United Nations University

WIDER

World Institute for Development Economics Research

Discussion Paper No. 2003/73

Adverse Geography and Differences in Welfare in Peru

Javier Escobal and Máximo Torero*

October 2003

Abstract

In Peru, a country with an astonishing variety of different ecological areas, with 84 different climate zones and landscapes, with rainforests, high mountain ranges and dry deserts, the geographical context may not be all that matters, but it could be very significant in explaining regional variations in income and poverty. The major question this paper tries to answer is: what role do geographic variables, both natural and man-made, play in explaining per capita expenditure differentials across regions within Peru? How have these influences changed over time, through what channels have they been transmitted, and has access to private and public assets compensated for the effects of an adverse geography?

We have shown that what seem to be sizeable geographic differences in poverty rates in Peru can be almost fully explained when one takes into account the spatial concentration of households with readily observable non-geographic characteristics, in particular public and private assets. In other words, the same observationally equivalent household has a similar expenditure level in one place as another with different geographic characteristics such as altitude or temperature. This does not mean, however, that geography is not important but that its influence on poverty, expenditure level and growth differential comes about through a spatially uneven provision of public infrastructure…/…

Keywords: regional economics, spatial distribution, welfare, poverty, Peru
JEL classification: D91, R11, Q12

*Grupo de Análisis para el Desarrollo (GRADE)

This study has been prepared within the UNU-WIDER project on Spatial Disparities in Human Development, directed by Ravi Kanbur and Tony Venables.

UNU-WIDER gratefully acknowledges the financial contributions to the 2002-2003 research programme by the governments of Denmark (Royal Ministry of Foreign Affairs), Finland (Ministry for Foreign Affairs), Norway (Royal Ministry of Foreign Affairs), Sweden (Swedish International Development Cooperation Agency-Sida) and the United Kingdom (Department for International Development).

Furthermore, when we measured the expected gain (or loss) in consumption from living in one geographic region (i.e. coast) as opposed to living in another (i.e. highlands), we found that most of the difference in log per capita expenditure between the highland and the coast can be accounted for by the differences in infrastructure endowments and private assets. This could be an indication that the availability of infrastructure could be limited by the geography and therefore the more adverse geographic regions are the ones with less access to public infrastructure. It is important to note that there appear to be non-geographic, spatially correlated, omitted variables that need to be taken into account in our expenditure growth model. Therefore poverty reduction programmes that use regional targeting do have a rationale even if geographic variables do not explain the bulk of the difference in regional growth, once we have taken into account differentials in access to private and public assets.

Acknowledgements

This research was supported by the IDB Research Network and the Global Development Network Award. We are greatly indebted to John Gallup, Eduardo Lora, Alejandro Gaviria and colleagues of the 'Geography and Development' project for numerous helpful comments to the different drafts of this work. Their comments and criticism improved it substantially. We are also grateful to Jorge Aguero for excellent research assistance in this project. All errors remaining are ours.

UNU World Institute for Development Economics Research (UNU-WIDER)
Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Camera-ready typescript prepared by Lorraine Telfer-Taivainen at UNU-WIDER
Printed at UNU-WIDER, Helsinki

ISSN 1609-5774
ISBN 92-9190-532-1 (printed publication)
ISBN 92-9190-533-X (internet publication)

1 Introduction

In *The Wealth and Poverty of Nations* David S. Landes argues that Europe's temperate climate encouraged hard work and capitalist development, while the heat of the tropics brought reliance on slaves (Eichengreen 1998). Engerman and Sokoloff (1997), trying to explain why the United States and Canada have been so much more successful over time than other New World economies suggest that the roots of these disparities lay in differences in the initial factor endowments of the respective colonies. Why do we see areas with persistently low living standards, even in growing economies? Will the legacy of these differences persist?

One view is that differences arise from persistent spatial concentrations of individuals with personal attributes inhibiting growth in their living standards. This view does not ascribe a causal role to geography per se. In other words, identical individuals will have the same growth prospects regardless of where they live. Alternatively, one might argue that geography has a causal role in determining how household welfare evolves over time, and that geographic externalities arising from natural geographic characteristics, local public assets, or local endowments of private assets, entail that living in a well endowed area means that a poor household can eventually escape poverty. Yet an otherwise identical household living in a poor area experiences stagnation or decline. If this is so, then it is important to understand which geographic factors matter for growth at the micro level (Jalan and Ravallion 1998; Engerman and Sokoloff 1997).

Peru has an astonishing variety of ecological areas. Only a few countries offer so many climate zones and landscapes, with rainforests, mountain ranges and deserts. Peru contains a total of 84 of the world's 104 known ecological regions and 28 different climates. This geographic diversity, its link to development, and the important differences in the welfare of the different regions makes Peru a good case study in attempting to ascertain what role geographic variables—both natural and man-made—play in explaining per capita expenditure differentials across regions within Peru.

As shown in Table 1, when comparing the income per capita and consumption per capita differences between the diverse regions of the country, it is clear that Peru has one of the highest degrees of inequality between regions in Latin America. According to the World Bank (1999) and our own estimates based on the Peruvian LSMS of 1997, Peru has a larger dispersion of per capita income by region than Colombia, Brazil, Chile or Mexico. Only Argentina is reported as having larger regional income disparities. Furthermore, this dispersion is also very large within the different geographical regions of Peru.

This paper attempts to understand whether geographic externalities arising from natural geographic characteristics have a causal role in determining how household welfare evolves. The paper is divided into six sections. Section 2 gives a detailed description of

Peru's geography and specifically the main areas in which geography might play a fundamental role in economic development. It also makes a first attempt to analyse whether there is a correlation between geographic variables and earning levels. Additionally, it analyses whether the differences observed across the different regions in Peru are also correlated to the changes in geography and therefore to geographic externalities.

Table 1: Regional income per capita dispersion in Latin American countries (selected years)

Country	Year	Dispersion
Colombia	1989	0.358
Brazil	1994	0.424
Chile	1994	0.470
Mexico	1993	0.502
Peru	1997	0.561
Argentina	1995	0.736

Note: Unweighted coefficient of variation.
Source: World Bank (1999) and LSMS (1997).

In Section 3 we try to formally answer whether geography is a determinant of the evolution of welfare across households over time. We developed a model of consumption growth at the province level. This model not only takes in the local effect of geographic variables but also includes spatial econometric techniques to ascertain the presence of persistent spatial concentrations forced by geography. In addition, we also analyse whether the presence of positive geographic externalities arising from local public assets, or local endowments of private assets, implies that the effect of natural geographic characteristics can be overcome and a poor household could eventually escape poverty. To be able to analyse the partial effects of each of these types of assets (geographic, private and public assets) we also develop a methodology to break down the partial effects of each of these variables.

Section 4 details the main databases constructed for this paper and the methodological issues regarding the databases. We use the National Population Census for 1972, 1981 and 1993, the Living Standard Measurement Study (LSMS) surveys for 1991, 1994, 1996, 1997, information from the district infrastructure census; geographical datasets, and information from the III National Agrarian Census of 1994. In Section 5, the results are presented and we detail the major conclusions of the study.

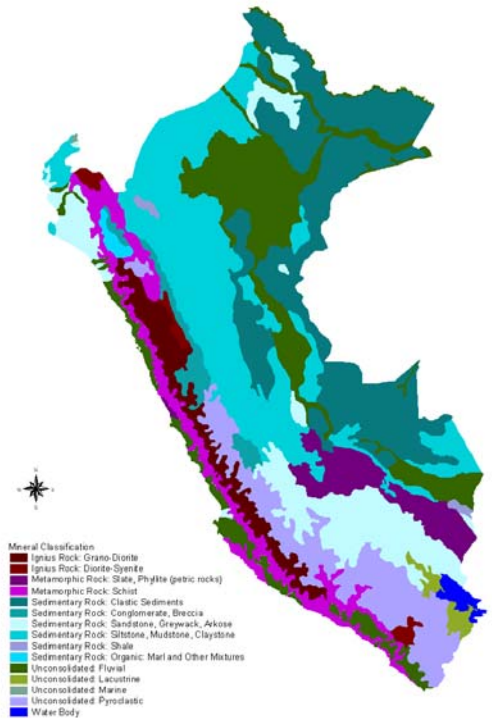
2 Basic characteristics of the Peruvian geography

Leading historians and economists have long recognized geography as having a crucial role in economic development, even though geography has been neglected in most recent empirical studies of comparative growth across countries and of comparative growth within the same country.[1] Specifically, in the case of Peru the enormous diversity of its geography makes it an extremely interesting case study to analyse the importance of these variables to economic growth within the country.[2] Peru is located in a tropical part of the globe, but because of variations in relief (particularly elevation as shown in Map 1) and such factors as rain shadows, bodies of water (i.e. marine currents such as El Niño and Humboldt) and wind patterns, it comprises a multitude of microclimates. Although many geographic factors interact, it can be said that throughout most of Peru the orography and the morphologic structure of the Andes have conditioned the local climate, the type and use of the land, and the agricultural activities of the country.

Map 1: Major landforms in Peru

Source: Authors.

Map 2: Underlying surface composition in Peru



Source: Authors.

The entire coastal area of Peru (around 11 per cent of its territory with 49 per cent of the total population)[3] is one of the world's driest regions. Cold waters off the coast and the proximity of the Andes, as well as high-pressure wind patterns, contribute to the virtual

---

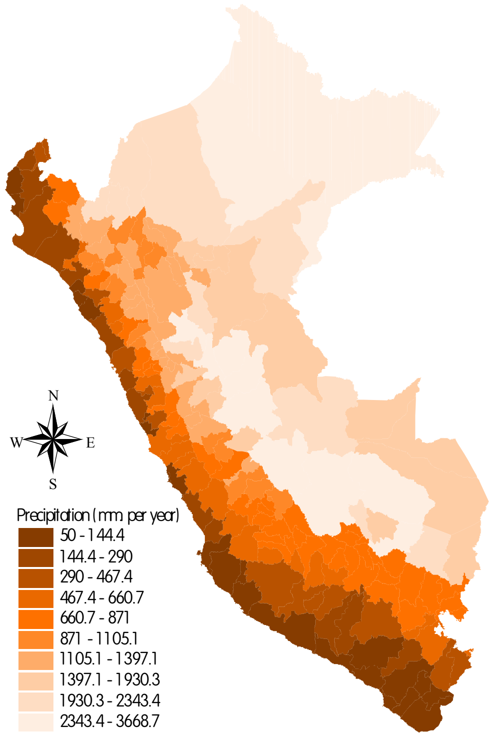
[1] There are few studies estimating the economic importance of geography within a region or a country, for example Bloom and Sachs (1998) make a great contribution for the case of Africa, and Engerman and Sokoloff (1997) for Canada and the USA.

[2] There are several papers—for example Hall and Jones (1997, 1998); Gallup *et al.* (1998); Moreno and Trehan (1997); Davies and Weinstein (1996)—that have tried to answer the question of the importance of geography in explaining the levels of economic activity across countries.

[3] Whereas the 'selva' (mountains) represents 58 per cent of the territory but holds only 7 per cent of the population.

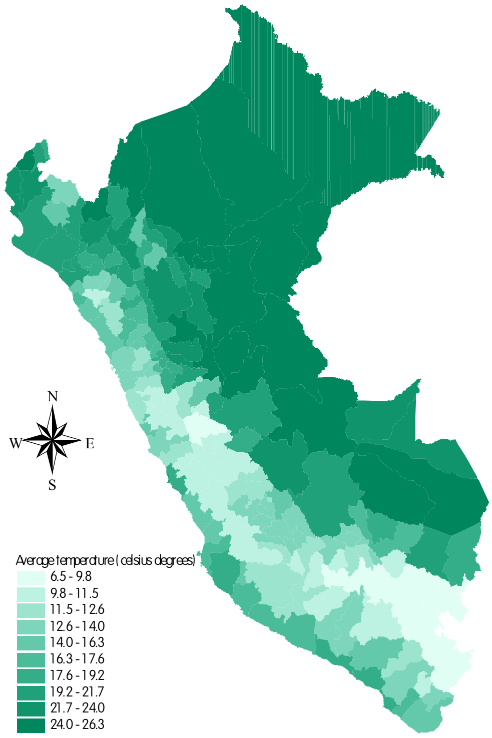
lack of rainfall in this region (see Map 3). However, this cold humid desert results in pleasant living conditions for those acclimatized to the local environment. Many separate ranges, surrounding several areas of high plateau, make up the Andes which account for 31 per cent of Peruvian territory. Passes through these mountains are usually high and difficult, especially, the southern Andes, which can be considered a barrier to trade and transportation. Climatic conditions also make vast areas of the Peruvian Andes relatively inhospitable (see Maps 3 and 4).

Map 3: Precipitation



Source: Authors.

Map 4: Temperature



Source: Authors.

A large part of Peruvian territory (about 58 per cent) lies in the Amazon Basin, most of this area is covered by dense forest that has slowed the development of the region. In some of these areas annual floods raise the water level more than 15 meters (50 feet) and inundate thousands of square miles of land. These floods deposit alluvial silts that renew the soils of the flooded areas (see Map 3). The distribution patterns of vegetation and soils in Peru are closely related to the distribution patterns of landforms and climate. Tropical forest types of vegetation and soils are found mainly in the Amazon Basin, while desert types are located mainly along the coast. Soils in most tropical forests are poorly developed and low in fertility except in areas subject to annual flooding. Peru is also well known for its mineral reserves. It has the world's second largest silver reserves, third largest of tin, fourth largest of lead, seventh largest of copper and eighth largest of gold. As can be seen in Map 2, a large proportion of Peru's mineral surface composition is sedimentary rock where petroleum deposits are usually found. Gold, silver, copper deposits are to be found in igneous and metamorphic rock.

Table 2: Eight natural regions of Peru

REGIONS	DESCRIPTIONS
Costa or Chala (coast or plain)	Territory below 500 m.o.s.l. on occidental side of Andes. Mainly desertic
Yunga (warm zone)	Both sides of Andean mountain range, located between 500 and 2,300 m.o.s.l. (occidental side) and 1,000 and 2,300 m.o.s.l. (oriental side). Typically formed by valleys
Quechua (temperate zone)	Both sides of the Andean mountain range, located between 2,300 and 3,500 m.o.s.l. Typically formed by knolls and steep hillsides
Suni or Jalca (cold lands)	Both sides of the Andean mountain range, located between 3,500 and 4,000 m.o.s.l. Typically formed by steep hills
Puna (high altitude plateau)	Both sides of the Andean mountain range, located between 4,000 and 4,800 m.o.s.l. Below snow mountains
Janca or Cordillera	At top of Andean mountain range, located between 4,800 and 6,768 m.o.s.l. Not a continuous area (only 1 district capital of 1879 districts in Peru is located at an altitude higher than 4,800 m.o.s.l.)
Selva Alta (high altitude jungle)	On oriental side of Andean mountain range, between 400 and 1,000 m.o.s.l. Mountainous forest with valleys
Selva Baja (low altitude jungle)	On oriental side of Andes, below 400 m.o.s.l.

Note: M.o.s.l. = metres above sea level.
Source: Pulgar Vidal (1946) and Peñaherrera (1986).

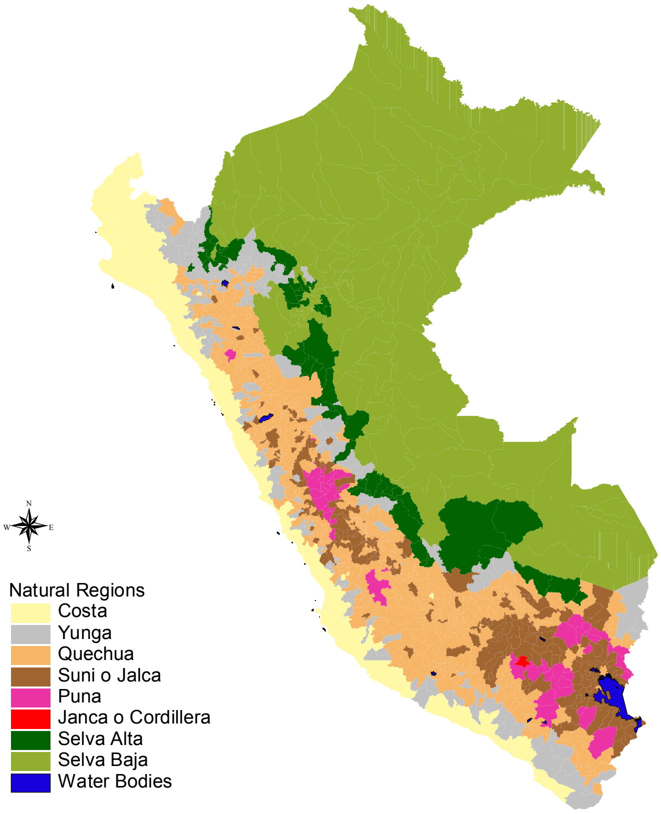
Peru has a long tradition of geographic analysis and its links with development. Initially, following the Spanish tradition, the country was classified into three distinct zones: the costa (coast or plains), the sierra (basically the Andean mountain range) and the selva (the jungle or Amazon). However, many authors[4] have shown that this classification scheme is not sufficient to encompass Peru's geographic diversity. Peru's geographic heterogeneity is quite high and landscapes can differ widely. Based on these findings, Pulgar Vidal (1946) divided Peruvian territory into eight distinct 'natural regions' (see Table 2). The geographical pattern of these zones is depicted in Map 5.

Despite the fact that there have been many efforts to link Peruvian geographical diversity to key issues as important as settlement location or construction of administrative or political regions, very little has been done to analyse the links between this geographic diversity and development, economic growth or poverty. The only exception is the government construction of 'poverty maps' to help target social programmes. One of the

[4] A literature review on this topic can be found in Pulgar Vidal (1946) and in Peñaherrera (1986:115-34).

most recent efforts in this regard is the construction of poverty indices at the provincial and district level by FONCODES (the public agency in charge of poverty alleviation programmes). Although these maps are geographic in nature, no effort has been made to link them to geographic variables, trying, for example, to find out whether there is any kind of poverty trap due to the negative externalities of certain 'geographic endowments'.

Map 5: Eight natural regions of Peru



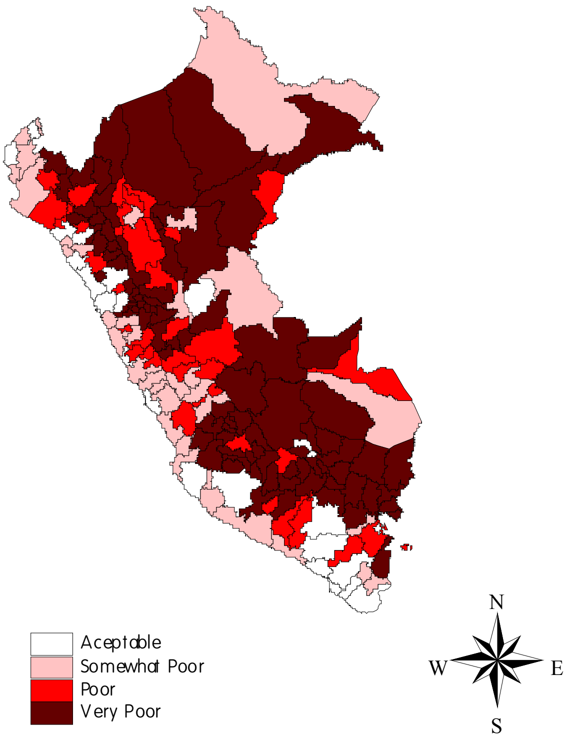
Source: Authors.

The next question to ask is, then, whether there is geographic concentration of poverty in Peru. Map 6 graphically answers this question by showing the poverty indices at the provincial level based on a 'poverty index' constructed by FONCODES.[5] As shown in the map there are huge welfare disparities across the country, and there is a heavy concentration of very poor people along the most geographically adverse regions, as in the sierra and selva. Table 3 also shows how there is a negative relation between the main

[5] This index was constructed at the district level by weighting socioeconomic indicators reflecting: extreme poverty (infant mortality, child malnutrition), indicators of education (illiteracy rate, school attendance rate), labour market indicators (proportion of working children, percentage of illiterate adults), housing indicators (percentage of overcrowded households, percentage of houses with inadequate roofing), and basic services indicators (access provided by public networks to water, sanitation and electricity); see FONCODES (1995)

geographic variables (altitude, rainfall, and temperature) and household economic welfare. The higher the altitude the larger the number of poor households in the specific region (districts). As expected, temperature shows a nonlinear relationship such that poverty increases in areas with extreme temperature levels (high and low). The precipitation variable however, does not display a clear relationship. These welfare disparities can also be attributed, at least in part, to a significant dispersion of asset ownership or access. As can be see in Tables 4 and 5, most of the access to public assets and services is at least 2 or 3 times as high in urban areas as compared to rural areas. In the case of access to sanitation connection, differences are even greater.[6]

Map 6: Poverty indices at the provincial level in Peru



Source: Authors.

Even though access to public goods and services has increased dramatically in rural areas during the last four years, new access continues to be biased in favour of urban areas. Two-thirds of the new electricity, sanitation and health services are placed in urban areas. Only in education does the pattern of new public goods placed in rural areas surpass that of urban areas.

[6] Poverty maps provide a detailed description of the spatial distribution of poverty within the country and are a crucial tool for research in trying to explain the relationship between poverty or inequality and indicators of development. On the other hand, it is important to mention that they must be interpreted within their limitations given that their quality is limited by is the sparseness of the desegregated data. Some improvements on these methodologies can be found in Hentschel *et al.* (1998).

Table 3: Geography and economic welfare (% of poor households)

	1985	1994	1997
Altitude (m.o.s.l)			
0-500	41.4	37.5	46.1
500-1000	43.5	38.2	48.6
1000-2300	51.9	37.0	53.8
2300-3500	57.7	43.7	59.7
> 3500	52.1	62.5	63.3
Precipitation (mm per year)			
0-100	35.3	33.2	40.7
100-200	54.0	33.4	42.8
200-400	46.0	65.3	58.7
400-600	59.4	69.8	61.9
600-1000	51.5	49.2	63.1
1000-1400	67.0	42.8	59.4
1400-2000	63.4	43.4	58.4
2000-2800	60.3	70.4	55.8
> 2800	42.7	34.4	54.7
Temperature (Celsius degrees)			
0-5	52.7	67.6	65.4
5-10	49.1	44.2	57.8
10-15	40.6	34.4	43.1
15-20	55.1	43.0	53.1
> 20	61.7	46.8	55.9

Source: Authors' calculations based on LSMS 1985/6, 1994 and 1997.

Note: Poverty line obtained from Escobal *et al.* (1998).

Table 4: Regional differences in access to services and assets, 1997

	Urban	Rural	Ratio
Family size	6.1	6.3	1.0
Years of education (head)	8.6	4.5	1.9
Years of education (adults)	8.1	5.0	1.6
Drop-out rates, secondary school	12%	15%	0.8
Access to electricity	97%	30%	3.2
Access to water, public network	89%	43%	2.1
Access to sanitation	84%	12%	7.3
Access to credit	37%	23%	1.6
Memo: poverty rate	40%	65%	

Source: LSMS (1997).

Table 5: Distribution of new access to basic and social services, 1994-97

	Urban	Rural	Ratio
Water, public network (%)	57	43	1.3
Electricity (%)	72	28	2.6
Sanitation connection (%)	78	22	3.5
Outpatient health (%)	74	26	2.8
School enrolment (%)	33	67	0.5

Source: LSMS (1994, 1997)

Given the above evidence, the major questions this research will try to answer are: what causal role do geographic variables, both natural and man-made, play in explaining per capita expenditure differentials across regions within Peru? How have these influences changed over time, how important will they be in the future, through what channels have those influences been transmitted and does access to private and public assets play a crucial role in reducing the negative effects of an adverse geography? The next section describes how we plan to formally answer these questions.

3 Analytical framework to test the effects of geography

The main question this paper tries to answer is whether geography has any effect on living standards after controlling for observable non-geographic characteristics of the households and whether access to public and private assets compensates for an adverse geography. To address this question, we have divided the analysis into three stages.

The first stage analyses the evidence of regional income differences and to what extent these differences had been hampered (or facilitated) by local or neighbouring, natural or man-made, geographic endowments. We analyse the evolution of geographic patterns and the importance of clustering in some areas by using spatial econometric techniques, such as the Moran I statistic (Moran 1950; see Annex 1). We measure for the presence, over time, of spatial concentration of per capita expenditure and geographical, private and public assets and test for their significance. In the second stage, to formally answer whether geography has a causal role in determining how household welfare evolves over time, we developed an estimable micro model of consumption levels and growth.

To model changes in consumption over time we use three census databases at the province level (see Annex 2 for details on how consumption is estimated for the census databases). This analysis also allows us to see what geographic factors matter to growth prospects at the micro level (Jalan and Ravallion 1998; Engerman and Sokoloff 1997). Our explanatory variables include a set of individual characteristics such as human assets (x), a set of private assets (z), a set of public assets at the district level (r) and a set of variables comprising specific geographic characteristics such as climate, soil characteristics and altitude (g). Specifically the change in consumption equation is:

$$\Delta c_p = a + \beta x_{p,0} + \phi z_{p,0} + \gamma r_{p,0} + \varphi g_p + \varepsilon_p \qquad (1)$$

in which the subscript p refers to province level averages of the respective variables, and the subscript zero refers to information of the initial period. We include each of the groups of regressors incrementally and lastly we estimate the full model. We run a set of models including, one by one, each of the groups of explanatory variables: geography (g), neighbouring public assets (r), private assets (z), and individual characteristics (x) and identify the direct externality effects of the presence of each of them. Additionally, according to the hypothesis of the presence of spatial concentration we analyse the importance of neighbouring province effects by measuring the significance of spatial autocorrelation[7] in each of our specifications and test how it decreases as we include additional groups of regressors (see Annex 1 for the spatial autocorrelation tests used).

We model the spatial dependence as a nuisance (a nuisance since it only pertains to the errors). Formally, this dependence is expressed by means of a spatial process for the error terms, either of an autoregressive or a moving average form (see Anselin 1988; and Anselin *et al.* 1996). Such an autoregressive process can be expressed as:

$$\begin{aligned} \Delta c_p &= \alpha + \beta x_{p,0} + \phi z_{p,o} + \gamma r_{p,0} + \varphi g_p + \varepsilon_p \\ \varepsilon_p &= \lambda W \varepsilon_p + \xi \end{aligned} \qquad (2)$$

with $W\varepsilon$[8] as a spatially lagged error term, λ as the autoregressive coefficient and ξ as a well-behaved (i.e. homoskedastic uncorrelated) error term. As a consequence of the spatial dependence, the error term no longer has the usual diagonal variance matrix but instead takes the following form (Anselin 1988):

$$E[\varepsilon\varepsilon'] = \Omega = \sigma^2[(I - \lambda W)'(I - \lambda W)]^{-1} \qquad (3)$$

Therefore, OLS estimates are no longer efficient but they are still unbiased. Furthermore, given that the lambda coefficient is unknown, the regression coefficients cannot be estimated using generalized least squares (GLS) and, therefore, in our last specification we

---

7 Spatial autocorrelation, or more generally, spatial dependence, is the situation where the dependent variable or error term at each location is correlated with observations on the dependent variable or values for the error term at other locations.

8 For N districts observed, W_i is the ith row of an (N*N) matrix W that assigns neighbouring districts to each district . The W used can be characterized by $W=\{w_{ij}\}$ such that $w_{ij}=1$ if i and j are neighbouring districts, $w_{ij}=0$ otherwise, and $w_{ii}=0$ for all i. The rows of W are then normalized such that each observation's neighbouring districts have the same amount of influence, that is $\sum_j W_{ij} = 1,$ for all i. In addition it will be assumed that each neighbouring district of a given district carries equal weight, $w_{ij=} w_{ik}$ for non-zero elements (neighbours) k and j for firm i. If more information were available about the amount of influence each district yields, this could be incorporated into the W matrix (regarding the different possible structures see Anselin 1988).

estimate the lambda coefficient jointly with the regression coefficients using full maximum likelihood estimation techniques.[9]

Lastly, the results of the previous specifications are used to break down the geographic effects into their component elements. For this purpose, we compute the expected gain (or loss) in consumption from living in one geographic region (coastal, for example) against living in another geographic region (mountainous, for example) specifying how much of the gain is explained by geographical variables, location (urban/rural), infrastructure and private assets:

$$(\overline{X}_M - \overline{X}_C)\hat{\beta} \qquad (4)$$

where $\overline{X}_{M,C}$ are the sample means for mountain and coastal regions for example, and $\hat{\beta}$ is the parameter of the respective variables under analysis (i.e. geographical, location, infrastructure and private assets). This breakdown represents the differential impact on a household's standard of all non-excluded variables in the two regions.

4 The data

To be able to answer the major questions outlined in the previous section we have developed three different databases—census and household surveys (LSMS) all of which were linked to a geographical database (see data sources).

Using the population and household census of 1972, 1981 and 1993 to construct a set of variables allows us to analyse the kind of changes that have emerged in the geographical pattern of Peru's most important socioeconomic variables during the last three decades. Additionally, using the methodology of Hentschel *et al.* (1998), we estimate a household-level expenditure equation using the information from the 1985/6 and 1994 LSMS surveys which allowed us to model the determinants of per capita expenditure growth at the provincial level. This, in turn, allows us to determine what role geographic variables play in explaining per capita expenditure differentials across regions in Peru.[10] This interpolation method is basically a nonlinear interpolation, and as such may be subject to the criticism that it may show more the structure of the model rather than the structure of the data. Although we fully acknowledge this, it is worthwhile mentioning that a similar model was tested against a different dataset to check whether or not the results were reasonable. The results, available in Escobal *et al.* (2001), show that when this model was fitted to the fourth quarter official household survey of 1998 (ENAHO) the results were

9 For a more extensive technical discussion of the relative merits of the various estimators suggested in the literature, see Anselin (1988, 1990).

10 For an example, see Borjas (1995) on effects of neighbourhood on schooling and wages in the US and Ravallion and Wodon (1997) on effects of geography on the level of poverty in Bangladesh, as well on the importance of public and private assets in explaining regional poverty variations.

quite good, as we were able to validate the model comparing the poverty rates it predicted with that of seven regions from which that survey was statistically representative. It should also be mentioned that in all of the seven regions, the interpolation allowed that more than 80 per cent of the household were correctly assigned as been poor or not poor.

Therefore, to estimate per capita expenditure at provincial levels for census years we regress per capita expenditure on private and public assets, allowing interactions between them.[11] Table A.2.1 in the Appendix shows the results of this procedure. The endogenous variable in each equation was the per capita expenditure in constant nuevos soles of 1994. From the coefficients obtained in Table A.2.1, we simulated the province-level per capita expenditure using the province-level variables obtained from the census data, and the means of the household surveys whenever there was not a counterpart variable in the census. For 1972 and 1981 we used the parameters of LSMS 1985/6 and for 1993 the calculations of LSMS 1994 due to the proximity of the sample surveys and census dates.

The province-level variables used in all census years were household size, percentage of houses without access to potable water, without drainage, without electricity, total illiteracy rate, schooling attendance rate, percentage of child labourers and percentage of the population living in urban areas. Additionally, for 1993 we included the percentage of the non-professional, economically active population, percentage of households headed by women, and college attendance rate. We complete the set of variables (to estimate province-level expenditure) using sample average values of the LSMS by regions. As we mention above, LSMS are divided in geographical regions to improve the quality of the sampling. These regions were included in the regression as dummy variables associated with location: northern coast, central highland, and Greater Lima, for example.

Per capita expenditure at the provincial level in each census year was adjusted to reproduce the aggregate consumption growth rate of national accounts within those years. Using 1981 as an anchor, we changed slightly the intercept coefficients of the other regressions to re-estimate the projected variables. Thus, we replace the OLS estimated coefficients 6.690 with 6.350, and 7.695 with 7.595, for 1993 and 1972 respectively. In this way the growth rate of the projected per capita expenditure (weighted by population in each year) is equal to the macroeconomic statistics. The coefficients reported in Table A.2.1 display the new values for the intercepts.

Finally, the number of provinces has not remained constant in the last 30 years. In 1972 the number of provinces was 150, in 1981 it was 153, and 188 in 1993, therefore we had to homogenize the province areas and shapes through time. With this purpose we decided to use the political-administrative division of Peru in 1993 because the Geographical Information System (GIS) was developed following the 1993 census. To impute the values in 1972 for new provinces we repeated the 'original' province information in each of its

[11] A more detailed discussion of these estimations can be found in Escobal *et al.* (1998).

new regions or areas. For 1981 we had district-level data and since the creation of a new province is basically a new clustering of districts we aggregate those district values to create data for the new provinces.

5 Empirical results

5.1 Peru's geography and its regional differences in expenditure

In this section we analyse the kind of changes that have emerged in the geographical pattern of Peru's most important socioeconomic variables during the last three decades. In addition we analyse changes in expenditure estimates, at the province level, between three census years of 1972, 1981 and 1993. We analyse 24 variables at the provincial level for a panel of the three given census years, as well as 160 additional variables at the provincial level and 88 additional variables at the district level for variables that were available only for 1993 and beyond. The data section described these variables as well as the databases that generate them.

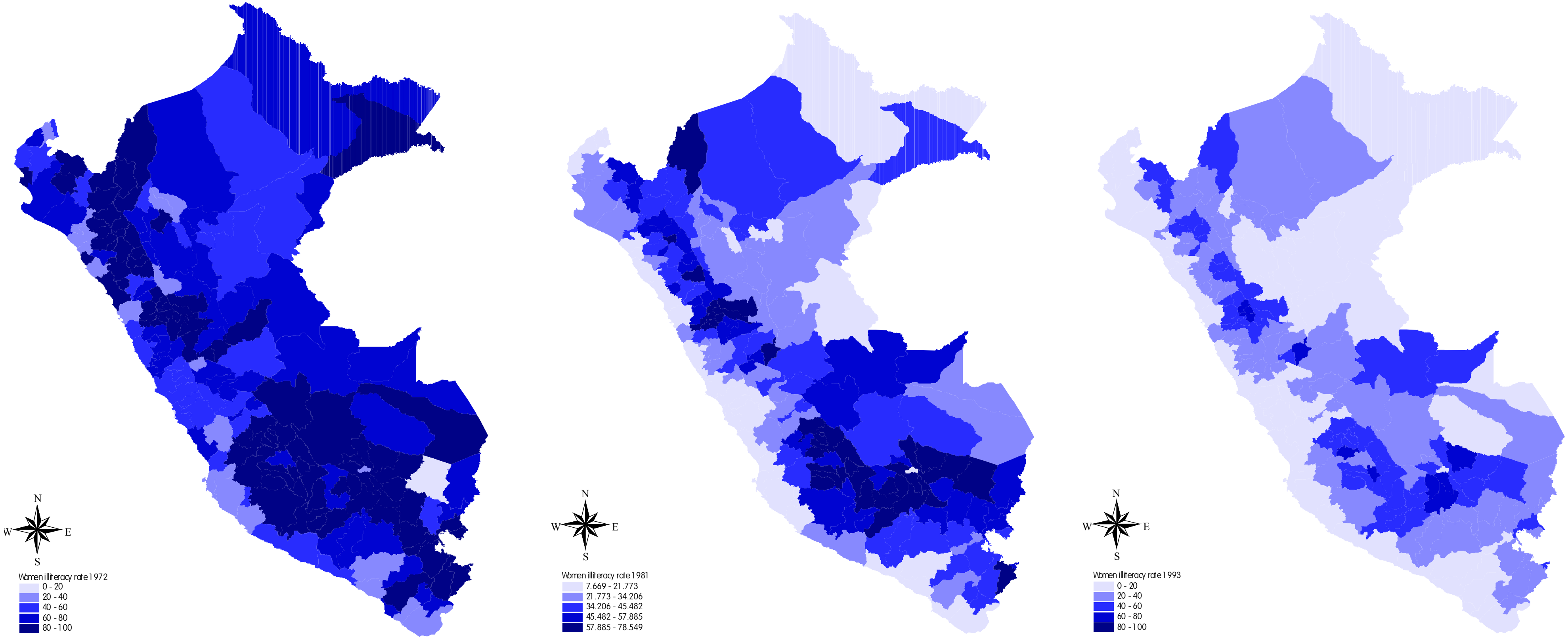
We have included in this section some of the maps generated with these variables. It is interesting to note that there are several types of evolution in the geographic patterns. There are cases such as the one depicted in Map 7 that show a dramatic reduction of illiteracy rates among women but, at the same time, the high rates are clustered in some areas (like the southern sierra and other high altitude zones). This kind of pattern can also be found in other key socioeconomic variables, such as total illiteracy rate or household size.

There are other variables, such as percentage of households without access to potable water, percentage of households without access to sanitation services, or percentage of households without access to electricity that display during the 1972-81 period a significant reduction in the coastal areas and afterward some clustering of high values specially in the southern sierra and high altitude jungle regions and no distinguishable pattern in the rest of the country (access to potable water is depicted in Map 8).

In order to more comprehensively analyse the changes that occurred in these geographic patterns we have constructed a per capita expenditure variable at the provincial level. Following a procedure similar to that of Hentschel *et al.* (1998), we used household data to construct expenditures functions using the Peruvian LSMS surveys of 1985 and 1994. We used the 1985 expenditure function to construct provincial-level expenditure estimates using data taken from the 1972 and 1981 census as explanatory variables. We used the 1994 expenditure function to construct the provincial-level expenditure estimates based on data taken from the 1993 census (see data section for details).

14

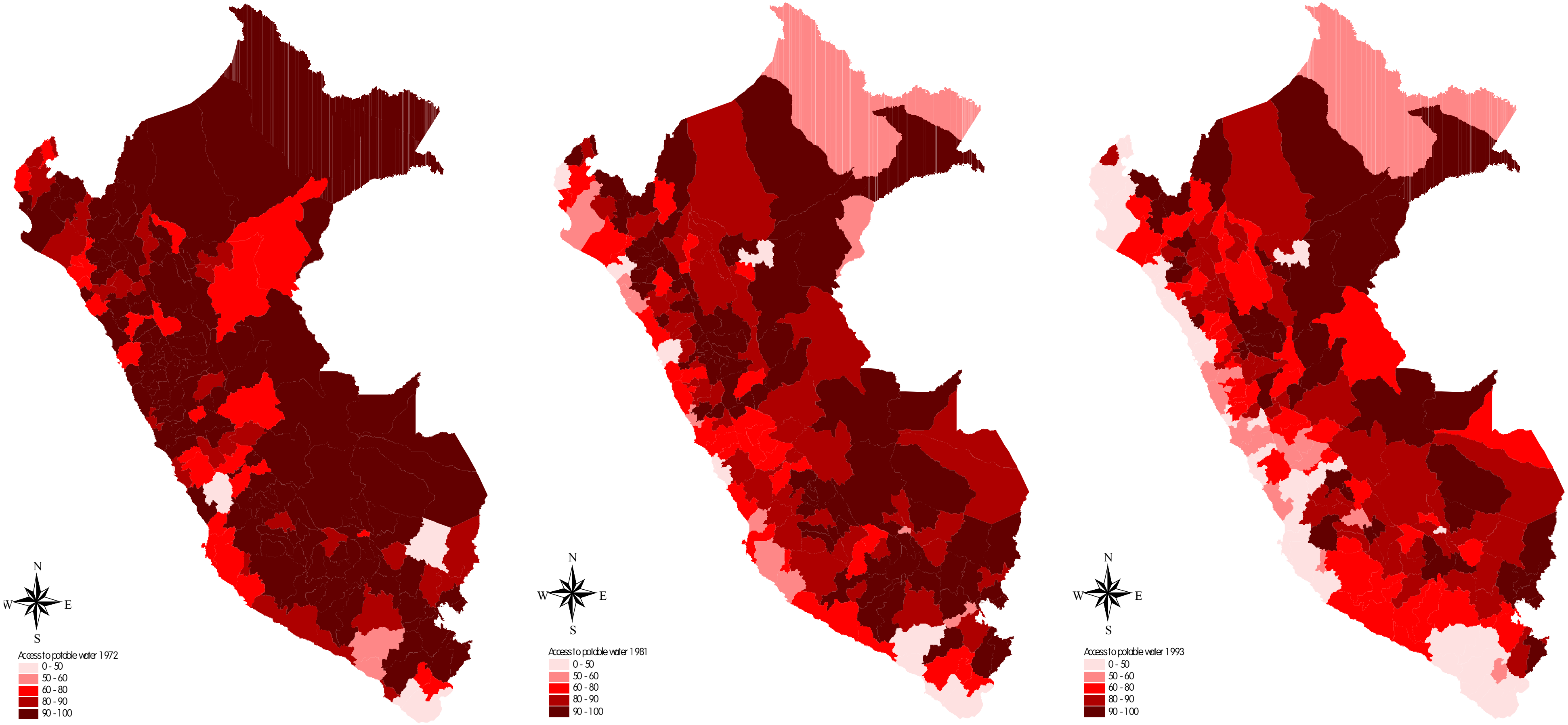
Map 7: Illiteracy rate of women



1972 | 1981 | 1993

Source: Authors.

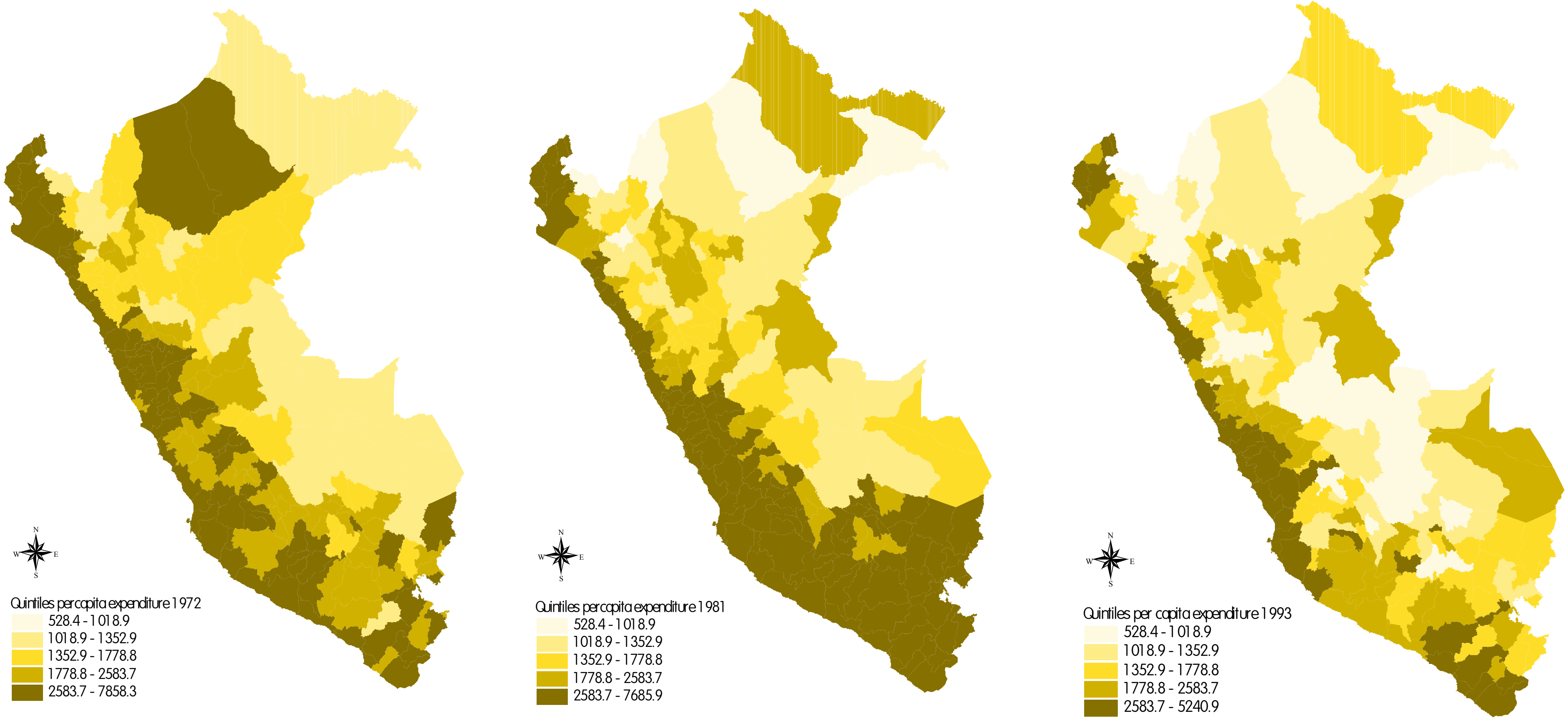
Map 8: Households without access to potable water



Source: Authors.

1972 1981 1993

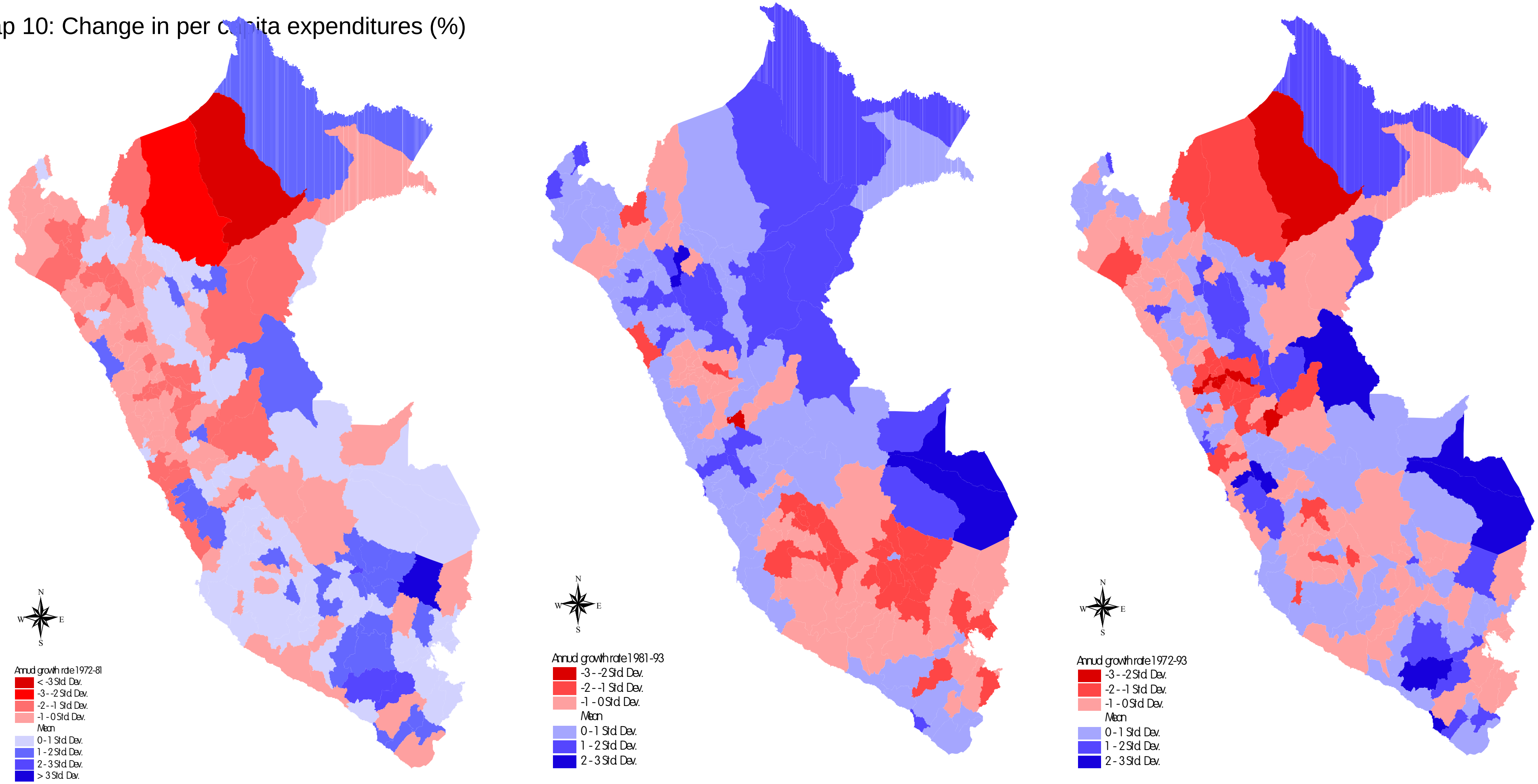
Map 9: Distribution of per capita expenditure



1972 1981 1993

Source: Authors.

Map 10: Change in per capita expenditures (%)



Source: Authors.

1972-81 1981-93 1972-93

The geographical evolution of Peru's per capita expenditures between 1972 and 1993 can be viewed in Map 9. Here it is evident that higher per capita expenditure is to be found along low altitude coastal regions. This pattern, which is already clear using 1972 data, is even more apparent as time passes. It is interesting to note that the Gini coefficients are extremely low (0.118 in 1972, 0.088 in 1981 and 0.187 in 1993). It must be noted however that interregional expenditure variance is very low, at least when compared to intraregion variance, making these Ginis perfectly consistent with a national Gini coefficient of 0.42 and 0.38 in 1985 and 1994, respectively.

Map 10 shows the pattern of distribution of interannual per capita expenditure growth rates between census years. Here it can be noted that the provinces whose per capita expenditures have grown faster tend to be clustered, as do those provinces showing little or even negative growth. Provinces showing high growth tend to be clustered in the higher altitude jungle. Table 6 confirms the graphical analysis showing high and statistically significant Moran Index and Geary Index values for all three census years. In addition, high Moran and Geary Index values can also be found for per capita expenditure growth.

Table 6: Spatial autocorrelation of province-level expenditure variables

Variables	Moran Index	Prob.[1]	Geary Index	Prob.[1]
Per capita expenditure				
1972	0.4131	0.00	0.6078	0.00
1981	0.5709	0.00	0.3993	0.00
1993	0.4888	0.00	0.4565	0.00
Change in per capita expenditure				
1972-81	0.3708	0.00	0.6186	0.00
1981-93	0.4990	0.00	0.4616	0.00
1972-93	0.2427	0.00	0.7308	0.00

Note: [1]Probability to reject null hypothesis (absence of spatial autocorrelation).

Source: Authors' calculations based on province estimates.

Table 7 shows some of the most significant spatially autocorrelated variables in our dataset. Using the Moran and Geary Indexes we find, aside from some obviously spatially correlated variables such as annual precipitation or altitude of the province or district capital, critical socioeconomic variables such as household size, percentage of households headed by women, or total and female illiteracy rates to be heavily clustered, high values in high altitude zones and low values in coastal areas. A similar situation can be found in other variables such as percentage of houses with inadequate floors or overcrowded housing, malnutrition rates and school drop-out rates and schooling years.

One summary welfare variable, per capita expenditure, for 1993 displays high and statistically significant Moran and Geary Index values. It is also interesting to note that the variable of soil depth, constructed to show agricultural land potential, also has a highly spatial autocorrelated pattern. Aside from some obvious variables, such as those related to urban areas (urban density or number of towns per province, for example) there are very few variables that do not show a clear geographical pattern. Only three variables deserve some mention—change in household size between 1972 and 1981; the growth of the illiteracy rate between 1981 and 1993; and the growth in per capita expenditures between 1972 and 1981—which do not show any geographical pattern measured by the Moran spatial autocorrelation index or the Geary Index. (see Appendix 3).

Table 7: High spatial autocorrelated variable

Variables	Moran Index	Z-Value	Geary Index	Z-Value
South latitude	0.9302	20.21 *	0.057	-18.76 *
North longitude	0.8870	19.27 *	0.093	-18.04 *
Precipitation	0.7573	16.47 *	0.259	-14.73 *
Household size 1993	0.7495	16.30 *	0.241	-15.10 *
Temperature (average)	0.7486	16.29 *	0.256	-14.79 *
Temperature (min.)	0.7469	16.25 *	0.255	-14.83 *
Temperature (max.)	0.7422	16.15 *	0.265	-14.62 *
Altitude of the district capital (meters above sea level)	0.6693	14.57 *	0.322	-13.47 *
% female-headed households 1993	0.6560	14.28 *	0.325	-13.43 *
Inadequate flooring	0.6518	14.19 *	0.339	-13.16 *
Soil depth	0.6422	13.99 *	0.328	-13.37 *
Total illiteracy rate 1981	0.6352	13.83 *	0.356	-12.82 *
Overcrowded houses 1993	0.6286	13.69 *	0.339	-13.15 *
Household size 1981	0.6130	13.35 *	0.377	-12.39 *
Per capita expenditure in 1981	0.6084	13.26 *	0.399	-11.95 *
Perimeter of the province	0.6032	13.14 *	0.390	-12.12 *

Note: *=p<0.01 where p is the probability to reject null hypothesis (absence of spatial autocorrelation).
Source: Authors' calculations based on National Population Census of 1972, 1981 and 1993.

5.2 Testing the causal role of geography on the evolution of welfare

As we read in Section 3, it is possible to derive a connection between the asset endowment of an individual household and its expenditure level. Following the same reasoning we can derive a connection between the level of private and public assets that can be found at some level of spatial aggregation (here the provincial level) and the per capita expenditure level that can be found in that area.

Table 8 shows the econometric results of what could be called the determinants of per capita expenditure growth at the provincial level. To reduce any possible endogeneity bias in explaining 1972-93 per capita expenditure growth rates we have chosen initial asset endowments as independent right hand side variables. To this basic dataset we have added several key geographical variables to check whether they can provide some explanation of

Table 8: Determinants of per capita expenditure growth rate, 1972-93 (OLS estimations with robust standard errors, at province level

Variables	Models				
at initial period	(1)	(2)	(3)	(4)	(5)
Intercept	4.8269 *	4.6892 *	4.3913 *	-0.0277	-0.3270
	(1.631)	(1.563)	(1.585)	(1.385)	(1.706)
Altitude	-1.1081 *	-0.7872 ~	-0.5096	0.2616	0.4580
	(0.385)	(0.377)	(0.447)	(0.385)	(0.389)
Latitude	-0.0226	-0.0308	-0.0288	-0.0231	-0.0170
	(0.017)	(0.017)	(0.017)	(0.019)	(0.019)
Longitude	-0.0561 *	-0.057 *	-0.0543 *	-0.0182	-0.0171
	(0.018)	(0.017)	(0.018)	(0.015)	(0.015)
Soil slope	-0.0012	0.0016	0.0021	0.0033	0.0035
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Soil depth	-0.003	-0.0017	-0.0018	0.002	0.0023
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)
Igneous rock	-0.2143	-0.2944 ~	-0.3102 *	-0.3197 *	-0.2757 *
	(0.126)	(0.123)	(0.123)	(0.100)	(0.106)
Metamorphic rock	0.0732	0.0536	0.0863	-0.1318	-0.1362
	(0.149)	(0.145)	(0.146)	(0.122)	(0.122)
Temperature	-0.0191	-0.0045	-0.0043	-0.0114	-0.0082
	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)
Basic needs		-0.0561 *	-0.0393 ~	-0.0222	-0.0225
		(0.013)	(0.020)	(0.017)	(0.016)
High x basic needs			-0.1110	0.0045	-0.0149
			(0.097)	(0.090)	(0.080)
School attendance rate				0.0143 *	0.0144 *
				(0.003)	(0.003)
Households headed by women (%)				-0.0109 ~	-0.0134 ~
				(0.005)	(0.005)
Working children (%)				0.0533 *	0.0462 ~
				(0.020)	(0.018)
Household size				0.0783	0.1057
				(0.133)	(0.128)
Household size growth[1]				-0.2624	-0.2208
				(0.140)	(0.136)
Number of migrants				0.0171	0.0101
				(0.029)	(0.029)
Spatial autocorrelation					0.2305 ~
					(0.102)
Number of observations	190	190	190	190	190
Adjusted R-squared	0.122	0.195	0.197	0.486	0.526

Note: [1]Instrumental variables are shown in the appendix. Standard deviation in parenthesis; *=p<0.01, ~=p<0.5. Model 1: Geography. Model 2: Geography + infrastructure. Model 3: Geography + infrastructure + Geo x infrastructure. Model 4: Geography + infrastructure + Geo x infrastructure + private assets. Model 5: Geography + infrastructure + Geo x infrastructure + private assets, modelling first-order spatial error autocorrelation.

Source: Authors' calculations based on Population and Household Census 1972 and 1993.

causes of expenditure growth. Table 8 shows the Moran spatial autocorrelation index for the four different specifications that were evaluated: (1) only private assets; (2) private assets plus geography variables; (3) the previous variables plus public assets; and (4) all the variables plus changes in access to key public assets.

We have used the log difference of per capita expenditures as a dependent variable. The reason for this choice (as opposed to using percentage changes) is related to functional form issues. If there is any misspecification in the per capita expenditure equations (which have been estimated as semi-log functions) the log difference of per capita expenditures will clean the bias, provided that these variables have similar effects over the years.

As can be seen in Table 8, when geographic variables are included as the only explanatory variables, altitude and longitude prove to be highly significant in explaining expenditure growth. In particular, it can be shown that the higher altitude provinces tend to have slower expenditure growth rates. When we add the variable of basic needs,[12] which encompasses the absence of critical public infrastructure (sanitation, water, telephone and electricity) we can see that altitude remains significant but its negative impact diminishes considerably. This effect can be viewed as demonstrating the importance of public infrastructure to lower negative geographic externalities. It is important to note that when we add private assets (some of which are obviously correlated with public assets) the importance of geography almost vanishes. This effect can be seen in Map 11 (Annexe 5) where we have graphed the pattern of geographic residuals of each model. This initial finding will be followed up more rigorously in the next section.

It is interesting to note that despite the fact that this expenditure growth function has included all relevant geographic variables at hand, the residuals continue to show spatial autocorrelation. As can be seen in Table 9, although the Moran Index diminishes as we include explanatory variables it remains significant. This fact suggests that there may be non-geographic non-observables affecting the provincial expenditure pattern. This is consistent with Ravallion and Wodon (1997) when they show that sizeable geographic differences in living standards can persist even if we take into account the spatial concentration of households with readily observable non-geographic characteristics conducive to poverty. The last column in Table 8 shows the estimated parameter values corrected for spatial autocorrelation.[13] The results confirm that when public and private

[12] Those defined as poor according to the unmet or unsatisfied 'basic needs' approach are those who suffer severe deprivation in one or more aspects of material well-being, such as poor housing, poor health, inadequate education or unemployment. This approach has a long history among development practitioners and, although seldom mentioned in the literature, is been used by most statistical agencies around the world. As Hicks and Streeten (1979) and Streeten (1984) mention, this approach tends to highlight those living in chronic poverty as oppose as short-term poverty that may be captured by other measures based on income or expenditure.

[13] The likelihood ratio test for spatial error dependence for the equation in the last column in Table 7 has a value of 3.67 with 1 degree of freedom, which confirms that the estimation has been properly corrected for spatial autocorrelation. For alternative methods of correcting for spatial autocorrelation see Annex 1 and for the empirical results Annex 4.

assets, as well as household characteristics, are included in the regression the impact of geographic variables is dampened.

Table 9: Spatial autocorrelation of growth regression residuals, by model

Type of	Regression Model Residuals			
Association	1	2	3	4
Moran Index	0.1091	0.1005	0.0973	0.0816
Z-value	3.1226	2.9658	2.9357	2.7877
Probability	0.0018	0.0030	0.0033	0.0053

Note: Model 1: Geography. Model 2: Geography + infrastructure. Model 3: Geography + infrastructure + Geo x infrastructure. Model 4: Geography + infrastructure + Geo x infra + private assets.

Source: Authors' calculations based on Table 7.

Table 10: Spatial association of growth regression residuals, by model (number of provinces)

Type of	Regression Model Residuals			
Association	1	2	3	4
Positive association	111	102	102	100
	(72.1)	(65.3)	(63.2)	(63.2)
Large values in large value areas	49	48	52	52
	(40.5)	(34.2)	(34.2)	(33.7)
Small values in small value areas	62	54	50	48
	(31.6)	(31.1)	(28.9)	(29.5)
Negative association	79	86	88	90
	(27.9)	(34.7)	(36.8)	(36.8)
Large values in small value areas	38	43	44	45
	(11.6)	(17.9)	(17.4)	(17.4)
Small values in large value areas	41	43	44	45
	(16.3)	(16.8)	(19.5)	(19.5)
Total	190	188	190	190
	(100.0)	(100.0)	(100.0)	(100.0)

Note: Column percentages are shown parenthesis. Model 1: Geography. Model 2: Geography + infrastructure. Model 3: Geography + infrastructure + Geo x infra. Model 4: Geography + infrastructure + Geo x infra + private assets.

Source: Authors' calculations based on Table 7.

Finally, in Table 10 we can find the spatial breakdown of the regression model according to the Anselin (1995a) technique (see Annex 1). Here residuals are clustered in four groups: large residual values clustered around large value areas; small residual values clustered around small value areas; large residual values located around small value areas; and, finally, small residual values located around large value areas. The results confirm that geography and public asset access variables tend to lower spatial autocorrelation, and geography variables are the ones that (at the marginal level) account most for per capita growth patterns.

5.3 Breakdown of regional per capita expenditure

To disentangle the effect of geography on regional expenditure and expenditure growth we have applied the breakdown technique described in Section 3 to the province per capita expenditure growth equations reported in Table 8. For this breakdown we have assumed that parameters are stable across the three main geographic areas: coastal, highland and jungle. This initial breakdown is shown in Table 11. In this case, per capita growth rate differentials between highland and coastal regions and between jungle and coastal regions can be broken down into their main determinants: geographical differences, infrastructure differences and asset endowment differences.

Table 11: Decomposition of regional per capita expenditure differences (growth rate differences at province level)

Group of variables	Highland–Coast		Jungle–Coast
Geography	**0.2126**		**0.1296**
Altitude level	0.1182		0.0055
Latitude	-0.0280		0.0471
Longitude	0.0437		0.0396
Soil slope	0.0518		-0.0159
Soil depth	-0.0020		0.0379
Igneous rock	-0.0329	*	0.0222
Metamorphic rock	0.0300		0.0399
Temperature	0.0319		-0.0467
Infrastructure	**-0.0431**		**-0.0920**
Basic needs	-0.0431		-0.0920
Geography x Infrastructure	**-0.0125**		**-0.0041**
Altitude x Basic needs	-0.0125		-0.0041
Private assets	**-0.3430**	*	**-0.0031**
School attendance rate	-0.1335	*	-0.0663
Female household head (%)	-0.0739	~	0.0147
Working children (%)	0.0278	~	0.0090
Household size	-0.0689		0.0580
Household size growth [1]	-0.0881	+	-0.0133
Number of migrants	-0.0063		-0.0051
Total explained	-0.1860		0.0304
Residual	0.1048		0.0989
Total	**-0.0812**		**0.1293**

Note: [1]Instrument variables are shown in the appendix. *=p<0.01, ~=p<=0.05, +=p<=0.1.

Source: Authors' calculations based on Population and Household Census 1972 and 1993.

Here, as was the case with the previous result, geography does not appear to significantly contribute to growth differentials, once infrastructure differences and private asset endowment differences are accounted for. In other words, once the main geographic variables are accounted for (altitude, temperature and surface characteristics), only private assets are needed to explain regional expenditure differences. Similarly, the second column shows the breakdown of the differences in log per capita expenditure between the jungle area and the coast, showing again that once main geographic variables are accounted for most of the regional expenditure differences can be explained by private asset composition.[14]

Obviously, the fact that geography has no additional impact on regional per capita expenditure differences has to do with the fact that key infrastructure variables—such as school and medical facilities, access to electricity, water and sanitation, as well as private assets—have dampened the effect of geography on regional expenditure differentials. To see this, Table 12 performs the same breakdown exercise introducing each set of variables sequentially. First, geography variables are entered in the model alone and the breakdown exercise is conducted only with these variables. In this case, geography is highly significant in explaining per capita expenditure growth differentials between the highland and coastal areas, as well as between the jungle and coastal areas. Geography remains highly significant even after we introduce location variables and their cross-products into the analysis. However, once infrastructure variables come into play in the analysis, the impact of geography disappears, as the coefficients associated with these types of variables are shown to be jointly non-significant. This could be because, in the models without infrastructure, the geography variables were choosing their effect and therefore when improving our specification the effect of these variables disappears. It must be noted that the analysis remains valid even if we correct for possible spatial autocorrelation due to possible omitted non-geographic spatially correlated variables.

14 When we carry out a similar exercise using household surveys similar results are obtained, i.e. once the main geographic variables are accounted for only private assets and infrastructure endowments are needed to explain regional expenditure differences (see Escobal and Torero 2000). Although is important to mention that given we cannot correct for spatial autocorrelation in the household level regressions households could be sorted geographically on the basis of some latent characteristics e.g., earning ability, land quality, and therefore geographic characteristics could be correlated with these latent characteristics not allowing us to genuinely identify a causal effect.

Table 12: Decomposition of regional per capita growth expenditure differences, by model (at province level)

Group of variables	Highland-Coast 1	1+2	1+2+3	1+2+3+4	1+2+3+4[1]	Jungle-Coast 1	1+2	1+2+3	1+2+3+4	1+2+3+4[1]
(1) Geography	-0.163 ~	-0.113	-0.047	0.158	0.213	0.023 ~	0.154	0.136	0.126	0.130
(2) Infrastructure		-0.108 *	-0.075 ~	-0.043	-0.043		-0.229 *	-0.161 ~	-0.091	-0.092
(3) Geo x infrastructure			-0.093	0.004	-0.013			-0.031	0.001	-0.004
(4) Private assets				-0.327 *	-0.343 *				-0.025 *	-0.003 *
Explained	-0.163	-0.221	-0.215	-0.208	-0.186	0.023	-0.075	-0.056	0.012	0.030
Residual	0.082	0.139	0.134	0.127	0.105	0.106	0.205	0.185	0.118	0.099
Total	-0.081	-0.081	-0.081	-0.081	-0.081	0.129	0.129	0.129	0.129	0.129

Note: [1]Modelling first-order spatial error autocorrelation. *=p<0.01, ~=p<=0.05, +=p<=0.1.

Source: Authors' calculations based on 1972 and 1993 Population and Household Census.

25

6 Conclusions

Peru's enormous geographic diversity makes it an extremely interesting case study to analyse whether geography has a causal role in determining how household welfare evolves over time. We know that there are huge welfare disparities across Peru, and there is a heavy concentration of very poor people throughout the most geographically adverse regions, as in the sierra and selva. Although, these welfare disparities can be attributed to geography they can also be related, at least in part, to a significant dispersion in access to infrastructure and other public assets. Therefore, there is no clear evidence that regional income differences can only be explained by geography or if they had been hampered (or facilitated) by local or neighbouring natural or man-made geographical endowments.

Despite the fact that there have been many efforts to link Peru's geographical diversity to key issues as important as settlement location or construction of administrative or political regions, very little has been done to analyse the links between this geographic diversity and development, economic growth or poverty. To reduce this gap, our research strategy consisted of describing how geography might play a fundamental role in regional economic growth and what relationship there is between geographic variables and expenditure levels and growth across regions within Peru. To formally answer whether geography is a determinant of the evolution of welfare over time, we developed a micro model of consumption which not only took in the local effect of geographic variables but also included public and private assets as variables that could reduce the potentially adverse effect of geography. For this purpose we used national census data for 1972, 1981 and 1993, the LSMS surveys for 1991, 1994, 1996, 1997, information from the district-level infrastructure census; geographical datasets, and information from the III National Agrarian Census of 1994. This cross-sectional analysis helped us in attempting to understand whether geographic externalities arising from local or neighbouring public assets, or local endowments of private goods, entail that living in or near a well-endowed area implies that a poor household can eventually escape poverty.

We have shown that what seem to be sizable geographic differences in living standards in Peru can be almost fully explained when one takes into account the spatial concentration of households with readily observable non-geographic characteristics, in particular public and private assets. In other words, the same observationally equivalent household has a similar expenditure level in one place as in another with different geographic characteristics such as altitude or temperature. This does not mean, however that geography is not important, but that its influence on expenditure levels and growth differentials comes about through a spatially uneven provision of public infrastructure. Furthermore, when we measure the expected gain (or loss) in consumption from living in a geographical region (i.e. coastal) as opposed to living in another geographic region (i.e. highlands), we found that most of the difference in log per capita expenditure between the highland and the coast can be accounted for by the differences in infrastructure endowments and private assets. This could be an indication that the availability of infrastructure could be limited by the geography and therefore the more adverse geographic regions are the ones with less access to public infrastructure.

Another interesting result is that despite the fact that in our models of expenditure growth we included all relevant geographic variables, as well as infrastructure and private assets variables, the residuals continue to show spatial autocorrelation. This fact suggests the idea that there may be non-geographic non-observables that could be affecting the provincial expenditure pattern. This is consistent with Ravallion and Wodon (1997) when they show that sizable geographic differences in living standards can persist even if we take into account the spatial concentration of households with readily observable non-geographic characteristics conducive to poverty.

It is important to note that there appear to be non-geographic, spatially correlated omitted variables that need to be taken into account in our expenditure growth model. Therefore policy programmes that use regional targeting do have a rationale even if geographic variables do not explain the bulk of the difference in regional growth, once we have taken into account differentials in access to private and public assets.

Lastly, an issue which we had not taken into account, and which could be very important for future research, is the fact that adverse geographic externalities can provide incentives to migration. This is something which we do not control for in this research. The migration effect could be twofold. On the one hand, it could be the reason why households with fewer private assets are the ones which choose to locate in the more adverse geographical regions. On the other hand, it could be very important for policy-making in developing infrastructure, in the sense that certain investments in infrastructure, such as education, are mobile with migration, while others are not. Therefore, it could be more profitable to invest in mobile infrastructure in the more adverse geographic regions, to give the individuals the necessary tools to migrate from these regions and therefore increase their probability of escaping a poverty trap.

Data Sources

At household level

Living Standard Measurement Surveys 1985-86 and 1994, Cuanto Institute.

At province level

- Population and Household Censuses 1972, 1981 and 1994, Instituto Nacional de Estadística e Informática: population and household characteristics.
- Third National Agrarian Census 1994, Instituto Nacional de Estadística e Informática: agricultural variables, cattle and land.
- Basic Needs Map 1994. Instituto Nacional de Estadística e Informática: basic needs and health variables
- Social Investment Map 1994, FONCODES: poverty index and its components, living standard.

Geographic variables

- Arc data online at: www.esri.com/data/online/esri/wothphysic.html. This information was afterwards overlaid on a map of Peru at provincial and district levels. The score for each province or district was selected according to the position of its centroid on the thematic map: earthquake zones, precipitation, soils and vegetation.
- Natural Resources in Peru 1995, Instituto Nacional de Recursos Naturales: bioclimate and land potential scores.
- Social Investment Map 1994, FONCODES: altitude and geographic location.

Annex 1

Measuring geographical association: theoretical framework

The importance of the spatial relationships began in the seventies with the works of Cliff and Ord (1972) in the United Kingdom, and Hordijk (1974), Hordijk and Pelinck (1976) and Hordijk and Nijkamp (1977). These studies created a great interest in the development of a methodology for the study of observations distributed in a specific geographical location and gave birth to what is called 'spatial econometrics'. Spatial autocorrelation says that what is observed in one place is in part determined by what is occurring in the other spatial locations. So, any observation of a variable *y* in *i* (where *i* is an element of a population *S*), is related formally through a function *f* to the magnitudes of the variable in other spatial units in the system.

$$yi = f\,(y_1, y_2, \ldots y_{i-1}, y_{i+1}, \ldots y_n) \qquad \text{(a1)}$$

There are a large number of tests to detect the presence of spatial correlation (Anselin 1988), but those that are most used are the 'Moran Statistic' *(I)* and the G statistics (Getis and Ord 1992).

The Moran Statistic

Formally, Moran's I is:

$$I = \frac{N}{S_o} * \frac{\sum_i \sum_j w_{ij}(x_i - \mu).(x_j - \mu)}{\sum_i (x_i - \mu)^2} \qquad \text{(a2)}$$

where N is the number of observations, x_i and x_j are observations for location i and j (with mean μ), w_{ij} is the element in the spatial weight matrix corresponding to the observation pair i, j. The W used here can be characterized $W=\{w_{ij}\}$ such that $w_{ij}=1$ if i and j are neighbours, $w_{ij}=0$ otherwise, and $w_{ii}=0$ for all i. The rows of W are then normalized such that each observation's neighbours have the same amount of influence, that is $\sum_j W_{ij}=1$, for all i. In addition it will be assumed that each neighbour of a given farm carries equal weight, $w_{ij=}\ w_{ik}$ for non zero elements (neighbours) k and j for farmer i. If more information were available about the amount of influence each household exercises, this could be incorporated into the W matrix (regarding the different structures see Anselin 1988). S_o is a scaling constant:

$$S_o = \sum_i \sum_j w_{ij} \qquad \text{(a3)}$$

i.e., the sum of all weights. For a row standardized spatial matrix, which is the preferred way to implement the test and the way it is done in this paper, S_o equals N (since each row sums to 1), and the statistic simplifies to the ratio of a spatial cross product to a variance:

$$I = \frac{\sum_i \sum_j w_{ij}(x_i - \mu).(x_j - \mu)}{\sum_i (x_i - \mu)^2} \qquad \text{(a4)}$$

Moran's I is similar but not equivalent to a correlation coefficient and is not centred around 0. In fact, the theoretical mean of Moran's I is -1/N-1. In other words, the expected value is negative and is only a function of the sample size (N). Note, however, that this mean will tend to zero as the sample size increases. Instead of using the I statistics by themselves, inference is typically based on a standardized z-value. This is computed by subtracting the theoretical mean and dividing the result by the theoretical standard deviation.

$$z_I = \frac{(I - E(I))}{SD(I)} \qquad \text{(a5)}$$

where $E(I)$ is the theoretical mean and $SD(I)$ is the theoretical standard deviation. For a technical discussion and detailed expressions for the moments see Cliff and Ord (1972, 1981). The most common approach is to assume that the variable in question follows a normal distribution. Based on asymptotic considerations (i.e. by assuming that the sample may became infinitely large) the z-value, using the proper measures for mean and standard deviation, follows a standard normal distribution (i.e. normal distribution with mean 0 and variance 1). Significance of the statistic can then be judged by comparing the computed z-value to its probability in a standard normal table (see Case 1992).

Deriving the G and G* statistic

The Getis and Ord (1992) statistic is used as a validation of the Moran I. Getis and Ord introduced a family of statistics, G, that can be used as measures of spatial association in a number of circumstances. Formally, the G statistic, for a chosen critical distance d, $G(d)$, is defined as:

$$G(d) = \frac{\sum_i \sum_j wij(d) x_i x_j}{\sum_i \sum_j x_i x_j} \qquad \text{(a6)}$$

where x_i is the value observed at location i, and $w_{ij}(d)$ stands for an element of the symmetric (non-standardized) spatial weights matrix for distance d. The numerator of the statistic is similar to that of Moran's I, but its denominator is different. Its significance is assessed by means of a standardized z-value, obtained in the usual fashion. The mean and variance of the G(d) statistic can be derived under a randomization assumption and the z-value can be shown to tend to a standard normal variable in the limit (see Getis and Ord 1992 for detailed derivations).

For each observation *i*, the G_i and G_i* statistics indicate the extent to which that location is surrounded by high values or low values for the variable under consideration, for a given distance d. Formally, the G_i and G_i* statistics are defined as:

$$G_i = \frac{\sum_j wij(d)x_j}{\sum_j x_j} \qquad \text{(a7)}$$

where $w_{ij}(d)$ are the elements from the contiguity matrix for distance d. The G_i and G_i* measures differ with respect to the number of observations that are included in the computation of the denominator. For G_i statistic, $j \neq i$ while for the G_i* statistic $j = i$ is included in the sum. In other words, the G_i* measure provides a measure of spatial clustering that includes the observation under consideration, while the G_i measure does not. Inference about the significance of the G_i and G_i* statistics is based on a standardized z-value, which is computed by substituting the theoretical mean and dividing by the theoretical standard deviation (for more details see Getis and Ord 1992).

A positive and significant z-value for a *G*, G_i or G_i* statistic indicates spatial clustering of high values, whereas a negative and significant z-value indicates spatial clustering of low values. Note that this interpretation is different from that of the more traditional measures of spatial autocorrelation, as the Moran I, where spatial clustering of like values, either high or low, are both indicated by positive autocorrelation.

Local indicators of spatial association (LISA)

In Anselin (1995), a local indicator of spatial association (LISA) is defined, and shows how they allow for the breakdown of the global indicators, such as Moran's I, into the contribution of each observation. The LISA statistics serve two purposes. On the one hand, they may be interpreted as indicators of local pockets of nonstationarity, or hot spots, similar to the G_i and G_i^* of the Getis and Ord (1992). On the other hand, they may be used to assess the influence of individual locations on the magnitude of the global statistic and to identify 'outliers,' as in Anselin's Moran scatterplot (1995). Both of these uses will help in determining which locations had the greatest correlation with their neighbours. The LISA for a variable y_i, observed at location *i*, can be expressed as a statistic L_i, such that:

$$L_i = f(y_i, y_{J_i}) \qquad \text{(a8)}$$

Where f is a function (possibly including additional parameters), and the y_{Ji} are the values observed in the neighbourhood J_i of i.

Similar to the rationale behind the significance tests for G_i and G_i^* statistics of Getis and Ord (1992), the general LISA can be used as the basis for a test on the null hypothesis of no local spatial association. However, in contrast to what holds for the G_i and G_i^* statistics, general results on the distribution of a generic LISA may be hard to obtain.

As a special case of the local Gamma,[15] a local Moran statistic for an observation I may be defined as:

$$I_i = z_i \sum_j w_{ij} z_j \qquad \text{(a9)}$$

where, analogous to the global Moran's I, the observations z_i, z_j are in deviations from the mean, and the summation over j is such that only neighbouring values j element of J_i are included. For ease of interpretation the weights w_{ij} may be in row-standardized form, though it is not necessary, and by convention, w_{ii}=0.

It can be easily seen that the corresponding global statistic is indeed the familiar Moran's I; the sum of the local Moran is:

$$\sum_i I_i = \sum_i z_i \sum_j w_{ij} z_j \qquad \text{(a10)}$$

The moments of I_i under the null hypothesis of no spatial association can be derived using the principles outlined by Cliff and Ord (1981:42-6) and a reasoning similar to the one by Getis and Ord (1992).

A test for significant local spatial association is based on these moments, although, as mentioned by Anselin (1995), the exact distribution of such a statistic is still unknown.

15 See Mantel (1967) and Luc Anselin (1995).

Annex 2: Data description

Table A2.1: Determinants of (log) per capita expenditure (OLS estimation with robust errors)

	Census year					
	1972[1]		1981[1]		1993[2]	
Variables	Coeff.	Std Dev.	Coeff.	Std Dev.	Coeff.	Std Dev.
Intercept	7.6959	(0.1954)	7.7777	(0.3271)	6.3502	(0.1377)
Access to credit	0.1384	(0.0399)	0.1351	(0.0364)	0.0826	(0.0366)
Access to drinking water	-0.1051	(0.0589)	-0.1316	(0.0535)		
Access to electricity	0.0846	(0.0541)	0.0788	(0.0497)	0.0021	(0.0004)
Access to in-house drainage services	0.1165	(0.1455)	0.1032	(0.1030)	0.0016	(0.0009)
Cattle	0.1288	(0.0827)	0.1368	(0.0800)	0.0913	(0.0788)
Durable goods	0.0680	(0.0092)	0.0681	(0.0087)	0.0051	(0.0046)
Fertilizers usage	0.1619	(0.0436)	0.1839	(0.0414)	0.1056	(0.0327)
Hh head gender	0.0278	(0.0627)	-0.0035	(0.0523)		
Hh members with secondary education (%)			0.0031	(0.0023)		
House with inadequate floor	-0.0042	(0.0009)	-0.0038	(0.0008)	-0.0021	(0.0003)
Hh size	-0.2760	(0.0341)	-0.3361	(0.0306)	-0.3253	(0.0283)
Illiteracy rate	-0.0017	(0.0008)	-0.0012	(0.0008)	-0.0016	(0.0007)
School attendance (children)	0.0010	(0.0006)	0.0006	(0.0006)		
Land size	0.0432	(0.0503)	0.0185	(0.0413)		
Number of migrants (household members)	-0.0061	(0.0410)	-0.0039	(0.0409)	0.1359	(0.0261)
Number of rooms in the house	0.0050	(0.0015)	0.0041	(0.0013)	0.0562	(0.0108)
Non-professional labor force			0.0002	(0.0028)		
Potential work experience	-0.0001	(0.0065)	0.0002	(0.0057)	0.0153	(0.0058)
Savings	0.0772	(0.0343)	0.0471	(0.0349)	0.0775	(0.0359)
Schooling attendance rate					0.0004	(0.0004)
Schooling years (hh head)	0.0167	(0.0119)	0.0168	(0.0114)	0.0310	(0.0073)
Schooling years (other members)	0.0372	(0.0188)	0.0388	(0.0160)	0.0326	(0.0070)
Seeds usage	0.1419	(0.0366)	0.1390	(0.0335)	0.0798	(0.0322)
Social networks	0.2282	(0.0601)	0.2197	(0.0620)	0.0862	(0.1102)
Spell of illness (hh head)	0.0153	(0.0299)	0.0268	(0.0299)	-0.0516	(0.0326)
Urban zone	0.0064	(0.0021)	0.0092	(0.0034)	0.0176	(0.1592)
Working children (%)	-0.0014	(0.0005)	-0.0013	(0.0005)		
Northern coast	-0.1374	(0.0334)	-0.1408	(0.0321)	-0.0460	(0.0257)
Central coast	-0.1991	(0.0375)	-0.2033	(0.0393)	-0.0304	(0.0332)
Southern coast	-0.0352	(0.0595)	-0.0552	(0.0642)	-0.0939	(0.0490)
Northern highlands	-0.5987	(0.0541)	-0.5789	(0.0508)	0.1185	(0.0358)
Central highlands	-0.3599	(0.0379)	-0.3670	(0.0374)	-0.0564	(0.0267)
Southern highlands	-0.7135	(0.0365)	-0.0413	(0.0356)	-0.0769	(0.0287)
Northern high altitude jungle	-0.4818	(0.0579)	-0.4313	(0.0583)	-0.2987	(0.0488)
Central high altitude jungle	-0.4875	(0.0547)	-0.4324	(0.0509)	-0.2745	(0.0501)

table continues…

Low altitude jungle					-0.2327	(0.0561)
Durable goods (squared)	-8.59E-04	(0.0003)	-8.07E-04	(0.0002)	-7.72E-06	(0.0000)
Hh size (squared)	0.0120	(0.0024)	0.0156	(0.0021)	0.0153	(0.0020)
Number of migrants (hh members) squared	0.0002	(0.0072)	-0.0019	(0.0073)		
potential work experience (squared)	1.07E-05	(0.0001)	-3.00E-05	(0.0001)	-1.63E-04	(0.0001)
Savings (squared)	0.0002	(0.0003)	0.0004	(0.0003)	-0.0015	(0.0007)
Schooling years (other members, squared)	-0.0020	(0.0022)	-0.0034	(0.0021)		
Spell of illness (hh head) squared					0.0002	(0.0063)
Durable goods x social networks	-0.0060	(0.0022)	-0.0035	(0.0021)	0.0007	(0.0037)
Hh size x potential work experience	0.0001	(0.0003)	0.0004	(0.0003)	0.0001	(0.0002)
Hh size x savings	-0.0065	(0.0033)	-0.0053	(0.0036)	-0.0032	(0.0017)
Hh size x spell of illness	0.0011	(0.0078)	0.0020	(0.0084)	0.0076	(0.0135)
Number of migrants x durable goods	-0.0002	(0.0005)	-0.0003	(0.0006)	0.0005	(0.0009)
Number of migrants x land size	0.0296	(0.0319)	0.0227	(0.0354)	0.0596	(0.0506)
Number of migrants x savings	0.0043	(0.0023)	0.0040	(0.0026)	-0.0004	(0.0030)
Potential work experience x durables goods	-0.0001	(0.0001)	-0.0001	(0.0001)	0.0000	(0.0001)
Potential work experience x number of migrants	-0.0003	(0.0006)	0.0001	(0.0006)	-0.0017	(0.0006)
Potential work experience x savings	-0.0005	(0.0004)	-0.0004	(0.0004)	0.0002	(0.0004)
Potential work experience x spells of illness	-0.0001	(0.0006)	-0.0003	(0.0006)	0.0007	(0.0006)
Savings x durable goods	-5.06E-05	(0.0002)	-2.19E-05	(0.0002)	-2.12E-04	(0.0001)
Schooling years (hh head) x durable goods	-0.0001	(0.0003)	-0.0003	(0.0003)	-0.0006	(0.0003)
Schooling years (hh head) x land size	-0.0113	(0.0120)	-0.0053	(0.0102)	0.0092	(0.0089)
Schooling years (hh head) x potential work experience	-0.0001	(0.0002)	0.0000	(0.0002)	-0.0002	(0.0002)
Schooling years (hh head) x potential work experience	0.0023	(0.0019)	0.0027	(0.0020)	-0.0067	(0.0016)
Schooling years (hh head) x savings	-0.0044	(0.0016)	-0.0044	(0.0017)	0.0003	(0.0013)
Schooling years (hh head) x spells of illness	-0.0026	(0.0023)	-0.0013	(0.0022)	0.0056	(0.0017)
Spell of illness x durable goods	0.0005	(0.0007)	0.0002	(0.0007)	-0.0001	(0.0006)
Spell of illness x number of migrants	-0.0024	(0.0044)	-0.0028	(0.0045)	-0.0014	(0.0057)
Spell of illness x savings	0.0042	(0.0024)	0.0024	(0.0026)	-0.0006	(0.0033)
Urban zone x hh head gender	-7.85E-05	(0.0007)	1.95E-04	(0.0006)		
Urban zone x land size	0.0007	(0.0013)	0.0001	(0.0012)		
Urban zone x savings (squared)	-6.82E-06	(0.0000)	-8.07E-06	(0.0000)	1.29E-03	(0.0006)
Urban zone x schooling years (hh head, squared)	7.18E-05	(0.0001)	4.79E-05	(0.0001)	6.57E-03	(0.0066)
Urban zone x schooling years (other member)	-0.0001	(0.0002)	-0.0002	(0.0002)	-0.0015	(0.0079)
Urban zone x schooling years (other member, squared)	2.20E-05	(0.0000)	3.07E-05	(0.0000)		
Urban zone x access to credit	0.0004	(0.0005)	0.0004	(0.0004)	0.0560	(0.0540)

table continues…

Urban zone x access to drinking water	0.0009	(0.0007)	0.0010	(0.0006)		
Urban zone x access to electricity	-1.31E-04	(0.0007)	-4.18E-05	(0.0006)	-7.86E-04	(0.0006)
Urban zone x access to in-house drainage services	-0.0003	(0.0015)	-0.0001	(0.0011)	-0.0006	(0.0009)
Urban zone x cattle	-0.0009	(0.0013)	-0.0004	(0.0012)	-0.0223	(0.1018)
Urban zone x durable goods	-0.0003	(0.0001)	-0.0003	(0.0001)	0.0519	(0.0056)
Urban zone x durable goods (squared)	6.12E-06	(0.0000)	5.38E-06	(0.0000)	-3.06E-04	(0.0000)
Urban zone x fertilizers usage	-0.0011	(0.0008)	-0.0011	(0.0008)	-0.1592	(0.0816)
Urban zone x hh size	0.0009	(0.0004)	0.0013	(0.0003)	0.0609	(0.0326)
Urban zone x hh size (squared)	-0.0001	(0.0000)	-0.0001	(0.0000)	-0.0054	(0.0024)
Urban zone x illiteracy rate	7.28E-06	(0.0000)	6.38E-06	(0.0000)	7.38E-04	(0.0010)
Urban zone x number of migrants	0.0001	(0.0001)	0.0001	(0.0001)		
Urban zone x number of migrants (squared)	-0.0001	(0.0004)	-0.0003	(0.0004)		
Urban zone x number of room in the house	-2.31E-05	(0.0000)	-3.27E-05	(0.0000)	-0.0004	(0.0122)
Urban zone x pesticides usage	0.2702	(0.0764)	0.3074	(0.0659)	0.1272	(0.0326)
Urban zone x potential work experience	0.0001	(0.0001)	0.0001	(0.0001)	-0.0032	(0.0059)
Urban zone x potential work experience (squared)	-7.84E-07	(0.0000)	-1.12E-06	(0.0000)	0.0001	(0.0001)
Urban zone x savings	0.0006	(0.0003)	0.0008	(0.0003)	-0.0535	(0.0255)
Urban zone x schooling attendance rate					0.0006	(0.0005)
Urban zone x seeds usage	-0.0024	(0.0008)	-0.0017	(0.0007)	0.0109	(0.0830)
Urban zone x social networks	-0.0009	(0.0005)	-0.0011	(0.0005)	0.0554	(0.0770)
Urban zone x spells of illness	0.0003	(0.0002)	0.0001	(0.0002)		
Urban zone x Urban zone x inadequate floor	4.02E-05	(0.0000)	3.51E-05	(0.0000)	0.0004	(0.0005)
Urban zone x working children	2.04E-05	(0.0000)	1.62E-05	(0.0000)	-0.0989	(0.0863)
Number of observation	4949		4949		3623	
R-squared	0.7546		0.7612		0.8596	

Note: Standard deviation in parenthesis. [1] Based on 1985-86 LSMS. [2]Based on 1994 LSMS.
Source: Authors' calculations based on LSMS 1985/6 and 1994.

Annex 3: Results of spatial autocorrelation at the province level

Table A3.1: Spatial correction at province level

Variables	Moran Index	Z-Value	Geary Index	Z-Value
South latitude	0.9302	20.21 *	0.057	-18.76 *
North longitude	0.8870	19.27 *	0.093	-18.04 *
Precipitation	0.7573	16.47 *	0.259	-14.73 *
Hh size 1993	0.7495	16.30 *	0.241	-15.10 *
Temperature (average)	0.7486	16.29 *	0.256	-14.79 *
Temperature (min.)	0.7469	16.25 *	0.255	-14.83 *
Temperature (max.)	0.7422	16.15 *	0.265	-14.62 *
Altitude of district capital (meters over sea level)	0.6693	14.57 *	0.322	-13.47 *
% female-headed hh 1993	0.6560	14.28 *	0.325	-13.43 *
Inadequate floor	0.6518	14.19 *	0.339	-13.16 *
Soil depth	0.6422	13.99 *	0.328	-13.37 *
Total illiteracy rate 1981	0.6352	13.83 *	0.356	-12.82 *
Overcrowded houses 1993	0.6286	13.69 *	0.339	-13.15 *
Hh size 1981	0.6130	13.35 *	0.377	-12.39 *
Per capita expenditure in 1981	0.6084	13.26 *	0.399	-11.95 *
Perimeter of the province	0.6032	13.14 *	0.390	-12.12 *
Female illiteracy rate 1993	0.6030	13.14 *	0.389	-12.16 *
Igneous rocks	0.5994	13.06 *	0.389	-12.14 *
Total illiteracy rate 1993	0.5977	13.02 *	0.397	-11.99 *
Female illiteracy rate 1981	0.5948	12.96 *	0.386	-12.20 *
Malnutrition rate 1993	0.5871	12.80 *	0.389	-12.14 *
Schooling years 1993	0.5833	12.71 *	0.396	-12.02 *
Potential bioclimate score	0.5798	12.64 *	0.412	-11.68 *
Forestry land potential score	0.5798	12.64 *	0.425	-11.43 *
% urban population in 1993	0.5781	12.60 *	0.437	-11.19 *
Soil slope	0.5750	12.53 *	0.395	-12.02 *
Population 1993	0.5740	12.51 *	0.440	-11.13 *
Forestry potential bioclimate score	0.5738	12.51 *	0.432	-11.30 *
Natural resources score	0.5721	12.47 *	0.413	-11.67 *
Total area of the province	0.5712	12.45 *	0.351	-12.91 *
Living standard 1993 according to FONCODES	0.5609	12.23 *	0.436	-11.22 *
% hh without electric appliances 1993	0.5577	12.16 *	0.426	-11.41 *
Male Illiteracy rate 1993	0.5558	12.12 *	0.441	-11.12 *
Rural basic needs: hh head with low schooling 1993	0.5536	12.07 *	0.419	-11.55 *
Number of rooms per house 1993	0.5521	12.04 *	0.424	-11.45 *
Urban basic needs: hh head with low schooling 1993	0.5392	11.76 *	0.464	-10.66 *
Urban basic needs: inadequate housing 1993	0.5382	11.74 *	0.450	-10.95 *
Foncodes poverty index 1996	0.5372	11.72 *	0.459	-10.75 *

table continues…

Total illiteracy rate 1972	0.5352	11.68 *	0.453	-10.87 *
Total land potential score	0.5344	11.66 *	0.447	-11.01 *
Change per capita expenditure 1981-93	0.5267	11.49 *	0.462	-10.71 *
Per capita expenditure in 1993	0.5265	11.49 *	0.457	-10.81 *
Hh size 1972	0.5183	11.31 *	0.471	-10.52 *
School attendance 1993	0.5074	11.07 *	0.475	-10.44 *
Child mortality rate 1993	0.5070	11.07 *	0.481	-10.31 *
Rate of migration 1988-93	0.5056	11.04 *	0.514	-9.66 *
Foncodes poverty ranking 1996	0.5023	10.96 *	0.491	-10.12 *
School attendance 1981	0.5004	10.92 *	0.481	-10.33 *
Elementary school attendance 1981	0.4940	10.78 *	0.496	-10.02 *
School attendance 1972	0.4861	10.61 *	0.493	-10.08 *
Agriculture land potential score	0.4833	10.55 *	0.490	-10.15 *
Agriculture potential bioclimate score	0.4825	10.54 *	0.501	-9.93 *
Climate II zones	0.4731	10.33 *	0.511	-9.72 *
Unsatisfied basic needs	0.4590	10.03 *	0.546	-9.02 *
Cattle potential bioclimate score	0.4446	9.72 *	0.539	-9.17 *
Per capita expenditure in 1972	0.4399	9.62 *	0.608	-7.80 *
% hh without electric appliances 1972	0.4398	9.61 *	0.548	-9.00 *
Access to drinking water 1993	0.4376	9.57 *	0.558	-8.79 *
Earthquake zone	0.4306	9.42 *	0.558	-8.79 *
Metamorphic rocks	0.4221	9.23 *	0.564	-8.66 *
Change in female-headed hh 1972-93	0.4214	9.22 *	0.596	-8.04 *
Climate I zones	0.4173	9.13 *	0.583	-8.30 *
% of child workers 1981	0.4124	9.02 *	0.571	-8.53 *
Total fallow crop land	0.4111	8.99 *	0.639	-7.18 *
Access to electricity 1993	0.4081	8.93 *	0.584	-8.28 *
Change in illiteracy rate 1972-93	0.3858	8.45 *	0.640	-7.16 *
School attendance 1972-93	0.3857	8.44 *	0.614	-7.67 *
Cattle land potential score	0.3842	8.41 *	0.614	-7.67 *
Change in illiteracy rate 1972-81	0.3814	8.35 *	0.614	-7.68 *
Access to electricity 1981	0.3812	8.35 *	0.600	-7.95 *
Access to sanitation services 1993	0.3811	8.35 *	0.608	-7.79 *
Types of natural resources	0.3804	8.33 *	0.615	-7.65 *
Change in per capita expenditure 1972-81	0.3762	8.24 *	0.619	-7.59 *
% of rural population 1993	0.3643	7.98 *	0.658	-6.80 *
Access to sanitation services 1981	0.3577	7.84 *	0.625	-7.46 *
% of permanent crops for own consumption	0.3540	7.76 *	0.635	-7.26 *
Access to drinking water 1981	0.3505	7.68 *	0.622	-7.51 *
Rural population 1993	0.3501	7.68 *	0.658	-6.79 *
Change in female illiteracy rate 1972-93	0.3486	7.65 *	0.675	-6.47 *
School attendance 1981-93	0.3329	7.31 *	0.643	-7.11 *

table continues…

Non-irrigated land for temporal crops	0.3327	7.30 *	0.663	-6.70 *
Total fallow land	0.3285	7.21 *	0.731	-5.34 *
Economically active pop. without profession 1993	0.3281	7.20 *	0.674	-6.48 *
Total temporal irrigated land	0.3241	7.11 *	0.680	-6.36 *
% female-headed hh 1972	0.3238	7.11 *	0.669	-6.59 *
Annual per-capita income 1981	0.3234	7.10 *	0.655	-6.86 *
Inadequate ceiling 1993	0.3173	6.97 *	0.665	-6.67 *
Total grassland	0.3148	6.91 *	0.742	-5.13 *
% of working children 1993	0.3050	6.70 *	0.639	-7.18 *
Temporal crops sold in farm	0.2984	6.56 *	0.683	-6.30 *
Number of medics 1993	0.2895	6.37 *	0.730	-5.38 *
Number of rural houses 1993	0.2883	6.34 *	0.722	-5.53 *
Total agricultural land	0.2880	6.34 *	0.728	-5.41 *
Total used land crop	0.2784	6.13 *	0.718	-5.61 *
Change in access to drinking water 1981-93	0.2755	6.07 *	0.717	-5.63 *
Total agrarian units	0.2751	6.06 *	0.731	-5.36 *
Change in access to sanitation services 1972-93	0.2713	5.97 *	0.790	-4.17 *
Change hh size 1981-93	0.2700	5.95 *	0.695	-6.06 *
Change in access to drinking water 1972-93	0.2698	5.94 *	0.781	-4.35 *
Change in access to sanitation services 1981-93	0.2691	5.93 *	0.723	-5.51 *
Number of rural towns	0.2685	5.91 *	0.786	-4.26 *
Land of crops sold in farm	0.2677	5.90 *	0.704	-5.88 *
Inadequate ceiling 1981	0.2643	5.82 *	0.727	-5.43 *
Total harvested land	0.2639	5.81 *	0.737	-5.22 *
Temporal crops for own consumption	0.2622	5.78 *	0.790	-4.18 *
Hospital beds per thousands inhabitants 1981	0.2600	5.73 *	0.751	-4.95 *
Change in non durable goods 1972-93	0.2557	5.64 *	0.799	-4.00 *
Change in women illiteracy rate 1981-93	0.2514	5.54 *	0.788	-4.21 *
Economically active pop. without profession 1981-93	0.2491	5.50 *	0.711	-5.75 *
Change in per capita expenditure 1972-93	0.2460	5.43 *	0.731	-5.35 *
Change in child workers 1981-93	0.2443	5.39 *	0.716	-5.65 *
Change access to drinking water 1981-93	0.2325	5.14 *	0.807	-3.84 *
Land for sale	0.2313	5.11 *	0.712	-5.72 *
Change in child workers 1981-93	0.2306	5.10 *	0.754	-4.90 *
Change hh size	0.2271	5.02 *	0.750	-4.96 *
Number of towns	0.2254	4.98 *	0.869	-2.60 *
Temporal crops sold in market	0.2184	4.83 *	0.766	-4.65 *
Change in access to electricity 1972-81	0.2110	4.67 *	0.844	-3.10 *
Permanent crop land	0.2095	4.64 *	0.844	-3.09 *
Change in female illiteracy rate 1972-81	0.2032	4.50 *	0.818	-3.62 *
% child workers 1972	0.2016	4.47 *	0.794	-4.09 *
Access to drinking water 1972	0.2006	4.45 *	0.793	-4.11 *

table continues…

Access to electricity 1972	0.1955	4.34 *	0.797	-4.03 *
Change in school attendance 1972-81	0.1938	4.30 *	0.806	-3.86 *
Land used for own crops/production	0.1920	4.26 *	0.878	-2.44 ~
Total no agricultural land	0.1872	4.16 *	0.828	-3.42 *
Population with college education	0.1817	4.04 *	0.835	-3.27 *
Total land	0.1763	3.92 *	0.837	-3.24 *
Total land used	0.1711	3.81 *	0.807	-3.84 *
Change in access to sanitation services 1981-72	0.1709	3.81 *	0.941	-1.18
Permanent crops sold in market	0.1659	3.70 *	0.858	-2.82 *
Permanent crops sold in farm	0.1650	3.68 *	0.948	-1.03
Inadequate ceiling 1972-93	0.1620	3.61 *	0.856	-2.87 *
Change in access to electricity 1972-93	0.1612	3.60 *	0.897	-2.05 ~
Economically active pop. without profession 1981	0.1556	3.47 *	0.812	-3.73 *
Hospital beds per 1,000 inhabitants 1993	0.1447	3.24 *	0.870	-2.59 *
% change working children 1972-81	0.1434	3.21 *	0.862	-2.75 *
Number of urban towns	0.1153	2.60 *	1.256	5.09 *
Urban population 1993	0.1127	2.55 ~	1.282	5.61 *
Urban houses 1993	0.1122	2.54 ~	1.285	5.66 *
Total houses 1993	0.1020	2.32 ~	1.291	5.78 *
Total occupied houses 1993	0.1016	2.31 ~	1.291	5.79 *
Rural town density 1993	0.0999	2.27 ~	0.864	-2.71 *
Access to sanitation services 1972	0.0786	1.81	0.921	-1.56
Change in access to drinking water 1972-81	0.0624	1.46	1.029	0.58
Change in illiteracy rate 1981-93	-0.0688	-1.37	1.060	1.20
Rural house density 1993	0.0554	1.31	0.900	-1.98 ~
Hospital beds 1993	0.0395	0.97	0.933	-1.34
Temporal crops used as seeds	0.0383	0.94	0.912	-1.74
Rural population density 1993	0.0309	0.78	0.934	-1.32
Land whose main output was used as seeds	0.0281	0.72	0.918	-1.63
Change in hh size 1972-81	0.0188	0.52	0.988	-0.24
Forest land	0.0157	0.45	1.081	1.61
Permanent crop with main output used as seed	0.0036	0.19	1.026	0.51
Towns density 1993	0.0019	0.16	0.870	-2.58 ~
Urban towns density 1993	-0.0106	-0.12	0.825	-3.49 *
Houses density 1993	-0.0094	-0.09	0.818	-3.62 *
Population density 1993	-0.0093	-0.09	0.818	-3.61 *
Vegetation zones typology	-0.0020	0.07	0.796	-4.06 *
Urban houses density 1993	-0.0081	-0.06	0.817	-3.63 *
Urban population density 1993	-0.0070	-0.04	0.816	-3.66 *

Note: *=p<0.01, ~=p<0.5=, where p is the probability to reject null hypothesis (absence of spatial autocorrelation).

Source: Authors calculation based on National Census of Populations 1972, 1981 and 1993.

Annex 4: Alternative methods for correcting for spatial autocorrelation

Table A4.1: Comparing methods—determinants of per capita expenditure growth rates 1972-93 (OLS estimations with robust standard errors, at province level)

Variables at initial period	Models		
	(1)	(2)	(3)
Intercept	-0.0277	-1.2303	-0.3270
	(1.385)	(7.537)	(1.706)
Altitude	0.2616	0.1863	0.4580
	(0.385)	(0.171)	(0.389)
Latitude	-0.0231	-0.0815	-0.0170
	(0.019)	(0.083)	(0.019)
Longitude	-0.0182	-0.0788	-0.0171
	(0.015)	(0.068)	(0.015)
Soil slope	0.0033	0.0154	0.0035
	(0.002)	(0.010)	(0.002)
Soil depth	0.002	0.0102	0.0023
	(0.002)	(0.010)	(0.002)
Igneous rock	-0.3197 *	-1.2763 *	-0.2757 *
	(0.100)	(0.463)	(0.106)
Metamorphic rock	-0.1318	-0.6157	-0.1362
	(0.122)	(0.540)	(0.122)
Temperature	-0.0114	-0.0382	-0.0082
	(0.009)	(0.039)	(0.009)
Basic needs	-0.0222	-0.1029	-0.0225
	(0.017)	(0.073)	(0.016)
High*basic needs	0.0045	-0.0347	-0.0149
	(0.090)	(0.358)	(0.080)
School attendance rate	0.0143 *	0.0649 *	0.0144 *
	(0.003)	(0.014)	(0.003)
Female-headed households (%)	-0.0109 ~	-0.0574 ~	-0.0134 ~
	(0.005)	(0.024)	(0.005)
Working children (%)	0.0533 *	0.2151 *	0.0462 ~
	(0.020)	(0.082)	(0.018)
Household size	0.0783	0.4608	0.1057
	(0.133)	(0.573)	(0.128)
Household size growth[1]	-0.2624	-1.0146	-0.2208
	(0.140)	(0.606)	(0.136)
Number of migrants	0.0171	0.0588	0.0101
	(0.029)	(0.128)	(0.029)
Spatial autocorrelation		0.1702 *	0.2305 ~

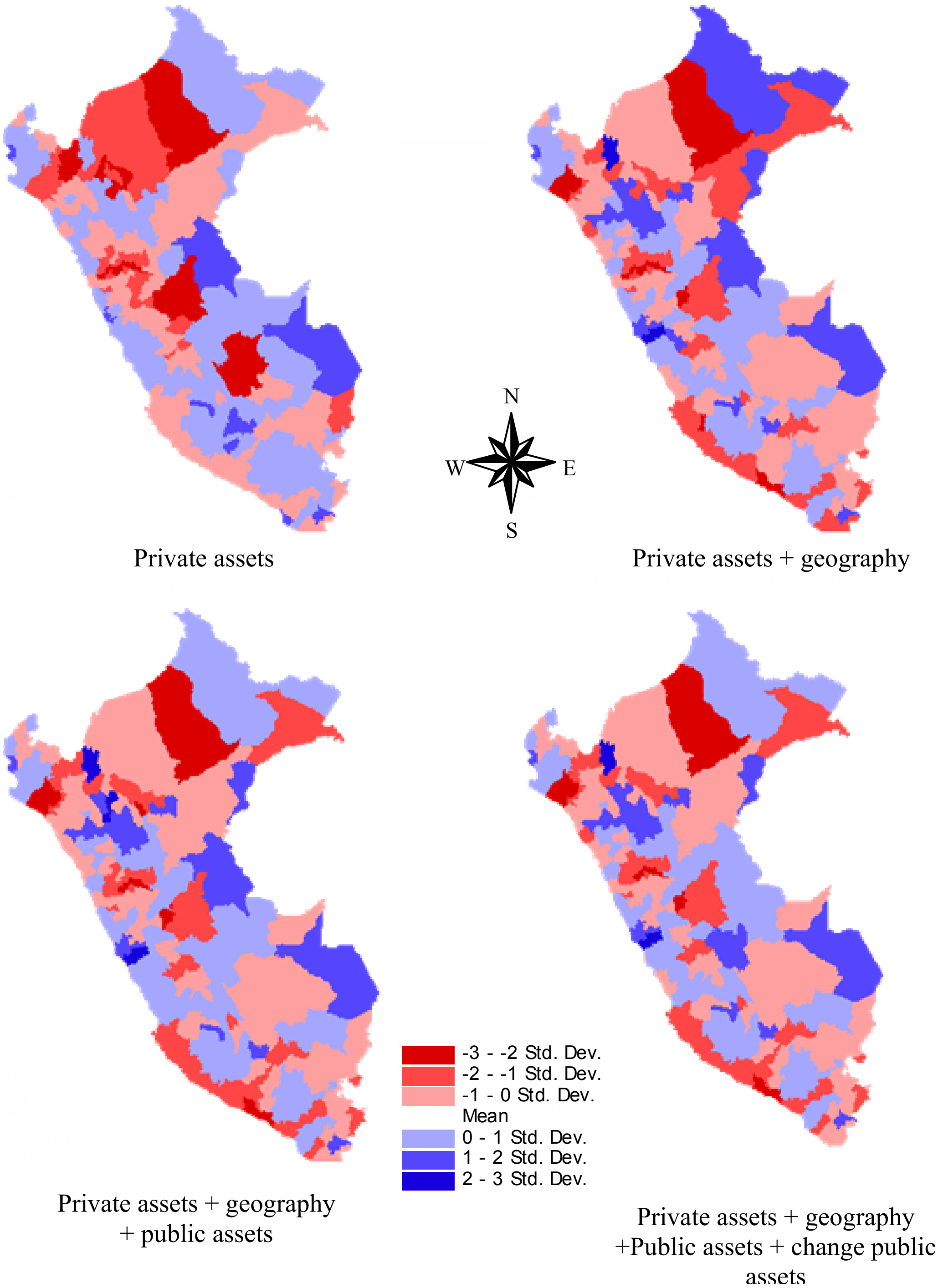
table continues…

		(0.000)	(0.102)
Number of observations	190	190	190
Adjusted R-squared	0.486	0.528	0.526

Note: Standard deviation in parenthesis and *=p<0.01, ~=p<0.5. [1]Instrumental variables are shown in the appendix Model 1: Geography + infrastructure + Geo x infra + private assets. Model 2: Geography + infrastructure + Geo x infra + private assets, modeling first-order spatial error autocorrelation. (GMM). Model 3: Geography + infrastructure + Geo x infra + private assets, modeling first-order spatial error autocorrelation (ML).

Source: Author's calculation based on Population and House Census 1972 and 1993.

Map 11: Spatial Distribution of Regression Residuals by Model



References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models,* Kluwer Academic Publishers: Amsterdam.

Anselin, L. (1995). 'Local Indicators of Spatial Association-LISA', *Geographical Analysis* 27(2):93-115.

Anselin, L., A. Varga and Z. Acs (1996). 'Local Geographic Spillovers Between University Research and High Technology Innovations: A Spatial Econometric Approach', *Regional Institute and Department of Economics Research Papers* 9606, West Virginia University: Morgantown.

Bloom, D.E. and J. Sachs (1998). *Geography, Demography, and Economic Growth in Africa*, Harvard Institute for International Development, Harvard University: Cambridge MA.

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

Cliff, A.D. and J.K. Ord (1972). 'Testing for Spatial Autocorrelation Among Regression Residuals', *Geographical Analysis* 4:267-84.

Cliff, A.D. and J.K. Ord (1981). *Spatial Processes*, Pion: London.

Eichengreen, B. (1998). 'Geography as Destiny: A Brief History of Economic Growth', *Foreign Affairs* 77(2):128-33.

Engerman, S.L. and K.L. Sokoloff (1997). 'Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies: A View from Economic Historians of the United States', in S. Haber (ed.) *How Latin American Fell Behind*, Stanford University Press: Palo Alto.

Escobal, J., J. Saavedra and M. Torero (1998). 'Los Activos de los Pobres en el Peru' (mimeo), report presented to IDB Research Network.

Escobal, J. and M. Torero (2000). 'Does Geography Explain Differences in Economic Growth in Peru?', *IDB Research Network Working Papers* 404, Inter-American Development Bank: Washington DC.

Escobal, J., M. Torero and C. Ponce (2001). 'Focalización Geográfica del Gasto Social: Mapas de Pobreza', Red CIES de Pobreza, GRADE-APOYO, Informe Final, available at www.grade/

Friedman, M. (1992). 'Communication: Do Old Fallacies Ever Die?', *Journal of Economic Literature* XXX(December):2129-32.

FONCODES (1995). *El Mapa de la Inversión Social: FONCODES frente a la Pobreza 1991-94*, UNICEF: Lima.

Gallup, J.L., J. Sachs and A.D. Mellinger (1998). *Geography and Economic Development*, Harvard Institute for International Development, Harvard University: Cambridge MA.

Getis, A. and K. Ord (1992). 'The Analysis of Spatial Association by Use of Distance Statistics', *Geographical Analysis* 24:189-206.

Hall, R.E. and C. Jones (1997). 'Levels of Economic Activity Across Countries', American Economic Association Annual Meeting, 4-7 January, New Orleans.

Hall, R.E. and C. Jones (1998). 'Why do Some Countries Produce So Much More Output per Worker than Others?' (mimeo), Stanford University: Stanford.

Hentschel, J., J.O. Lanjouw, P. Lanjouw and J. Poggi (1998). 'Combining Census and Survey Data to Study Spatial Dimensions of Poverty: A Case Study of Ecuador' (mimeo), Poverty Division, World Bank: Washington DC.

Hicks, N. and P. Streeten (1979). 'Indicators of Development: The Search for a Basic Needs Yardstick', *World Development* 7(6):567-80.

Hordijk, L. (1974). 'Spatial Correlation in the Disturbance of a Linear Interregional Model', *Regional Science and Urban Economics* 4:117-40.

Hordijk, L and P. Nijkamp (1977). 'Dynamic Models of Spatial Autocorrelation', *Environment and Planning* 9:505-19.

Hordijk, L. and J. Paekinck (1976). 'Some Principles and Results in Spatial Econometrics', *Recherches Economiques de Louvain* 42:175-97.

Jalan, J. and M. Ravallion (1998). 'Geographic Poverty Traps?', *Institute for Economic Development Discussion Papers* 86, IED, Boston University: Boston.

Moran, P.A.P. (1950). 'Notes on Continuous Stochastic Processes' *Biometrika* 37:17-23.

Moreno, R. and B. Trehan (1997). *Location and the Growth of Nations*, Economic Research Department, Federal Reserve Bank of San Francisco: San Francisco.

Peñaherrera, C. (1986). 'El desarrollo de la geografía en el Perú', in E. Yepes (ed.) *Estudios de historia de la ciencia en el Perú,* Ciencias Sociales, Sociedad Peruana de Historia de la Ciencia y la Tecnología: Lima.

Pulgar Vidal, J. (1946). *Geografía del Perú: las ocho regiones naturales*, Décima Edición 1986, Lima Editorial Universo: Lima.

Ravallion, M. and Q. Wodon (1997). 'Poor Areas, Or Only Poor People?', *Policy Research Working Papers* 1798, World Bank. Washington DC.

Streeten, P. (1984). 'Basic Needs: Some Unsettled Questions', *World Development* 12:973–82)

World Bank (1999). *Poverty and Social Development in Peru: 1994-97,* World Bank Country Study, World Bank: Washington DC.