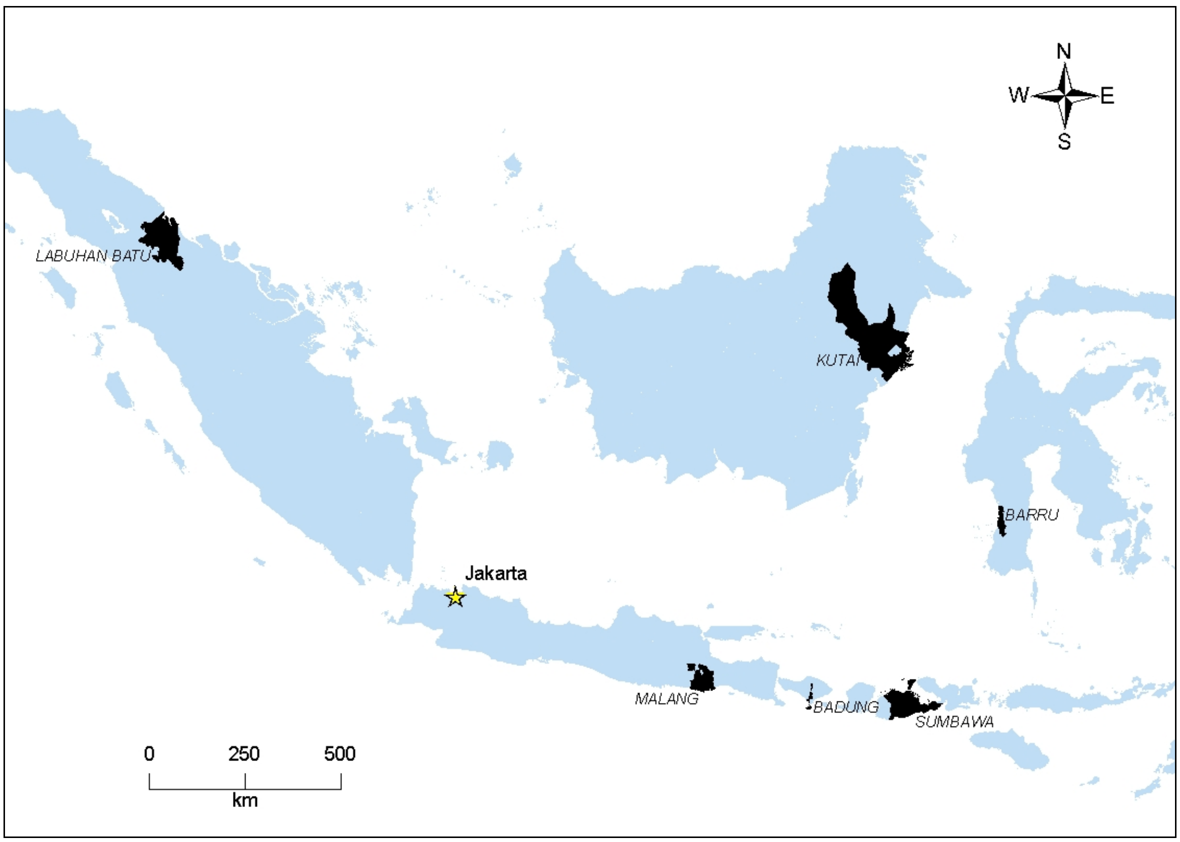
However, the literature has been less careful in reacting to a potential concern about the clustered nature of the survey evidence on NFRE, which is the possibility that incorrect inferences are drawn from econometric results. The response of rural households to exogenous factors, and especially those that are location-specific like infrastructure, may depend on the response of nearby households. This is a form of *spatial autocorrelation*, which can arise either because nearby locations have unobserved factors in common (e.g., access to markets, infrastructure quality) or because of interaction between one household and another (e.g. coordination problems when deciding to switch from farm to non-farm production). The first model, of unobserved common factors, is known as a *spatial error* model while the second, of unaccounted for interactions, is a *spatial lag* model. Ignoring a spatial error structure can cause inference problems while ignoring spatial lags can bias coefficient estimates since the omitted autocorrelation in the lag model enters through the systematic part rather than the random part of the model (Anselin, 1988).

Thus far, these spatial effects have not been accounted for in the literature on NFRE. Instead, some studies base their inferences on standard errors that are robust to survey clustering but there is no evidence to show whether this is sufficient precaution. Some clusters cover a large area while others are small but cluster correction methods are applied equally since location *within* a cluster is rarely known. Moreover, there is no allowance for spatial correlations *between* households in different clusters. To investigate these issues we use rural investment climate survey data from Indonesia that allow distances between each household to be measured. These additional distance variables allow spatial effects to be modeled so that any bias and inferential errors from ignoring such effects can be assessed.

2. The Rural Investment Climate Survey in Indonesia

The Indonesian Rural Investment Climate Survey (RICS) is an in-depth, quantitative survey of 2549 formal and informal non-farm enterprises and 2782 households located in 149 communities (clusters).[1] The survey was conducted in January/February 2006. The sample frame was six selected Kabupaten (districts) that are largely rural and that were designed to represent each of six broad types of economic geography, ranging from rich agricultural areas to forest margin areas. Figure 1 shows the locations of these six districts.

Figure 1: Location of the Sampled Kabupaten in the RICS

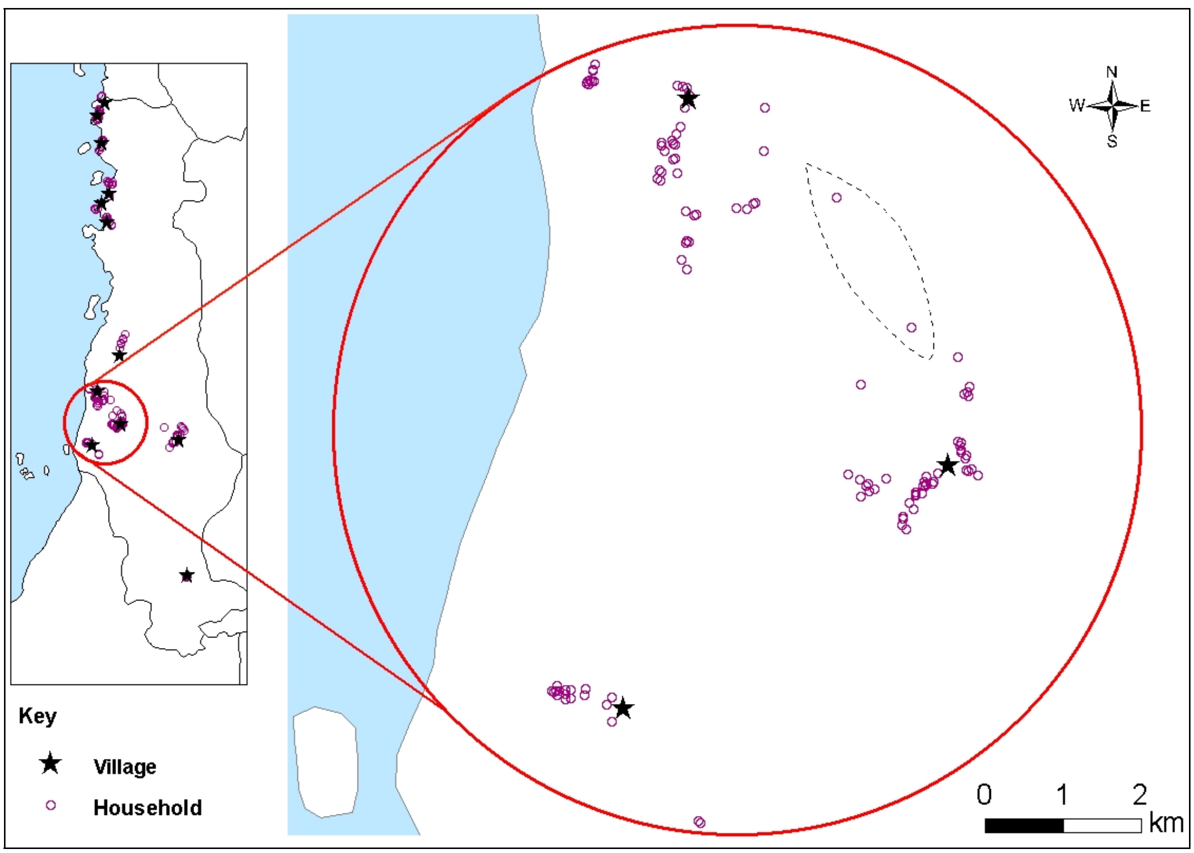


The survey is not designed to be representative of all of Indonesia, which does not matter for the purposes of the current paper. Instead, the key feature of the survey for the current study is that

[1] Over one-half of the sampled enterprises were located in the sampled households.

the locations of each household were recorded using a Global Positioning System (GPS) receiver and these coordinates allow the distance of every household to every other household to be measured. Moreover, because of the variation in the area of each Kabupaten and the approximately equal sample size in each one, there is variation in how far apart the sample clusters are, which may help in estimating between cluster spatial autocorrelation. Indeed, there are some cases of clusters being sufficiently close together that the distance between households in different clusters is smaller than the distance within a cluster (Figure 2). This feature may matter since previous studies ignore any spatial autocorrelation that occurs between households in separate clusters.

Figure 2: Example of Intra-cluster Distances Exceeding Inter-cluster Distances



To maintain comparability with much of the literature, which uses household surveys rather than enterprise surveys to study NFRE, we only use the household sample and restrict attention to the households in rural areas of the selected Kabupaten. This gives a sample of 1600 households in 97 clusters.[2] This average of 16 sampled households per cluster is similar to the clustering in many other rural household surveys.

The topical coverage of the survey includes household demographic and economic characteristics, business experience, business constraints, and community characteristics. There is a particular focus on different aspects of the rural investment climate including: infrastructure, credit, the diffusion of technical knowledge, marketing and competition, and local governance. A variety of household level variables (including those indicating characteristics of the household head) and community level variables are included in the regression model reported below. The community variables are (by definition) common for households in the same cluster and we are interested to see whether this makes them especially susceptible to misspecification from a failure to include relevant spatial error or lag terms in the regression model.

Table 1 shows that for the RICS rural households, the share of total household income from NFRE is 37 percent. The means of the household and household head characteristics for the overall sample are also presented in Table 4. It is notable that even with the high NFRE income share, three-quarters of the households owned land. In addition, Table 4 also shows the community characteristics, in the form of access to various forms of infrastructure and the

[2] Two clusters were dropped prior to the selection of the final sample because the community center was more than 20 kilometres from the location of most of the households in each cluster, suggesting a problem in the GPS measurements.

quality of that infrastructure as well as variables that may reflect the quality of the local business environment (such as crime and whether there is a local business association).

Table 1. Descriptive Characteristics for the Estimation Sample

Variable	Mean	Std. Dev
Dependent variable		
Share of Household income from non-farm enterprises	0.37	0.38
Household Characteristics		
Age of household head	45.68	13.00
Female household head	0.10	0.30
Married household head	0.87	0.34
Household head with tertiary education	0.07	0.26
Muslim household head	0.91	0.29
Household size	4.36	1.79
% of household who are adults 17+	0.68	0.22
% of household who are children 0 – 10 yrs	0.19	0.19
Dummy if household owns land	0.75	0.43
Per capita household income (log)[a]	7.99	1.08
Location and infrastructure characteristics		
Number of households in the village (log)	6.76	0.52
Local business association (=1,0 otherwise)	0.14	0.35
Crime/dispute occurred in village last year (=1,0 otherwise)	0.73	0.45
Co-operative present in village (=1,0 otherwise)	0.34	0.47
Distance to co-operative[b]	7.45	41.77
Distance to sub-district headquarters[b]	9.25	14.45
Electricity blackout < 30 minutes per day (=1, 0 otherwise)	0.82	0.38
No landline or cell phone access in village (=1, 0 otherwise)	0.10	0.29
Roads inside and out of village are unsealed (=1, 0 otherwise)	0.20	0.40
Total observation	1,600	

Notes: [a] in Rupiah; [b] in kilometer.

3. Testing for Spatial Autocorrelation

3.1 Methods of Detecting Spatial Autocorrelation

As noted above, the importance of non-farm enterprises as a rural household income source is likely to be (positively) spatially autocorrelated. More generally, positive spatial autocorrelation occurs when high or low values for a random variable tend to cluster in space. Such a sample contains less information than an uncorrelated one so inference errors may occur if this is not accounted for. Moreover, ignoring these spatial interactions may also cause omitted variable bias (Anselin and Bera, 1998).

A key issue in adjusting for this spatial autocorrelation is that some structure has to be imposed on the data. While it is hypothetically possible for the decisions of a household regarding non-farm enterprises to be influenced by the decisions of all other households in the sample, as a practical matter many of these bilateral relationships are likely to be zero. Moreover, not all of these potential interactions can be estimated. For example, with a cross-sectional sample of size N there would be $N \times N$ correlations to estimate along with the β and σ^2 parameters of a standard regression model, which exceeds the number of observations.

A *spatial weight matrix* is one way of imposing the required structure on the study of spatial autocorrelation. This is an $N \times N$ positive and symmetric matrix which exogenously determines for each observation (row) which locations (columns) belong in its neighbourhood. For non-neighbours, $w_{ij}=0$, while for neighbours the weights are either $w_{ij}=1$ (binary weights) or a function of something else, such as: $w_{ij} = 1/d_{ij}$ where d_{ij} is the distance between observations i and j (inverse distance weights). The diagonal elements of the weights matrix are conventionally set to zero, and typically standardised such that the elements of a row sum to one (Anselin and Bera, 1998). Hence, the spatial weight matrix allows all of the interactions between observation i and each of its neighbours to be parameterized in the form of a weighted average. Specifically, for some random variable of interest z, each element of the spatially lagged variable Wz equals $\sum_j w_{ij} z_j$ which is a weighted average of the z values in the neighbourhood of point i.

The spatial weight matrix is used by both main approaches for incorporating spatial effects into regression models: the *spatial lag* model and the *spatial error* model. Formally, the spatial lag model is defined as:

$$Y = \rho WY + X\beta + \varepsilon \tag{1}$$

where Y is an $N\times1$ vector of observations on the dependent variable, WY is the spatially lagged dependent variable, X is an $N\times k$ matrix of explanatory variables, ε is a vector of errors, β is the vector of regression parameters and ρ is the spatial autoregressive parameter. Although equation (1) looks like a dynamic model from time-series econometrics, one key difference causes OLS to always be an inconsistent estimator of the spatial lag model. In the time-series context, if there is no serial correlation in the errors, ε_t there will be no correlation between y_{t-1} and ε_t and OLS will be a consistent estimator. In contrast, $(WY)_i$ is always correlated with both ε_i and the error term at all other locations. Hence, OLS is not consistent for the spatial lag model and either a maximum likelihood or instrumental variables estimator is needed (Anselin, 1988).

In contrast to the spatial lag model, the spatial error model is defined as:

$$\begin{aligned} Y &= X\beta + \varepsilon \\ \varepsilon &= \lambda W\varepsilon + \mu \end{aligned} \tag{2}$$

where λ is the spatial autoregressive coefficient, μ is a vector of errors that are assumed to be independently and identically distributed and the other variables and parameters are as defined in equation (1). In this model, the error for one observation depends on a weighted average of the errors for neighbouring observations, with λ measuring the strength of this relationship.

It is clear that both equations (1) and (2) are restricted versions of a more general spatial autoregressive model with autoregressive disturbances:

$$\begin{aligned} Y &= \rho W_1 Y + X\beta + \varepsilon \\ \varepsilon &= \lambda W_2 \varepsilon + u \end{aligned} \tag{3}$$

It may therefore seem preferable to always begin with a model like equation (3) and test in a general-to-specific way to see if either equation (2) or equation (1) are data-acceptable. Indeed, equation (3) could always be the starting point for cross-sectional regressions because the standard OLS regression model:

$$Y = X\beta + \varepsilon \tag{4}$$

is just a special case with ρ=λ=0. However, spatial models are much more computationally demanding and for most econometric software there are limits on the sample sizes that they can accommodate (due to the need to form a weights matrix of order $N \times N$). Moreover, they have to be estimated by methods such as instrumental variables and maximum likelihood that require additional assumptions.

It is possible to use Lagrange Multiplier (LM) tests for spatial autocorrelation, which only need the restricted model to be estimated. Therefore it is common in the spatial econometrics literature to start with an OLS model and use the residuals from that model to test against spatial alternatives. In addition to these LM tests, Moran's *I* test, which has some parallels with the Durbin-Watson statistic, is also widely used (Anselin and Bera, 1998). For a row-standardized spatial weight matrix, Moran's *I* can be expressed as:

$$I = \frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{\mathbf{e}'\mathbf{e}} \tag{5}$$

where $\mathbf{e}$ is a vector of OLS residuals and $\mathbf{W}$ is the spatial weight matrix. Moran's I is asymptotically normally distributed with mean $-1/(N-1)$ and its statistical significance can be evaluated from a standardized normal table. A feature of Moran's I is that the alternative hypothesis does not specify the process generating the autocorrelated disturbances. However, there is a simple intuition for Moran's I because for any variable $\mathbf{z}$ in deviation from mean form, I is equivalent to the slope coefficient in a linear regression of $\mathbf{Wz}$ on $\mathbf{z}$ (Anselin, 1995).

The LM tests are based on explicitly specified alternative hypotheses. For testing OLS against the spatial error model (λ=0) the test statistic is:

$$LM_{\lambda} = \left[\mathbf{e}'\mathbf{We}/\hat{\sigma}^2\right]^2 / T \tag{6}$$

where $T = tr(W' + W)W$ and LM_{λ} is distributed as χ^2 with 1 degree of freedom. For testing OLS against the spatial lag model (ρ=0) the test statistic is:

$$LM_{\rho} = \left[\mathbf{e}'\mathbf{WY}/\hat{\sigma}^2\right]^2 / T_1 \tag{7}$$

where $T_1 = (\mathbf{WX}\hat{\beta})'\mathbf{M}(\mathbf{WX}\hat{\beta})/\hat{\sigma}^2 + T$ and $\mathbf{M} = \mathbf{I} - \mathbf{X(X'X)^{-1}X'}$. One difficulty with both LM_{λ} and LM_{ρ} is that they each have power against the other alternative. In other words, when testing λ=0, LM_{λ} responds to nonzero ρ and when testing ρ=0, LM_{ρ} responds to nonzero λ. To test in the possible presence of both spatial error and spatial lags, Anselin et al. (1996) develop specification tests for spatial lags that are robust to ignored spatial errors and tests for spatial errors that are robust to ignored spatial lags. These tests denoted LM_{λ}^* and LM_{ρ}^* should be used when both LM_{λ} and LM_{ρ} are statistically significant.

All five of the spatial autocorrelation tests described here are used in the current study. Depending on the outcome of the specification tests, the regression model for the household income shares from rural non-farm enterprises will be re-estimated in either the spatial lag or spatial error framework.

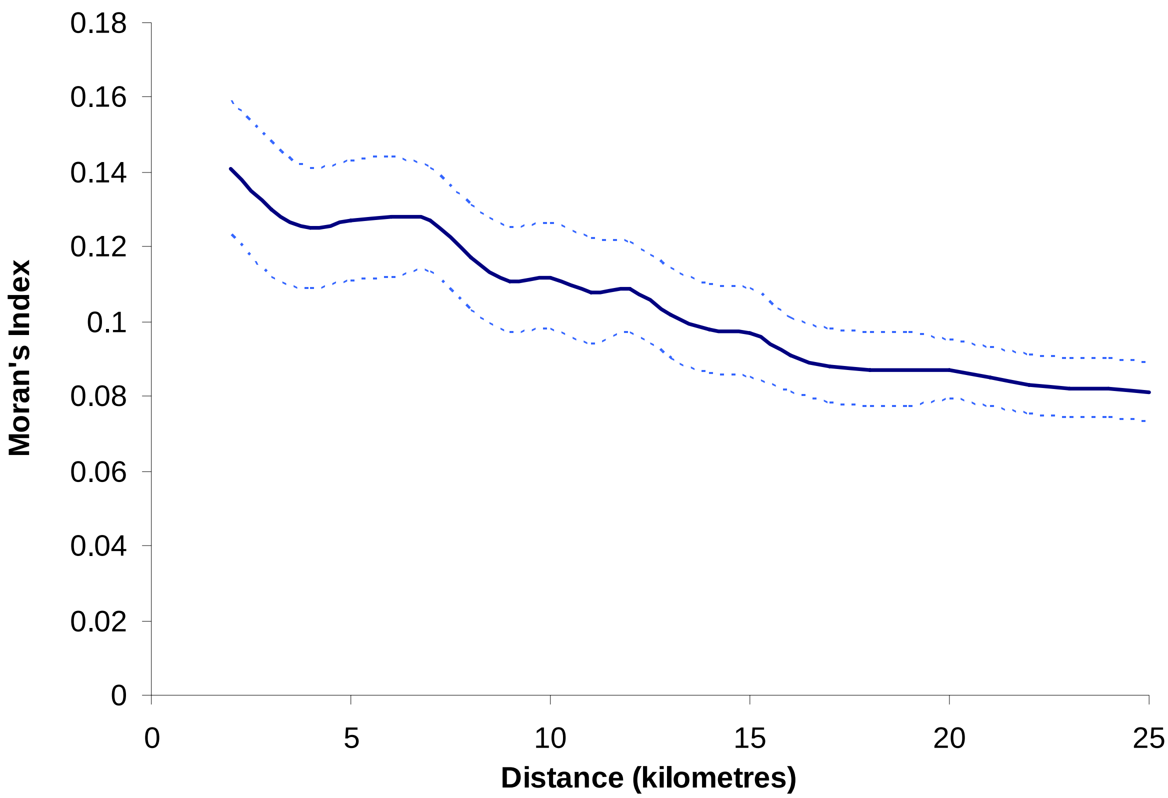
3.2 Results of Testing for Spatial Autocorrelation Effects

Results are presented initially for tests of spatial autocorrelation in the household income shares coming from rural non-farm enterprises. This spatial autocorrelation in a dependent variable need not cause any concern if it is removed by the covariates in the regression model. Therefore, the results of testing the residuals of an OLS income share equation are also reported.

A spatial weights matrix is needed to test for spatial autocorrelation and in turn this requires a measure of distance between households. Latitude and longitude coordinates for each household were used to calculate this and then Moran's Index for income shares from non-farm rural enterprises was calculated for varying neighbourhood sizes of 1-25 km. Note that the average distance within clusters from each household to the village centre is only 0.8 km and the largest distance between any two households in a given cluster averages 1.9 km. Hence this range allows for spatial autocorrelation that extends far beyond the boundary used in previous studies that can account only for correlations within clusters.

When households within a one kilometre radius of location *i* are considered as the neighbourhood, Moran's I=0.15 and it is statistically significant (p<0.01). In other words, a regression of the spatially weighted average share of non-farm enterprise income in household total income within this neighbourhood, *Ws* on the income share of each household, *s*, would have a statistically significant coefficient of 0.15. The strength of the spatial autocorrelation declines only slightly as the neighbourhood is defined to include a larger area, equalling 0.11 at a ten kilometre radius and 0.08 at 25 km. For all neighbourhood sizes considered Moran's *I* is still statistically significant.

Figure 3: Spatial Correlation in Rural Household Income Shares from Non-Farm Enterprises



Source: Author's calculation using Moran's *I* statistic described in the text. Broken lines are +/- 2 std errors.

To see whether this spatial autocorrelation is also transmitted to the residuals of an OLS regression, an income share model was estimated using variables typically found in NFRE studies in the literature. These variables included characteristics of the household head (age, gender, religion, marital status and education), characteristics of the household (size, composition, land ownership and income), and community characteristics.

Table 2: OLS, Spatial Error and Spatial Lag Estimates of the Equation for Non-Farm Enterprise Income Shares for Rural Households

	Robust std errors	Clustered std errors	Spatial error model	Spatial lag model
Age of household head	-0.002	-0.002	-0.001	-0.002
	(2.26)*	(1.75)+	(1.85)+	(2.08)*
Married household head	0.090	0.090	0.109	0.096
	(2.31)*	(2.24)*	(2.90)**	(2.56)*
Female household head	0.125	0.125	0.130	0.125
	(2.82)**	(3.03)**	(3.04)**	(2.90)**
Tertiary educated household head	-0.102	-0.102	-0.104	-0.103
	(2.82)**	(3.27)**	(2.95)**	(2.93)**
Household head is Muslim	-0.089	-0.089	-0.085	-0.082
	(1.60)	(1.65)	(1.49)	(1.48)
Household size	0.023	0.023	0.023	0.022
	(3.71)**	(3.29)**	(3.85)**	(3.75)**
% of household who are adults 17+	-0.063	-0.063	-0.054	-0.060
	(0.96)	(0.91)	(0.86)	(0.94)
% of household children 0-10 years	-0.018	-0.018	-0.000	-0.009
	(0.27)	(0.27)	(0.00)	(0.13)
Household owns land (=1, else 0)	-0.044	-0.044	-0.040	-0.040
	(1.97)*	(1.42)	(1.71)+	(1.86)+
log (per capita household income)	0.112	0.112	0.119	0.111
	(11.65)**	(8.60)**	(12.15)**	(11.86)**
log (# of households in village)	0.102	0.102	0.097	0.073
	(4.41)**	(2.31)*	(3.24)**	(3.18)**
Village has business association	0.103	0.103	0.117	0.083
	(3.44)**	(1.62)	(2.99)**	(2.94)**
Village had crime/dispute last year	-0.080	-0.080	-0.078	-0.061
	(3.57)**	(2.55)*	(2.61)**	(2.71)**
Village has a cooperative	0.040	0.040	0.039	0.031
	(1.63)	(1.07)	(1.19)	(1.31)
Distance to cooperative (km)	-0.490	-0.490	-0.451	-0.377
	(2.74)**	(2.09)*	(1.83)+	(2.08)*
Distance to sub-district (km)	-1.284	-1.284	-1.314	-1.045
	(1.92)+	(1.32)	(1.45)	(1.59)
Low blackouts (< 30 minutes/day)	-0.051	-0.051	-0.054	-0.040
	(1.81)+	(1.06)	(1.45)	(1.48)
Village has no telephones	0.057	0.057	0.056	0.044
	(1.41)	(0.90)	(1.03)	(1.12)
Roads in and out of village unsealed	0.041	0.041	0.044	0.034
	(1.40)	(0.92)	(1.14)	(1.18)
Constant	-0.939	-0.939	-1.011	-0.889
	(4.53)**	(2.44)*	(4.21)**	(4.47)**
Lambda (spatial error model)			0.285	
			(7.01)**	
Rho (spatial lag model				0.247
				(6.13)**
R-squared	0.16	0.16		
Log-likelihood function	-566.32	-566.32	-541.68	-546.77

Note: Spatial models use inverse distance weights for a 10 kilometre neighbourhood. **=sig at 1% level, * = 5%, +=10%.

When OLS is used as the estimator for the regression and robust (but not clustered) standard errors are calculated, it appears that the non-farm enterprises make up a larger share of rural household income for households with younger heads who are either married or female and lack tertiary education, and for larger and richer households who do not own land (Table 2, column 1). In terms of community characteristics, households in larger villages with a village business association appear to have higher NFRE income shares, while those further from both cooperatives and the sub-district headquarters and with an experience of crime or some other dispute have lower NFRE income shares. One community variable has an unexpected (and weakly significant) sign; the NFRE income share is lower for households living in villages where electricity blackouts are not very long-lasting (< 30 minutes/day) compared with similar households in villages with longer lasting blackouts.

The inferences change somewhat when the standard errors are re-calculated to take account of the sample clustering. In general, the clustered standard errors are larger and four variables that appeared statistically significant with the robust standard errors become insignificant; land ownership, whether the village has a local business association, the distance to the sub-district office and the prevalence of blackouts. It is notable that the correction for clustering has most effect on the locational variables, since these are inherently clustered by construction. This suggests that inferences about the effect of infrastructure and other location-specific attributes on NFRE may be sensitive to the treatment of clustering and more general spatial effects.

How reliable are the inferences coming from the OLS results in the first two columns of Table 2 in terms of ignored spatial autocorrelation? Tests using the methods described in Section 3.1

were used, with two different types of weights – binary and inverse distance – for neighbourhoods varying from 10 to 40 kilometres. According to these tests there is substantial evidence of misspecification in the OLS results (Table 3).

Table 3: Specification Tests for Spatial Autocorrelation in the OLS Residuals of the NFRE Income Share Regression

Type of weighting matrix	Moran's *I*	LM_λ	LM^*_λ	LM_ρ	LM^*_ρ
Inverse distance weights					
10 kilometre neighbourhood	8.64***	59.09***	17.26***	45.84***	4.00**
20 kilometre neighbourhood	8.60***	58.24***	16.68***	45.36***	3.80*
30 kilometre neighbourhood	8.54***	57.27***	16.86***	44.36***	3.95**
40 kilometre neighbourhood	8.47***	56.18***	16.85***	43.36***	4.02**
Binary weights					
10 kilometre neighbourhood	11.01***	71.70***	7.94***	64.84***	1.08
20 kilometre neighbourhood	14.17***	92.83***	17.66***	75.96***	0.79
30 kilometre neighbourhood	14.05***	67.68***	5.42***	67.12***	4.86**
40 kilometre neighbourhood	12.12***	32.02***	0.76	35.97***	3.73*

Note: ***=p<0.01, **=p<0.05, *=p<0.10.

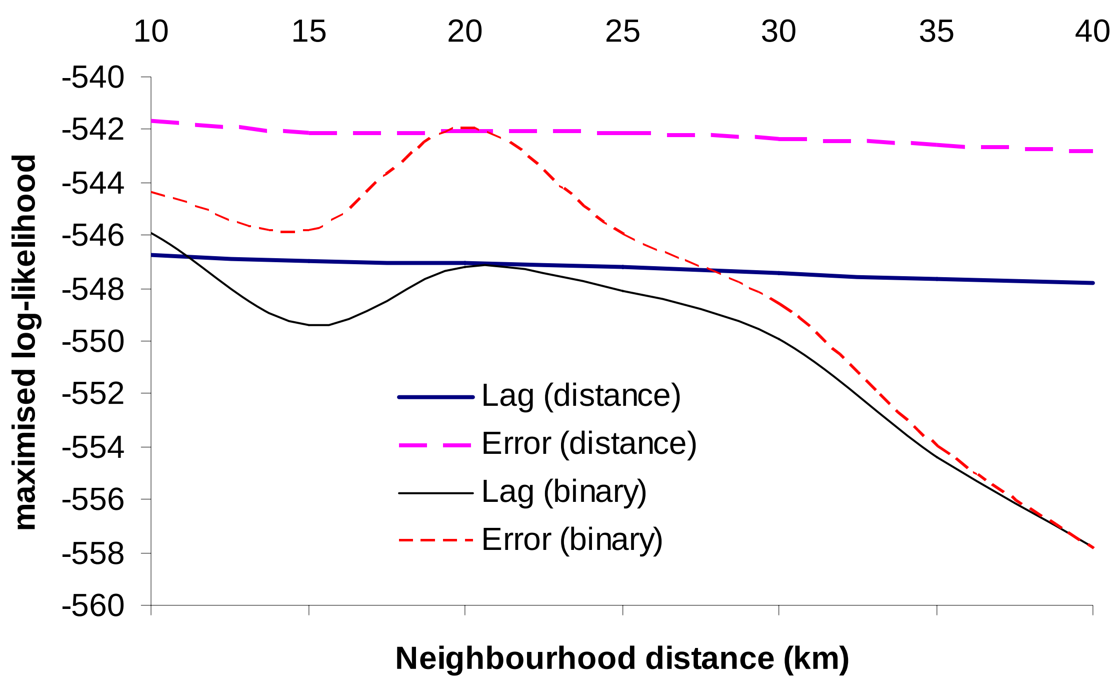
Both the LM_λ and the LM_ρ tests are statistically significant, so it is necessary to use the specification tests for spatial lags that are robust to unaccounted for spatial errors, and the tests for spatial errors that are robust to ignored spatial lags. According to these robust tests there is less evidence in favour of the spatial lag model (i.e., the values of LM^*_λ are almost always above the threshold for statistical significance while those for LM^*_ρ are sometimes below the threshold). In other words, inferences from the OLS estimates of the income share model in Table 2 are likely to be incorrect because of the unmodeled spatial autocorrelation in the residuals.

4. Spatial Regression Results

In light of the above results about the misspecification when OLS is used, a variety of spatial lag and spatial error models were estimated. Seven neighbourhood sizes were chosen, ranging from the minimum feasible (to prevent “islands” with no neighbours) of 10 kilometres to a maximum of 40 kilometres. Both binary and inverse distance weights were considered.

A comparison of the maximised log likelihoods of the resulting models indicated that there was better performance for the spatial error model than the spatial lag model (Figure 4). This is consistent with the results of the specification testing in Table 3. The log-likelihood was also higher when the spatial weight matrix was based on inverse distance rather than a simple 0/1 set of weights. When distance weights were used the model fit declined as the neighbourhood size increased and this was also clear for the binary weights for neighbourhoods greater than 20 kilometres.

Figure 4: Log-Likelihood Values for Spatial Regression Models with Binary and Inverse Distance Spatial Weight Matrices and Neighbourhoods of Different Sizes



Source: Author’s calculation using spatial regression methods described in the text.

The last two columns of Table 2 contain results of the spatial error and spatial lag models (based on inverse distance weights and a neighbourhood size of ten kilometres). According to the maximum likelihood estimates, in the preferred spatial error specification λ=0.29, with a standard error of 0.04.[3] In other words, the spatially weighted residual NFRE share within a ten kilometre radius is significantly associated with the residual income share for a particular household even after controlling for household characteristics and limited set of location attributes (infrastructure access and quality and the local business environment).

When the spatial error model is used, standard errors are generally smaller than for the clustered standard errors. In other words, the formula for clustered standard errors which is based on heteroscedasticity and autocorrelation within clusters of an unknown form appears to be too conservative. These differences could matter since two variables once again become statistically significant once inferences are based on the spatial error model; whether the household owns land and whether the village has a business association. Although the spatial lag model is not favoured over the spatial error model, it would also give similar inferences but with different coefficient values (since the spatial autocorrelation enters through the systematic part of the model).

5. Conclusions

Inferences about the household and locational factors that affect the importance of non-farm enterprises to rural households in developing countries may be sensitive to spatial effects. In the current example, use of clustered standard errors appears to lead to inferences that are too conservative. Further work is needed to see if this pattern holds elsewhere.

[3] In the spatial lag model λ=0.25, with a standard error of 0.04.

References

Berdegué, J., Ramirez, E., Reardon, T., and Escobar, G. (2001). "Rural nonfarm employment and incomes in Chile" *World Development* 29(3): 411-425.

Corral, L. and Reardon, T. (2001). "Rural nonfarm incomes in Nicaragua" *World Development* 29(3): 427-442.

Ellis, F. (2000). "The determinants of rural livelihood diversification in developing countries" *Journal of Agricultural Economics* 51(2): 289-302.

Escobal, J. (2001). "The determinants of nonfarm income diversification in rural Peru" *World Development* 29(3): 497-508.

Haggblade, S., Hazell, P., and Reardon, T. (2005). "The rural nonfarm economy: Pathway out of poverty or pathway in?" Proceedings of a Research Workshop on: *The Future of Small Farms*, Wye, UK, June 26-29, 2005. International Food Policy Research Institute.

Isgut, A. (2004). "Non-farm income and employment in rural Honduras: assessing the role of locational factors" *Journal of Development Studies* 40(3): 59-86.

Lanjouw, P. (1999). "Rural nonagricultural employment and poverty in Ecuador" *Economic Development and Cultural Change* 48(1): 91-122.

Lanjouw, P. (2001). "Nonfarm employment and poverty in rural El Salvador" *World Development* 29(3): 529-547.

Reardon, T. (1997). "Using evidence of household income diversification to inform study of the rural nonfarm labour market in Africa" *World Development* 25(5): 735-747.

Reardon, T., Berdegué, J., and Escobar, G. (2001). "Rural non-farm employment and incomes in Latin America: Overview and policy implications" *World Development* 29(3): 395-409.

Zhu, N. and Luo, X. (2006). "Nonfarm activity and rural income inequality: A case study of two provinces in China" *World Bank Research Working Paper* No. 3811, The World Bank, Washington DC.