cepe
Centre for Energy Policy and Economics
Swiss Federal Institutes of Technology

Regional disparities in electrification of India – do geographic factors matter?

Kemmler Andreas

CEPE Working Paper No. 51
November 2006

CEPE
Zurichbergstrasse 18 (ZUE)
CH-8032 Zürich
www.cepe.ethz.ch

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
PAUL SCHERRER INSTITUT
PSI

Regional disparities in electrification of India – do geographic factors matter?

Kemmler Andreas

Centre for Energy Policy and Economics (CEPE), ETH Zurich

Abstract

Modern energy sources are important input factors for human development. Although official estimates indicate that 85% of Indian villages are electrified, fewer than 60% of Indian households actually consume electricity. Therefore, one observes a considerable spatial heterogeneity in electrification rate.

This paper examines the factors that influence household and village electrification, with particular attention given to the influence of geographic factors. The analysis shows that village electrification is constrained by state area and village structure. In addition, a high share of agricultural areas seems to have a positive effect. Household electrification depends on household characteristics, the degree of community electrification, and the quality of electricity supply, and it is independent of geographic factors. Surprisingly, household expenditure and, in particular, the electricity tariff show only a relatively small effect on a household‘s choice for electricity.

1. Introduction

Access to modern energy sources is important for human development. According to Sen's capability framework (e.g. Sen, 1993, 1997), energy carriers can be understood as commodities or input factors that frame an individual's capability set and thus enable his functioning in society. In particular, electricity expands one's set of capabilities as it provides lighting, motive power and access to mass media and telecommunications, and permits cooling of rooms and the preservation of edibles. In this way, effective access is probably more important than consumed electricity quantity. Generally, indicators of wellbeing such as income, education or access to clean water increase with access to electricity, whereas the absence of any electricity use is often associated with poverty (IEA, 2002; Pachauri et al., 2004). Consequently, the relationship between household electricity consumption and poverty is bi-directional. On the one hand, access to electricity can contribute to poverty alleviation; on the other hand, lack of access is a sign of poverty.

Although electrification is an important development goal, a large share of the rural population in developing countries still lacks access to electricity. One observes remarkable regional differences in electrification, with areas in which the rate of electrified households is lower than in others. The aim of this paper is to investigate the causes of the spatial disparities in electrification rates in India. To this end, factors that affect household access to electricity are analysed, and particular attention is given to the effects of geographic factors upon the village electrification process. If geographic factors indeed influence the village electrification process, then there would be a causal relationship between the geographic endowment of a region and its level of electrification. This could explain why certain states have more difficulty completing village electrification, and a determination of the barriers to electricity use may lead to improved household access and more reasonable tariffs.

The rest of the paper is organised as follows: Section 2 provides some background on the current state of electrification in India, theoretical considerations about poor areas and a presentation of the applied analysis framework. In sections 3 and 4 two separate models for village and household electrification are introduced and discussed. Section 5 concludes with a summary of major findings and some general lessons that emerge from this work.

2. Background and analysis framework

2.1. Electrification of India

In developing countries there are still about 1.5 - 2 billion people who lack access to electricity, and 450 million of these individuals are in India alone. Despite the striking increase in power generation capabilities, India has been unable to keep up with its domestic demand for electricity. Besides the shortfall in power generation capability, India's transmission and distribution (T&D) infrastructure is inadequate to meet future demand. Moreover, due to high T&D losses, non-rational tariffs, and the fact that farmers are commonly provided with free electricity for irrigation pump-sets, the financial situation of state-owned utilities, the State Electricity Boards (SEB), dramatically worsened from the late 1980s to the 1990s. The importance that the utilities attribute to turning themselves around financially leads to their focusing on paying customers, who essentially are urban and industrial, while neglecting rural supply and electrification (Balasubramaniam and Shukla, 2003). In light of these disparities, new policies were introduced to restructure and reform the electricity sector: These include the Electricity Regulatory Commissions Act (1998), designed to promote investment friendliness and provide transparency in tariff-setting, and the Electricity Act (2003), which serves as a basis for a liberalised electricity market. By implementing these measures, the government aims to complete village electrification by 2007 and household electrification by 2012.

Figure 1 shows the village electrification rate in the 16 big states from 1970 – 2000. The states Kerala, Tamil Nadu, Haryana and Punjab already had a high level of electrified villages in the early 1970s and completed all village electrification before 1980. The other states seem to have a similar rate of electrification and differ merely in their initial levels. However, in the states Orissa, Uttar Pradesh, West Bengal, Bihar and Assam, the village electrification process began to stagnate in the 1990s before being completed.

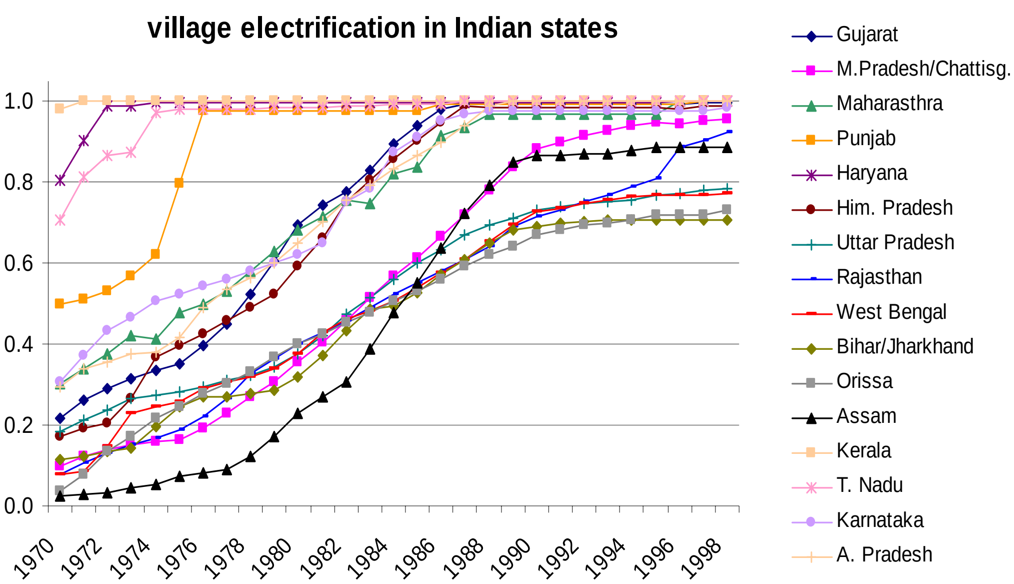

Figure 1. The electrification rate, or share of electrified villages, in the 16 big states from 1970 – 2000. Data taken from CMIE (1995, 1999, 2002).

Today, 85% of Indian villages are electrified (Srivastava and Rehman, 2006). However, fewer than 60% of households actually consume electricity. In this way, one observes large spatial differences in electrification rate. This difference in village electrification rates among the states is illustrated in Figure 1. Moreover, there is a large difference between rural and urban areas. Calculations based on NSS data show that about 81.5% of urban households are electrified, whereas in rural areas this rate is only 46.2% (NSS data for 2000). Even within rural households there is a remarkable spatial heterogeneity in electrification rates, with some areas having a higher share of households without access to electricity than others (Figure 2). Electrification rates are particularly low in the eastern and north-eastern regions.

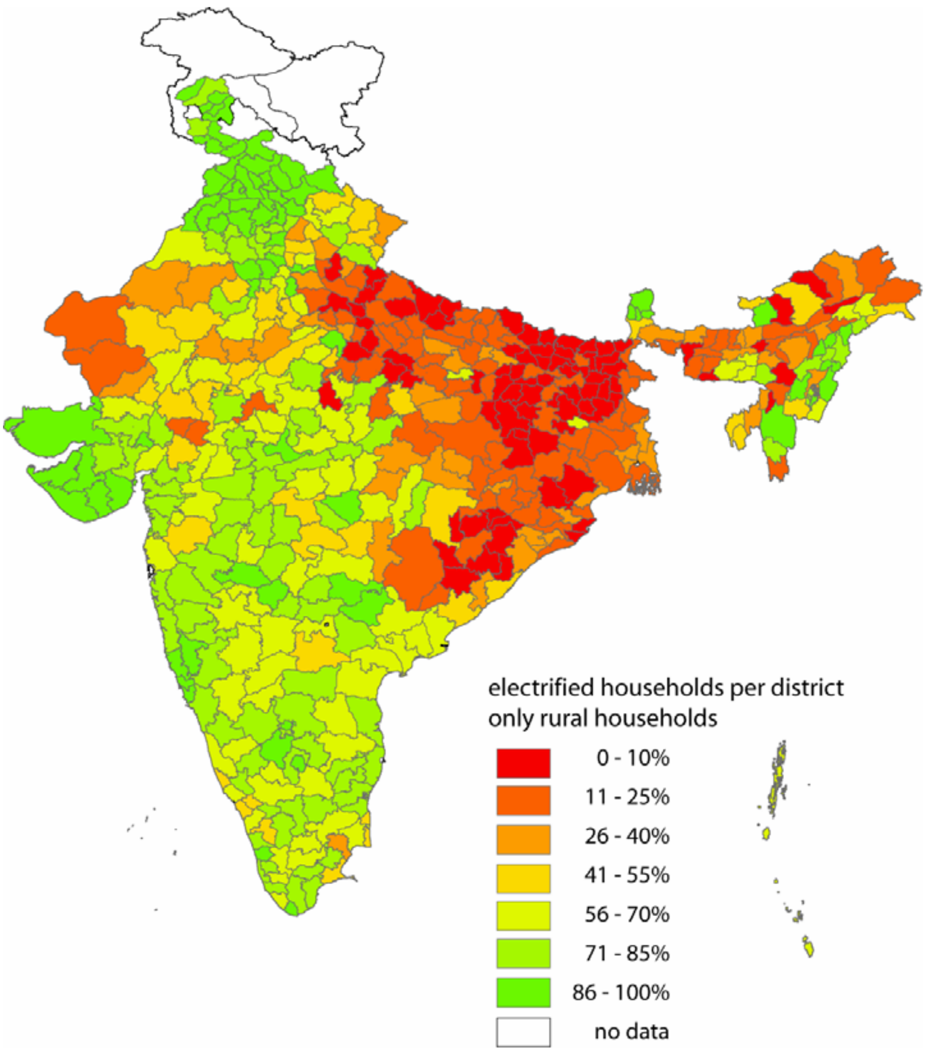

Figure 2. Percentage of electrified rural households per district. Calculations based on NSS data, round 55 (year 1999-2000).

2.2. Poor areas, geography, and electrification

Lack of access to electricity is generally related to energy poverty. Consequently, a spatial heterogeneity in access to electricity leads to spatial differences in energy poverty. In economic theory, two lines of explanations for regional differences in income poverty and poor areas are discussed (Crump, 1997). Researchers who employ an individualistic model assume that people are highly mobile and attribute no causal significance to spatial inequalities in resource endowments (geographic capital), although differences in geographic endowment may function as a sorting mechanism that leads to spatial poverty concentration (Henninger, 1998). Causes of poverty are identified at the individual level, and poor areas are described as consequences of personal decisions. On the other hand, researchers who use structural explanations suggest a causal link between the geographic endowment of a region and the general level of wellbeing of the people living in that area. It is assumed that local factors like land-use type, climate, infrastructure and access to services influence the marginal returns on investments. Because of limited mobility, structural differences in terms of natural resource endowment tend to persist and intensify between regions (Ravaillon, 1996). Each of the two theoretical models has shortcomings in explaining the spatial clustering of the poor, with a combination of individual and structural factors often identified as the cause of poverty and its spatial concentration[1] (Miller, 1996). The degree to which individual or structural factors cause poverty has implications for developing a strategy to improve the situation of the poor.

Are individual or infrastructural and geographic factors causing the regional differences in household electrification illustrated in Figure 2? Unlike income, the use of electricity traditionally requires a grid infrastructure. If a village is not electrified (that is, the village is not connected to a regional power grid), then no household within that village is able to consume electricity, irrespective of its income or status[2]. Even though the accessibility of electricity generally does not depend upon the availability of local resources, geographic factors are likely to influence the construction of the grid infrastructure and thus are relevant for explaining regional differences in village electrification rate. For instance, Chaurey reports that, within a given district, the electrification of a village may take place solely on account of its physical location (Chaurey et al., 2004). Also, certain land-use types may complicate the erection of the grid infrastructure, thus increasing costs and making the village's connection financially unattractive. In this way, geographic endowment acts as a sorting mechanism by influencing the decision process as to which villages get electrified before others.

[1] Structural factors include geographic endowment and infrastructure.
[2] Of course, this is only true disregarding stand-alone systems, which up to now have not been very widespread.

Although village electrification is traditionally a prerequisite for household access, there is a large gap between the share of electrified villages per state and the share of electrified households (Figure 3). The rate of village and household electrification is particularly low in the states of Assam, Bihar, Orissa, Uttar Pradesh and West Bengal. But even in states in which all the villages are officially connected to the power grid, large shares of the rural households do not use electricity. Moreover, in electrified villages, the proportion of households that actually consume electricity often varies considerably. Obviously, village electrification is an essential prerequisite for household electricity access, but it is not enough to guarantee it (Srivastava and Rehman, 2006). For there to be the possibility of electricity access, a grid not only has to reach a village, but it must also reach the neighbourhood and the street where the household resides. According to the literature on fuel switching, households climb up the rungs of the *energy ladder* by switching to or adding more convenient and more efficient energy sources in relation to their rising household income, assuming that the possibility of access is presented (Hosier and Dowd, 1987; Masera et al., 2000). Accordingly, income (expenditure), education, household size and fuel price are commonly the most significant factors for explaining the energy source choice of households, c.f. UNDP/WB, 2003; Heltberg, 2004. Generally, it is assumed that natural endowment has no effect on the utility of electricity use and thus does not influence the household's decision whether to use electricity.

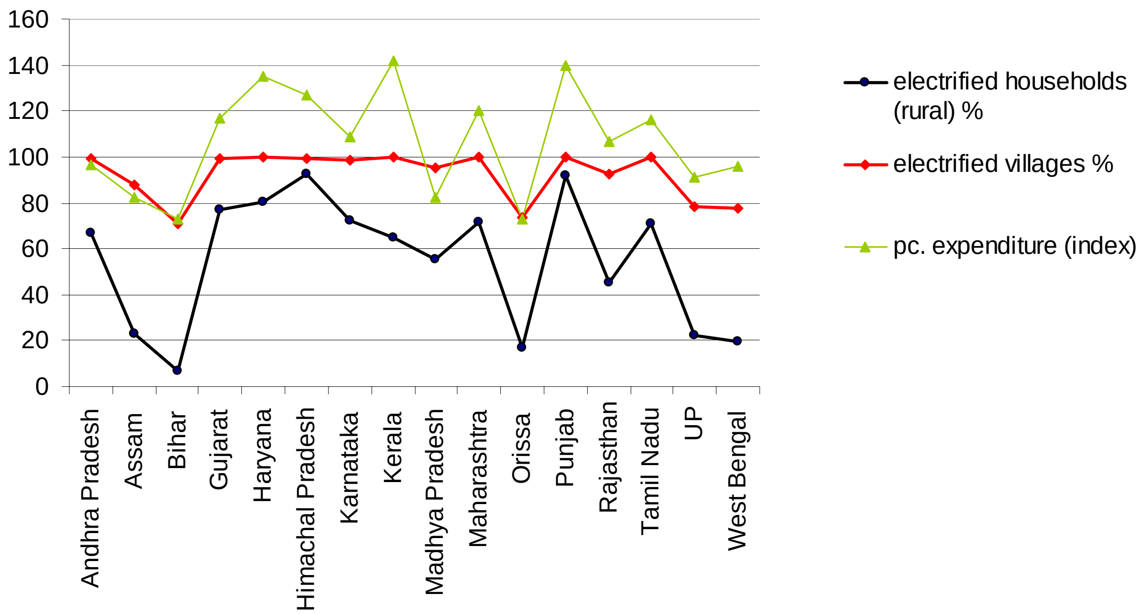

Figure 3. Comparison of village and household electrification rates per state. Both correlate well with per capita expenditure. Calculations based on NSS data, round 55 (year 1999-2000).

2.3. Analysis framework

Having observed the significance of household characteristics for household fuel choice, the dependence on grid infrastructure and the influence of geographic endowment on infrastructure erection, we see that a mix of individual, infrastructural, and geographical factors seems to cause the observed spatial disparities in electrification. We drew a distinction between the electrification of villages and the electrification of households, as the former is a prerequisite for the latter. Our hypothesis is that the electrification of villages is influenced by geographic endowment, whereas the use of electricity in electrified villages depends on household characteristics, the attributes of the electricity supply, and community electrification, but not on geographic factors themselves. The degree to which these factors cause the low electrification rates in certain regions has implications for developing a strategy to improve the situation of the people who lack access to electricity.

In order to take into account both village and household electrification, our analysis is based on a village electrification model as well as a household electrification model. Herein, village electrification is defined as the connection of a village to a regional power grid and is the traditional prerequisite for household electrification. Household electrification is defined as the connection of a household to its local community grid. In our study, the availability of data had a strong influence on the choice and the specification of the models. When setting out to do this research, we hoped to make use of the Census of India data, which is an immense household-level data set containing information on each and every household including its precise geographic location (village name). This would have afforded the construction of geographic variables that refer to single villages and their inhabitants and identify the decisive factors for village and household electrification. Unfortunately, the Census of India data seems to be inaccessible to researchers[3].

For this reason it was decided to make use of the NSS data set for the household electrification model. These data contain a large amount of information on households' characteristics and their energy consumption. However, this does not permit the localisation of the villages in which the households exist, and thus it is not possible for us to link the households with precise geographic variables. The NSS data only allow for the identification of the district containing a household. Therefore, the household data are combined with aggregated geographic district-level variables. Then, to estimate the factors affecting the household's choice for using electricity, a discrete choice model is employed. Because the NSS data do not contain information on villages, they are not appropriate for the analysis of village electrification. Instead, we decided to use state level panel data that contain information on the share of electrified villages. The employed village

[3] Aggregated Census of India data is available, but not at the household level.

electrification model is a panel data model that allows for the analysis of the effect of geographic endowment of the states on the process of village electrification, and is expressed in the rate of electrified villages. This approach is not without drawbacks, however, as it does not permit the identification of the factors that are relevant for the connection of single villages, and the high level of aggregation may blur effects of geographical variations within states.

Despite village electrification being a prerequisite for household electrification, no physical linkage between the two models was made. However, the NSS data allow for the construction of a proxy measure of community electrification, which is used in the household access model. This measure indicates whether the village in which a household resides is electrified as well as how comprehensively the village is electrified. Despite the drawbacks outlined above, the household and the village model together allow for an assessment of the factors influencing access to electricity and an explanation of the regional disparities in the electrification rate.

3. Village electrification

3.1. Model specification

The village electrification model is outlined first. What primarily concerns us is whether geographic endowment influences the process of village electrification. The principal actors in this process are the State Electricity Boards. They are responsible for power generation, transmission and distribution, and they own the intrastate lines. In the proposed model, it is assumed that the rate of electrified villages of a state (ER) – that is, the rate of electrified villages to the total number of villages in the state – depends mainly on the SEB's built grid infrastructure (SEB), and, to a lesser extent, on the state's general development and structure (S), and geographic endowment (Geo):

$$ER = f(SEB,\ S,\ Geo)$$

The SEB built infrastructure is represented by the variables *length of installed T&D lines per state area* and the *T&D losses*. Every connection of a village to the regional power grid requires the erection of additional grid infrastructure. Generally, the easily-accessible villages (those close to the power plants) are connected first, while the remote villages are connected later[4]. T&D losses are an indicator of the condition of grid infrastructure and show how well an SEB can

[4] The connection of remote villages may require a proportionately greater number of additional transmission lines and thus the effect of new build transmission lines on village electrification might decline. To capture this effect, we tried to include the variable *squared length of transmission lines per area* in the model. However, due to multi-collinearity, this was not feasible.

maintain its grid[5]. The T&D losses cause a loss of earnings and lessen the available electricity quantity, thus potentially resulting in fewer households being supplied. Initially, the variables *per capita available electricity, installed capacity* and *length of the railway net* were also considered in the model but were dropped later due to high correlation with other variables in the model or because they were insignificant (in the case of the railway).

The state development and structure vector (S) considers the per capita state domestic product (SDP) and the shares of the three main economic sectors at the SDP. Rural village electrification might have a higher priority in states that depend on the agriculture sector, and it is assumed that wealthier states can more easily afford to connect remote and less financially-attractive villages. In addition, the per capita SDP correlates highly with per capita available electricity (p=0.92) and thus corresponds well with electricity supply.

The geographic endowment (Geo) of the states is described by the variables *state area*, *the number of villages per state area*, *the share of agriculture area* and *the difference in altitude within the state*. Larger areas with a higher number of villages require longer transmission lines for the interconnection and thus cause higher costs. Therefore, it is assumed that these variables have a negative effect on the village electrification process. On the other hand, a high share of agricultural area might have a positive effect on electrification process when the modernisation of the agriculture sector (irrigation and crop processing) goes along with rural electrification. Some argue that agriculture electrification as opposed to village electrification was the main driver for rural electrification. For instance, according to Bhattacharyya, the energisation of the irrigation pump sets was for a long time a principal aim of rural electrification. Consequently, the level of electrification was not measured as a percentage of electrified households but in the extension of electricity lines to a particular area expressed by the percentage of electrified villages (Bhattacharyya, 2006). Mountains may form physical barriers and hamper the erection of power grids. The variable share of mountain area itself was not applicable, because in most of the states this land-use type exists only to a marginal extent. Therefore, the *variable altitude difference within a state* is used as a proxy variable. This measure correlates well with share of mountain area but is distributed more evenly among the states.

Within this paper it is assumed that the considered land-use types do not change over the observed time span. Thus, the geographic variables are treated as constants. To capture any common tendency of growing over time, a linear time trend is included in the model. It is likely that a region has a higher electrification rate if its adjacent regions have high a rate of electrification. A variogram analysis, which measures the difference of a characteristic between

[5] The poor condition of grid infrastructure is not the only reason for high T&D losses in India: theft is very common, but there were not adequate data available to consider it in the model. However, electricity theft seems to be more prevalent in urban than rural areas.

two locations in relation to their distance, shows some spatial correlation at the district level, and it is likely that such a correlation also exists on the state level. While the neighbourhood relationships between the states are not modelled explicitly, we will nevertheless allow for mutual correlation between the states, as described below.

3.2. Econometric method

The problem of village electrification shows some kind of censoring, since additional built power lines do not lead to a higher rate of electrified villages once all the villages are electrified. One way to deal with a censoring problem is to employ Tobit or Logit models. These approaches were not chosen here for two reasons. First, although there are a couple of observations with an electrification rate of around 0.99, there are only a few observations where all villages are electrified (~7%). Second, the data show high serial correlation, which cannot be dealt with straightforward using Tobit and Logit models. Instead, an arcos-sinus-root transformation ($\arcsin(\sqrt{ER})$) was applied to improve the normality of the dependent variable (Mosteller and Tukey, 1977; Stahel, 2002). For the same reason, the logarithmic value of the variable length of the T&D lines per state area was employed. The model to estimate is therefore of the form:

$$TER_{it} = \alpha_0 + \beta_1\, lntr_{it} + \beta_2 l_{it} + \beta_3\, pcsdp_{it} + \beta_4\, sa_{it} + \beta_5\, ss_{it} + \beta_6\, area_{it} + \beta_7\, vpa_{it} + \beta_8 aa_{it} + \beta_9 alti + \beta_{10} t_t + \varepsilon_{it}$$

where TER_{it} is the arcos-sinus-root transformed rate of village electrification, subscripts *i* and *t* denote the state and year, and ε_{it} is an *iid* error term.

variable	**definition**
TER	arcos-sinus-root transformed rate of village electrification
lntr	natural logarithmic transformed length of the transmission lines per state area
l	T&D losses in percentage of production
pcsdp	per capita state domestic product (SDP)
sa	share of SDP generated in the agriculture sector
ss	share of SDP generated in the service sector
area	area of the state
vpa	number of villages per state area (village density)
aa	share of agriculture area
alti	difference in altitude within a state
t	linear time trend

Table 1. Variable definitions

The above statistical model is estimated for a balanced panel data set consisting of 16 states over 29 years (464 observations). The repeated observations of a same state allow the use of panel data models that can account for unobserved heterogeneity across states. However, the number of states is considerably smaller than the number of periods (N<T). Such a data set, sometimes called time-series-cross-section data (TSCS), is an unusual case for widely used panel data specifications such as fixed effects and random effects models, in which T, the number of periods, is small relative to N, the number of units (Greene, 2003; Wooldridge, 2003). When the sample period is relatively short, one can assume that the individual effects remain constant. However, in long panel data these effects might change over time, resulting in the serial correlation of errors. The significant test statistics from an autocorrelation test in panel data indicate the presence of serial correlation in the data (Wooldridge, 2002).

For the above reasons, it was decided to pool the data across different states and use a heteroscedastic model with autoregressive errors that considers contemporaneous correlation between the cross-sections, as was proposed first by Parks (1967) and then discussed by Kmenta (1986). The Parks-Kmenta approach is attractive when N < T, or when the within-variation of many explanatory variables is very low (Farsi et al., 2006). Both conditions hold here as T is significantly larger than N and the employed geographic variables are assumed to be time-constant. In the Parks-Kmenta model the cross-sectional heteroscedasticity captures the unobserved heterogeneity across states, while the serial correlation is modelled through an autoregressive error structure. Geographic entities like regions or states are generally not mutually independent of each other but show contemporaneous correlation[6]. When this correlation is taken into account, the model may be termed a cross-sectionally-correlated and first-order autoregressive model. The particular characterization of this model is:

$$E(\varepsilon_{it}^2) = \sigma_{ii} \qquad \text{(heteroscedasticity)}$$

$$E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij} \qquad \text{(contemporaneous correlation)}$$

$$\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + u_{it} \qquad \text{(autoregressive errors)}$$

A likelihood ratio test indicates the use of state-specific first-order autocorrelation parameters ρ_i. The Parks-Kmenta method consists of two sequential feasible generalized least squares (FGLS) transformations. First, autocorrelation is removed and then the contemporaneous correlation of errors is eliminated. In this way, the correction for the contemporaneous correlation automatically corrects for any cross-sectional heteroscedasticity.

[6] A likelihood ratio test provides evidence for a correlation between the states (cross-sections).

3.3. Data

The employed state level panel data covers yearly data for the 16 big Indian states over the years 1970 –1999. The yearly data on electrified villages and the SEB indicators are taken from the *Energy* statistic books from the Centre for Monitoring Indian Economy (CMIE, 1995, 1999, 2002). The data on the state domestic product relies on work of the Economic and Political Weekly Research Foundation (EPWRF, 2003) and the information on the number of villages per state is taken from the Census 1991 (Census of India, 2006). Although these numbers seem to change slightly over time when compared with the results of other surveys, we decided to consider them constant over time[7]. The state level geographic variables were generated in a GIS.

variable	mean	std. dev.	minimum	maximum	n
rate of electrified villages	0.692	0.299	0.025	1	464
T&D lines per area [km/1000km2]	1999.08	1562.73	90.28	6757.04	464
losses [% of production]	0.210	0.063	0.047	0.58	464
pc SDP in 1000 Rs.	0.182	0.082	0.060	0.487	464
share of agricultural sector	0.425	0.109	0.161	0.656	464
share of service sector	0.347	0.071	0.200	0.521	464
share of industry sector	0.228	0.062	0.067	0.397	464
area [Mio. km2]	0.157	0.105	0.032	0.391	16
villages per km2	0.241	0.134	0.043	0.507	16
share agriculture area	0.608	0.180	0.229	0.903	16
difference altitude [1000m]	1.956	1.179	0.875	5.675	16

Table 2. Descriptive statistics of the parameters used in the village electrification analysis. (SDP in constant prices, base year 1981)

3.4. Results

The estimation results for the two village electrification models are given in Table 3[8]. The coefficients of both SEB variables are significant and show the expected direction signs: electrification increases with additional installed power lines but is constrained by T&D losses. The per capita SDP coefficient also has the expected positive sign and is significant.

---

[7] This restriction allows avoiding decreasing electrification rates when the number of villages is "increasing," particularly with regard to the possibility that the actual number of villages is not increasing, but merely the number noted in the official statistic.

[8] Beck and Katz argue that the estimated standard errors in the Parks-Kmenta model may be underestimated (Beck and Katz, 1995). Therefore, the model has also been estimated using an OLS model with panel- specific first-order autoregressive errors and panel-corrected standard errors as proposed by Beck and Katz for TSCS data. The standard errors become larger and the variables *T&D losses* and *per capita SDP* are no longer significant. However, all other variables, including the geographic variables, stay significant.

parameter	estimate	se	
ln trans.lines/area	0.103	0.008	***
losses	-0.022	0.012	*
pc. SDP	0.166	0.066	**
share agriculture	-0.145	0.029	***
share service	-0.025	0.045	-
area	-0.567	0.091	***
village density	-1.032	0.058	***
share agriculture area	0.394	0.078	***
difference altitude	0.058	0.010	***
time trend	0.018	0.001	***
intercept	0.012	0.055	-

*Table 3. Regression results of the village electrification model. ***, ** and * refer to 1%, 5% and 10% levels of significance, respectively.*

Economic structure also shows a significant effect on village electrification, but the nature of the effect is somewhat different than expected. First, there seems to be no difference between the service and industrial sectors. Although the coefficient share service sector is slightly negative, this is not significant. The share of agriculture, however, shows a significant and negative effect. Consequently, village electrification seems to be lower in states with an SDP depending heavily on agriculture. On the other hand, the model shows a significant and positive effect for the share of agricultural area, demonstrating that agriculture electrification is an important driver for rural electrification. This means that the level of village electrification is higher in regions with large agricultural areas where electricity is used for irrigation and crop processing. Therefore, village electrification is higher in states that, despite having large agricultural areas, have modern economies that do not depend on agriculture.

The variable *difference in altitude* was employed as a proxy variable for mountainous areas, as it was assumed that mountains form physical barriers that hamper village electrification. The coefficient difference in altitude is significant but, in contrast to expectations, has a positive sign. Perhaps difference in altitude is a bad proxy variable for mountainous areas, or the unexpected direction sign is due to the high level of regional aggregation[9].

The coefficients of the variables *state area* and *number of villages per area* are both highly significant and show the expected negative sign, indicating a constraining effect on the electrification process. An additional descriptive analysis reveals that those five states that have the lowest electrification rates by far (Bihar, Orissa, Assam, West Bengal and Uttar Pradesh) show the highest number of villages per area. Moreover, a look at the village composition of

[9] The direction signs and significance levels do not change if the model is estimated without this variable but, because the coefficient was significant, the variable was not removed from the model.

these states indicates that these states also have a high proportion of small villages (Figure 4). Two of the other states with a high proportion of small villages, Madhya Pradesh and Rajasthan, could not electrify all their villages yet either. Obviously, it is more difficult and less economically attractive to connect many small villages with few paying customers than to connect only a few larger villages with lots of potential customers. In addition, the unfavourable village structure may have aggravated the financial misery of these SEBs. The constraining effect of state area is smaller and thus less relevant than village structure for explaining the regional differences in village electrification.

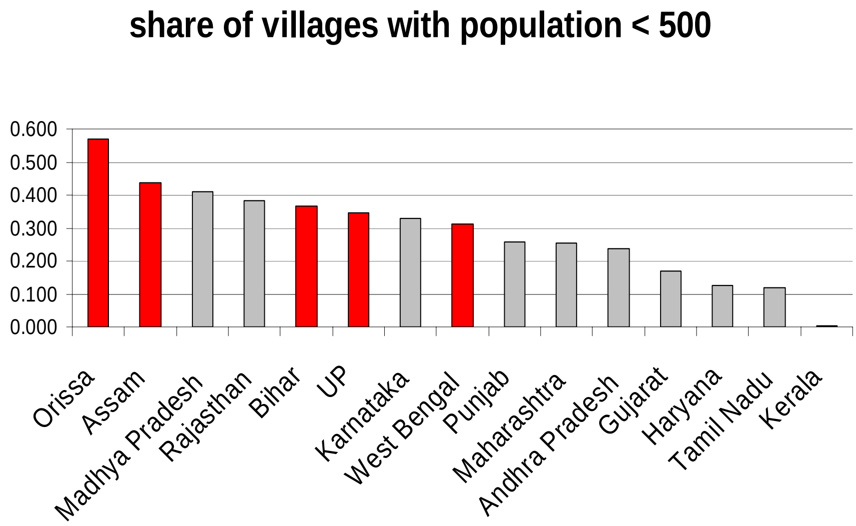

Figure 4. Share of villages with population below 500 in the larger Indian states, based on data from Census of India.

4. Household electrification

4.1 Model specification

In the second model the focus is on the electrification of households. As the focus here is on access, the factors that determine a rural household's choice whether to use electricity are analysed without taking the consumed electricity quantity into account. In the proposed binary choice model it is assumed that the choice of the households is based on the utility of the alternatives of using or not using electricity. The restricted utility (U) can be represented by the function:

$$U = f(H, E, C, L)$$

where H is a vector of household characteristics, E refers to a vector describing the attributes of the electricity supply, C refers to a vector of the community electrification and L refers to a vector of geographic location variables. The vector of household characteristics (H) considers the variables of per capita expenditure, household size, education levels of husband and wife, age and sex of the head of the household, access to liquefied petroleum gas (LPG), and information on employment type category and social group affiliation. The attributes of electricity supply (E) are described using the electricity tariff, the percentage of forced outages as a measure of the supply quality, and the supply of the alternative fuel kerosene. Low supply quality may be an important factor for not using electricity. For instance, Alam reports that, despite the growing importance of electricity, its supply is the most erratic among all major energy sources (Alam et al., 1998). Because the variable *percentage of forced outages* refers only to outages of fossil power plants, the variable *share of fossil production* was considered in the model to control for the relevance of the fossil production. Unfortunately, information on costs of connection and internal wiring was not available and could therefore not be included. To facilitate electricity access for the poor, the SEBs offer initial electricity units at a reduced tariff. This subsidised, non-cost-effective electricity tariff is the tariff considered in the model[10]. Other policies including the "Bright Home Programme," the national policy scheme to facilitate household electricity access (Kutir Jyoty)[11], are not taken into account.

As access to electricity requires the availability of grid infrastructure, it is necessary to control for whether or not a household can effectively choose to use electricity. Because this information is not directly recorded in the NSS data set, a proxy was created. As a grid has not only to reach a village but the neighbourhood and the street where the household exists as well, it follows that the likelihood that access possibility is given is higher the better the electrification is within the village. For this reason, the vector community electrification (C) is employed to control for the potential access possibility. This vector considers two variables: the availability of other infrastructure, expressed by the length of the railroad and highway net per district area, and the share of neighbouring households from the same sample village that are using electricity[12]. For instance, if none of the neighbouring households is using electricity, then the village is probably not electrified and it is unlikely that the household has potential access. On the other hand, the

[10] Average tariff for 1 kW rural (50kWh/mth) in Paises/kWh.

[11] Households below the poverty line are eligible for a single point connection. The central government bears the entire cost of service connection and internal wiring. Only about 550,000 households have participated in the programme (REC, 2006).

[12] The employed NSS data is structured into first sample units (FSU), usually a village or a city block each containing 12 household observations. Although the NSS data does not permit the identification of the exact localisation of the FSU, it is possible to identify which households belong to the same FSU and to calculate for every household observation the share of electrified neighbouring households in the FSU. In other studies, e.g. Heltberg (2004), a community is considered to be electrified if one of the households in the sample unit is electrified and not electrified otherwise. We think it is more appropriate to describe the community electrification as something continuous rather than with a dummy variable, particularly when the proxy variable is based on a small sample of 12 observations.

greater the number of neighbouring households using electricity, the more comprehensively the community is electrified and the higher the likelihood that access possibility is given.

In order to test whether the geographic endowment indeed has no direct effect on household choice, some geographical variables were included in the analysis as well; they include average yearly rainfall, temperature, altitude and shares of land-use types, all variables aggregated at the district level[13]. Additionally, state dummies were considered in the model to capture other regional effects that are not caused by geographic factors (L). To avoid multi-collinearity between the explanatory variables, some states had to be grouped into mini-regions.

Following the model, a household *i* does choose to use electricity if the utility of using it (U_1) is larger than the utility of not using it (U_0). In random utility models, the net utility for individual *i* is described by a latent variable y_i*:

$$y_i^* = U_{i1} - U_{i0} \begin{cases} > 0 & \rightarrow \text{choose } 1 \\ \leq 0 & \rightarrow \text{choose } 0 \end{cases}$$

$$= X_i \beta + u_i \text{ with } u_i = \varepsilon_{i1} - \varepsilon_{i0}$$

where *X* is the vector of all the explanatory variables of the vectors H, E, C and L; β is the corresponding vector of coefficients; and u_i the stochastic part, capturing the uncertainty. In order to estimate the vector of coefficients, a Probit model is employed. As an alternative, a Logit model is estimated and the results compared. To avoid heteroscedasticity, robust standard errors are employed.

4.2. Data

The analysis of household electrification depends mainly on unit-level budget survey data from the National Sample Survey Organisation (NSSO) of India's household consumer expenditure survey (round 55, year 1999/2000). Data from this survey include information on monetary expenditures and physical quantities of consumption of a number of household items, including electricity. The data also include information on a host of socio-economic and demographic characteristics of households. The survey collects information from a cross-section of households covering the entire area of the country over a period of one year. The employed rural sub-sample of the 16 big Indian states contains almost 60,000 household observations and, being this large, is representative of the rural population as a whole[14].

---

[13] As mentioned in section 2.3., the NSS data do not allow a further spatial disaggregation.

[14] Some states created after 1999 are included within these 16 big states; Uttar Pradesh includes Uttaranchal, Madhya Pradesh includes Chattisgarh and Bihar includes Jharkhand, respectively.

The state-level information on electricity tariffs, forced outages, energy mix and number of kerosene dealers is taken out of the *Energy* statistic book by the Centre for Monitoring Indian Economy (CMIE, 2002). The data are complemented by geographic district-level variables generated in a GIS. The employed geographic variables describe average values for each of the 428 districts of the 16 states considered. Table 4 and 5 show the descriptive statistics. The estimates are unweighted and thus do not describe the rural population but the applied sample.

variable	mean	std. dev.	minimum	maximum	n
user electricity	0.49	0.50	0	1	59543
per capita expenditure [Rs./month]*	331.94	249.93	25.8	16376.4	59543
household size	5.34	2.78	1	36	59543
share of electrified neighbours	48.73	36.14	0	100	59543
transport infrastructure length /area [km/100km^2]	4.57	2.88	0	15.41	428
electricity tariff [Paise/kWh]	135.60	56.02	59.2	246.14	16
kerosene dealers / population in Mio.	6.83	2.64	2.58	13.39	16
forced outages (%)	16.76	12.68	4.56	44.92	16
share of thermal production	60.69	29.68	4.64	99.7	16
share forest area	21.63	21.29	0	92.19	428
share mountain area	1.16	6.76	0	92.32	428
share irrigated crop area	14.45	24.82	0	99.31	428
share of non-irrigated crop area	51.05	25.01	0	99.57	428
share grazing area	5.48	8.02	0	68.45	428
share unproductive area	1.38	6.23	0	93.37	428
share water area	1.66	2.84	0	19.28	428
share other area	3.19	7.23	0	62.09	428

*Table 4) Descriptive statistics for variables. *monthly expenditure, real values with base year 1993/94.*

dummy variables	1 if, 0 otherwise	frequency	n
household variables			59543
husband illiterate		0.41	
husband primary education		0.26	
husband secondary or higher education		0.26	
wife illiterate		0.65	
wife primary education		0.17	
wife secondary or higher education		0.11	
no wife in household		0.07	
no husband in household		0.07	
age head < 30 years		0.11	
age head > 50 years		0.39	
household with Access to LPG		0.08	

social group code is tribe	0.11	
social group code is caste	0.19	
self-employed	0.15	
self-employed in agriculture sector	0.37	
wage labourer	0.07	
wage agriculture labourer	0.29	
other employment type	0.11	
district level variables		428
average temperature < 25°C	0.10	
average temperature > 27.5°C	0.09	
average yearly rainfall < 650mm	0.23	
average yearly rainfall > 1650mm	0.12	
average altitude < 75m	0.13	
average altitude > 400m	0.25	

Table 5. Descriptives for dummy variables.

4.3. Results

The estimation results of the household choice model are given in Table 6[15]. The R^2 proves satisfying for such a large and heterogeneous cross-section sample, and the coefficients show the expected direction signs. All household variables are highly significant, apart from the employment type *self-employed*. The marginal probability effects at the mean (MPE) are shown in the last row of Table 6[16]. A comparison of the MPE reveals a high correlation between household educational level and household electricity decisions. The probability of electricity use increases considerably as the education levels of the husband and wife in a household rise. On the other hand, the probability is lower in households in which the head is widowed, single or young, and in smaller households. Generally, a close relationship between access to electricity and access to LPG is observed. Only about 6-7% of LPG users have no access to electricity (year 2000). This relationship is reproduced in the high MPE of the variable access to LPG. Although this close relationship is observed, the reason for its existence is not quite clear[17]. Some possible explanations for the electricity – LPG nexus are given in the UN/WB study (2003) and in Barnes et al. (2005). For instance, in the UN/WB study it is stated that *"areas that are in some sense more "modern" (for example large as opposed to small towns and places with better*

[15] Estimations based on a Logit model show very similar results.

[16] The MPE measures the marginal change in the probability of observing electricity use in the household given a marginal change in the explaining variable. For the logarithmic variables of *expenditure* and *household size*, reported numbers can be interpreted directly as a change in percentage points. For instance, an increase in expenditure by one percent corresponds to an increase in logarithmic expenditure by 0.01.

[17] Because the effect is not clear, the model has also been estimated without this variable. The estimated coefficients hardly changed.

infrastructure) get connected first to the electricity grid", whereby the availability of an LPG market can be considered a sign of better infrastructure.

variables	coefficient	robust std.err.	sign. level	MPE at mean dy/dx
ln pc expenditure	0.734	0.021	***	0.292
ln household size	0.572	0.016	***	0.228
husband illiterate	-0.215	0.018	***	-0.085
husband education sec./ higher	0.190	0.021	***	0.076
no husband	-0.106	0.031	***	-0.042
wife illiterate	-0.152	0.021	***	-0.060
wife education sec./ higher	0.089	0.029	***	0.036
no wife	-0.165	0.032	***	-0.065
age group young	-0.065	0.025	***	-0.026
age group old	0.148	0.015	***	0.059
user LPG	0.528	0.042	***	0.206
tribe	-0.157	0.025	***	-0.062
caste	-0.145	0.019	***	-0.057
self employed	-0.023	0.022	-	-0.009
labour	-0.239	0.029	***	-0.093
labour in agriculture sector	-0.396	0.020	***	-0.155
other employment type	0.111	0.026	***	0.044
neighbourhood electrification	0.026	0.0003	***	0.010
transport infrastructure	0.012	0.003	***	0.005
minimum electricity tariff	-0.001	0.0003	***	-0.0005
kerosene dealers/pop	0.051	0.007	***	0.020
% outages	-0.033	0.001	***	-0.013
share fossil production	0.011	0.001	***	0.004
average temperature < 25°C	0.035	0.033	-	0.014
average temperature > 27.5°C	0.044	0.036	-	0.017
average yearly rainfall < 650mm	0.021	0.025	-	0.008
average yearly rainfall > 1650mm	-0.026	0.040	-	-0.010
average altitude < 75m	-0.036	0.028	-	-0.014
average altitude > 400m	-0.005	0.022	-	-0.002
share forest area	0.0001	0.001	-	0.00004
share mountain area	-0.001	0.002	-	-0.0005
share irrigated crop area	-0.0005	0.0004	-	-0.0002
share grazing area	0.001	0.001	-	0.0003
share unproductive area	-0.001	0.002	-	-0.001
share water area	-0.001	0.003		-0.0003
share other area	-0.005	0.001	***	-0.002
state dummies				
const.	-6.480	0.158	***	

n=59543 R2: 0.4787
Log pseudolikelihood = -21504.657

*Table 6. Results of the Probit model for household electrification. Omitted categories include education: primary education of man and woman; employment type: self-employed agriculture; land- use type: non-irrigated crop area. *** refers to a 1% level of significance*

Although economically poor areas largely coincide with those with low household electrification rates, the causal relationship between the two appears to be weak. That is, per capita expenditure shows only a relatively small effect on a household's decision to have electricity. A rise in expenditure of 1% increases the probability of electricity use by only 0.29%. The effect would possibly be larger if the access cost for connection and internal wiring were included in the model, or if electricity consumption were less subsidised. Furthermore, the effect could be quite different for households far away from the population mean which include, for instance, the poorest segment. However, this seems not to be the case in that an estimation of the MPE for lower income groups does not show larger effects. Moreover, the subsidised electricity tariff shows only a small effect (MPE: -0.0005)[18]. This means that a reduction in the mean tariff by 10Rs. (-7.5%) would result in an increase in probability by 0.5%. The quality of the supply seems to be more relevant than the electricity price; a decrease of 1% in outages increases the probability by about 1.3%. The availability of kerosene does not show a negative effect on electricity use. This demonstrates the fact that kerosene is often used as a complementary fuel to compensate for the erratic electricity supply rather than as an alternative energy source. Furthermore, kerosene is also used for cooking, a use for which electricity is not available as a substitute.

The social groups' scheduled castes and tribes, and in particular the employment type groups' labour and labour in agriculture, use significantly less electricity. For instance, the probability of electricity use is 15.5% lower in households in which the head is working in agricultural labour as opposed to being self-employed in agriculture. It is unclear whether these people value the benefit of electricity less or if they suffer some sort of access discrimination. These lower social groups generally live in poorer and thus potentially less-electrified neighbourhoods. Therefore, it might be more difficult for them to obtain a household connection even if they were able to afford it. On the other hand, if farm land were made accessible to electricity for irrigation, then the farm of the land-owning, self-employed agricultural worker most probably would have access as well. In any case, community electrification proves to be a crucial factor for household access. It was assumed that a higher share of electrified neighbours signifies a better access situation and a higher likelihood that the household gets connected itself. The estimations confirm this hypothesis: the coefficient share of electrified neighbours is clearly positive and highly significant. Density of highway and railroad infrastructure, the other applied proxy variable for community electrification, is less relevant despite being highly significant. On the other hand, as

[18] In the model, all households within one state pay the same electricity tariff and enjoy the same supply quality. Therefore, for the state level variables, only 16 different values were available for the statistical analysis. In an additional estimation, the option robust cluster was applied, which allows for correcting the standard errors for intragroup correlation in STATA. The significance level of the coefficients, however, did not change.

expected, none of the geographic variables is significant except for share of other area. This variable stands for a mix of different minor land-use types which are found only in certain states, and thus the effect is difficult to interpret. Nevertheless, the hypothesis that the coefficients of geographic factors are zero cannot be rejected, and we therefore conclude that geographic factors have no direct effect on the utility of electricity.

5. Conclusion

This research set out to combine conventional household- and village-level data with a GIS in order to identify geographic factors which could potentially effect electrification rates and cause regional disparities in access. While this approach offers great potential for gaining new insights, the necessary conditions were not yet fully available to exploit its entire potential. The Census of India is not made accessible for public research and therefore more highly aggregated data had to be used. The NSS and SEB data employed as an alternative permit linking geographic data only to the district and state levels respectively. Despite these limitations, the presented analysis allows for an explanation of the observed regional disparities in electrification according to a combination of factors influencing household electrification and grid availability. Areas experiencing the lowest electrification rates are such as a result of poor household characteristics and low local grid availability. In this way, some geographic variables are relevant for grid availability but not for household access. A region's having a high proportion of agricultural area correlates positively with village electrification, which demonstrates the importance of agriculture electrification as a driver for rural electrification. On the other hand, an unfavourable village structure and a large state area constrain the village electrification process. In particular, areas with small but numerous villages seem to have lower village electrification rates. Thus, this analysis provides some evidence for a causal relationship between the man-made geographic endowment of a state and its level of village electrification. However, geographic factors influence only the speed of the erection of regional infrastructure and act temporally as a sorting mechanism; they seem not to affect electrification inside the villages, as they do not change the utility of electricity use.

Even though economically poor areas largely coincide with areas with low household electrification, an analysis of household choice has shown that expenditure has an attenuated effect. Indeed, an increase in expenditure alone would hardly improve low household access rates, although a higher household expenditure in a region might increase the incentive for the utilities to expand grid infrastructure to that area. In any case, the village electrification model provided some evidence for a positive effect of income (pc SDP) on village electrification at the state level. Other factors besides expenditure, in particular community electrification and education of household members, are probably more relevant for household electrification. Furthermore, the model suggests that electrification is better extended by improving supply quality rather than subsidising consumption by a non-cost-effective tariff. The influence of the present electricity tariffs on the household decision to use electricity is small, and the undifferentiated subsidies benefit those who are already connected to the grid rather than those who are still seeking a connection. The high negative MPE of the social groups' tribe and caste, as well as the employment type groups' labour and labour in agriculture, could be a sign of large intra-village differences in community electrification. As these social groups probably live in poorer and therefore less-electrified neighbourhoods, they might suffer from some sort of access discrimination. Unfortunately, access to electricity still seems hardly a given in the hamlets surrounding the outskirts of villages, even those in regions noted for their high village electrification rates.

6. References

Alam, M., Sathaye, J. and Barnes, D. (1998). "Urban household energy use in India: efficiency and policy implications." Energy Policy. 26(11): 885-891.

Balasubramaniam, S. and Shukla, P.R. (2003). "Developing country experience with electricity reforms: case of India." Working Paper, Indian Institute if Management, Vastrapur, Ahmedabad

Barnes, D., Krutilla, K. and Hyde W. (2005). The urban household energy transition social and environmental impacts in the developing world. Publisher: Johns Hopkins University Press (June 2005).

Beck, N. and Katz, J. N. (1995). "What to do (and not to do) with time-series cross-section data." American Political Science Review 89: 634-647.

Bhattacharyya, S. C. (2006). "Energy access problem of the poor in India: is rural electrification a remedy?" Viewpoint. Energy Policy 34: 3387-3397.

Census of India (2006). Accessed 2006: http://www.censusindia.net/

Chaurey, A., Ranganathan, M. and Mohanty, P. (2004). "Electricity access for geographically disadvantaged rural communities - technology and policy insights." Energy Policy. 32: 1693-1705.

CMIE (1995). Energy. Centre for monitoring Indian Economy Pvt Ltd. Mumbai.

CMIE (1999). Energy. Centre for monitoring Indian Economy Pvt Ltd. Mumbai.

CMIE (2002). Energy. Centre for monitoring Indian Economy Pvt Ltd. Mumbai.

Crump, J. R. (1997). "Teaching the political geography of poverty." Journal of Geography. 96(2): 98-104.

EPWRF (2003). Economic and Political Weekly Research Foundation, accessed 2006: http://www.epwrf.res.in/DisplayCategories.asp?CatId=9

Farsi, M., Fetz, A. and Fillipini, M. (2006). "Economics of scale and scope in local public transportation". CEPE working paper No. 48, April 2006.

Greene, W. H. (2003). Econometric analysis. Upper Saddle River, Prentice Hall Pearson Education International.

Heltberg, R. (2004). "Fuel switching: evidence from eight developing countries." Energy Economics. 26: 869-887.

Henninger, N. (1998). "Mapping and geographic analysis of human welfare and poverty - review and assessment." World Resource Institute. Washington, D.C., USA, accessed 2006: http://www.povertymap.net/publications/doc/henninger/

Hosier, R. H. and Dowd, J. (1987). "Household fuel choice in Zimbabwe. An empirical test of the energy ladder hypothesis." Resources and Energy 9: 347-361.

IEA (2002). World energy outlook 2002. Chapter 13 on energy and poverty. International Energy Agency, Paris.

Kmenta, J. (1986). Elements of econometrics. New York, Macmillan.

Masera, O. R., Saatkamp, B. D. and Kammen, D. M. (2000). “From linear fuel switching to multiple cooking strategies: A critique and alternative to the energy ladder model.” World Development 28(12): 2083-2103.

Miller, S.M. (1996). “The great chain of poverty explanations.” In Poverty: A Global review, handbook of international poverty research, Else Øyen, S.M. Miller, and Syed Abdus Samad, eds., Scandinavian University Press and UNESCO Publishing, Oslo.

Mosteller, F. and Tukey, J. W. (1977). Data analysis and regression: a second course in statistics, A.W. Series in Behavioral Science: Quantitative Methods, Addison-Wesley, Reading.

Pachauri, S., Mueller, A., Kemmler, A. and Spreng, D. (2004). „On measuring energy poverty in Indian households.” World Development 32(12): 2083-2104.

Parks, R.W. (1967). “Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated.” Journal of the American Statistical Association, 62, 500-509.

Ravaillon, M. (1996). “Issues in measuring and modelling poverty.” The Economic Journal 106 (September): 1328-1342.

REC (2006). Rural Electrification Corporation, accessed 2006: http://recindia.nic.in/operkutir.htm

Sen, A. (1993). “Capability and well-being.” in The Quality of Life, Nussbaum, M.C. and Sen, A. (Eds.), Oxford University Press (30-53)

Sen, A. (1997). “Human capital and human capability.” Editorial, World Development. 25(12): 1959-1961.

Srivastava, L. and Rehman, I. H. (2006). “Energy for Sustainable Development in India: Linkages and Strategic Direction.” Energy Policy. 34: 643-654

Stahel, W. A. (2002). Statistische Datenanalyse - Eine Einführung für Naturwissenschaftler. 4.Auflage, vieweg Verlag, Braunschweig.

UNDP / WB (2003). Household energy use in developing countries. A multi country study. Energy Sector Management Assistance Program (ESMAP) of the UNDP and World Bank. Washington, DC.

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. Cambridge and London: MIT Press.

Wooldridge, J. M. (2003). Introductory econometrics: A modern approach." Michigan State University, Thomson South-Western.

CEPE Reports and CEPE Working Papers can mostly be downloaded free of charge in pdf-format from the CEPE Website (www.cepe.ethz.ch). Alternatively, they may be ordered from: CEPE, Secretariat, Zurichbergstrasse 18 (ZUE E), CH-8032 Zurich, Switzerland.