On a Method of Calculating Regional Price Differentials with Illustrative Evidence from India

by

D. Coondoo

Economic Research Unit
Indian Statistical Institute
203 BT Road
Calcutta 700 035
India

dcoondoo@isical.ac.in

A. Majumder

Economic Research Unit
Indian Statistical Institute
203 BT Road
Calcutta 700 035
India

amita@isical.ac.in

R. Ray

School of Economics
University of Tasmania
GPO Box 252-85
Hobart Tasmania 7001
Australia

ranjan.ray@utas.edu.au

May 2001

Acknowledgments: The authors are grateful to Professors Nikhilesh Bhattacharya and Probal Chaudhuri for their helpful advice during the work on this paper. The authors, also, thank Geoffrey Lancaster for his painstaking and skilful research assistance. Ranjan Ray acknowledges the financial assistance from an Australian Research Council (Large) Grant. The disclaimer applies.

ABSTRACT

In this paper we propose a method of estimating multilateral regional price index numbers from a given household level data set on item-wise unit values/prices. The method is closely related to the Country-Product Dummy variable model of Summers (1973). This method is likely to be particularly useful in studies of regional comparisons of poverty and inequality, optimal commodity taxes and tax reforms. To illustrate the method, we use it to calculate the regional consumer price index numbers for Eastern, Western and Southern India (taking Northern India as the reference region) separately for three categories of rural and urban households, viz., all households and those below and above the poverty line, using household level unit records of the NSS 50th round (1993–94) Consumer Expenditure Survey.

Key words: Regional price differentials, Multilateral index numbers, Country-Product Dummy model.

JEL Classification Number: C43, O18

1. INTRODUCTION

The measurement of regional differences in consumer price levels is important to policy makers in business, government and academics as well as to individual citizens faced with decisions on where to live. Estimates of the magnitude of regional price differences are needed in comparisons of real income, levels of living or consumer expenditure patterns across regions. In large Federal countries such as India and the US, with considerable heterogeneity in preferences, quality of items and household characteristics between regions, the calculation of regional price differentials, hence, acquires considerable importance. There is, therefore, a significant literature, mostly based on US data, on the measurement of regional cost of living [see, for example, Moulton (1995), Kokoski, Moulton and Zeischang (1999), Koo, Phillips and Sigalla (2000)].

When the number of regions compared is more than two, the price index number problem involved in such inter-regional real income comparisons are resolved in one of two major ways. The first and the most straightforward approach is to use binary price index numbers for pair-wise comparison of real income/level of living and then attempt to get a consistent ordinal ranking of the regions so as to obey transitivity. Examples include Sen (1976), Bhattacharya, Joshi and Roychowdhury (1980), Bhattacharya, Chatterjee and Pal (1988), and Coondoo and Saha (1990). Use of the methodology of binary price index numbers in such multilateral real income comparisons does not, however, ensure transitivity of price level comparisons except under trivial and simplifying assumptions. The International Comparison Project (ICP) of the United Nations Statistical Office and the World Bank popularised an alternative methodology of multilateral price comparisons whereby a set of internally consistent price indices, known as Purchasing Power Parities (PPP), are obtained from a set of country-wise price and quality data for a common set of items/item groups [see Geary (1958), Khamis (1972) and Kravis, Heston and Summers (1978)].

The methodology of multilateral price comparison has thrived over time both theoretically and in terms of its application to a wide variety of problem areas [see, for example, Prasada Rao (1997)]. However, like its binary price index number counterpart, the computation of a set of pure multilateral price index numbers requires a set of country-specific prices and quantities of items of uniform quality specifications, which is difficult to obtain. To resolve data problems arising from quality variation of items across regions and from gaps in the available country-wise price data, the Country Product Dummy (CPD) regression methodology is often used [see Summers (1973)]. The CPD, which is essentially an implementation of the *hedonic* approach accounting for quality variations present in the price data, offers a regression analysis-based econometric methodology of construction of multilateral price index numbers that takes care of the quality variations present in the cross-region price data [see Kokoski, Moulton and Zeischang (1999)]. The CPD methodology has undergone immense theoretical improvements – see, for example, Prasada Rao (2001) where the equivalence between a generalised CPD procedure and some standard multilateral price index number formulations has been discussed.

The literature on multilateral price index numbers is mostly concerned with the construction of PPP’s/exchange rates from item/group-wise price and quantity/expenditure/ share data available at the level of region/country. There is no reference to the use of micro-level data (for example, household level data on commodity prices/unit values available from countrywide consumer expenditure surveys) for estimation of multilateral price index numbers reflecting regional price differentials. However, given the fact that such micro-level data often contain valuable price information, it is worth while to explore if such data can be utilised to measure regional price differentials by estimating multilateral (consumer) price index numbers when the data set covers households belonging to more than one region (namely, districts within a region, states/provinces within a country or a set of countries).

The purpose of the present paper is to attempt and report such an exercise. To be precise, given a set of cross-sectional household level expenditure data obtained from a nation-wide survey, we consider the subset of items/item groups for which household level price/unit value and quantity measurements are both available. We then specify a price equation [ie., a 'quality equation' in the terminology of Prais and Houthakker (1955)] for each of these items/item groups by relating its price/unit value to the household's level of living (as measured by the household's per capita total consumer expenditure (PCE)) and a set of relevant household attributes (for example, household age-sex composition) together with two sets of dummies – one set relating to the items/products and the other set relating to the regions. The proposed methodology employs a two-stage estimation procedure. In the first stage, the item/item group-specific price equations are estimated and, hence, the region-specific estimates of slope and of the intercepts of the item-specific price equations are obtained. In the second stage, the set of multilateral regional price index numbers are estimated by regressing the region-specific intercept differentials on the corresponding slope differentials of individual items/item groups using another dummy variables based regression equation. This procedure is closely related to the CPD methodology because the price equation described above shares the *hedonic* feature which is central to the idea of the CPD model. There is, however, a basic difference – viz., we use the household PCE and attributes as *surrogates* for quality of items/item groups consumed by a sample household, rather than the information of item quality (which is usually not recorded in great detail in consumer expenditure surveys).

The paper is organised as follows: Section 2 specifies the price equation, explains via a reference to the CPD model the rationale of the proposed regression based procedure for estimating multilateral price index numbers from household level price/unit value data (Section 2.1) and describes the estimation method (Section 2.2). Section 3 presents a brief description of the data used (Section 3.1) and reports the results of the estimation (Section 3.2). The paper ends on the concluding note of Section 4.

2. METHODOLOGY

2.1 Specification of the Price Equation

In the basic CPD model, prices are regressed on two sets of dummy variables – viz, one relating to the item specifications and the other to the countries covered in the price data used (other than a country chosen as the *numeraire* country) [see Summers (1973)]. The specification of the linear regression equation of a typical CPD model is thus as follows:

$$p_{jr} = \sum_{r=1}^{R} b_r S_r + \sum_{j=1}^{M} z_j D_j + \varepsilon_{jr} \qquad (1)$$

where there are R + 1 countries with r (=0, 1, 2,…,R) denoting the individual countries and r = 0 denoting the *numeraire* country; S_r's are the country dummies; M is the number of items in a basic heading, Dj's (j = 1, 2,…,M) are the item dummies and p_{jr} is the natural logarithm of the price of item j in country r. The country coefficients, namely, the b's are the natural logarithms of the estimated country parity for the heading, and the item coefficients, namely, the z's are the natural logarithms of the prices in the currency of the *numeraire* country. It may be pointed out that here the disaggregation of consumer expenditure is reasonably detailed so that the terms *commodity* and *item* (i.e., groups of commodities) may be used interchangeably.

The CPD model was originally used for *filling gaps in available price information rather than for estimating purchasing power parities (PPP)* since it does not make use of any quantity or value data. Prasada Rao (1996) generalised the estimation procedure of this model by making use of quantity and value data[1]. Kokoski, Moulton and Zieschang (1999), Hill, Knight and Sirmans (1997) and Triplett (2000) proposed the use of CPD model to incorporate quality adjustment in estimation of PPP for regional price comparison [see also Prasada Rao (2001)]. The basic CPD model used for making quality adjustments is given by the following regression equation:

$$p_{jr} = \sum_{r=1}^{R} b_r S_r + \sum_{j=1}^{M} z_j D_j + \sum_{q=1}^{Q} \theta_q C_{qjr} + \varepsilon_{jr} \qquad (2)$$

where C_{qjr}'s, q = 1, 2,…,Q, are the set of quality characteristics that are deemed to be relevant for a given price comparison problem.

The present study proposes the use of a variant of (2) to measure interstate/regional price differentials from a given set of household level cross-sectional data on item/item group-wise prices/unit values covering two or more states/regions. The logarithmic price equation for the jth item/item group is specified as follows:

$$\left(p_{jrh} - \Pi_r\right) = \alpha_j + \sum_i \delta_{ji} n_{irh} + \left(\lambda_j + \eta_{jr}\right)\left(y_{rh} - \Pi_r\right) + \varepsilon_{jrh} \qquad (3)$$

j = 1, 2,…,M

where p_{jrh} is the natural logarithm of nominal price of jth item for the sample household h of region r (= 0, 1, 2,…,R; 0 being the *numeraire* region), y_{rh} is the natural logarithm of the corresponding nominal per capita total consumer expenditure (PCE), n_{irh} is the number of household members of the i^{th} age-sex category (i = 1, 2, 3, 4 denote adult male, adult female, male child and female child, respectively), ε_{jrh} is the associated random disturbance term and α_j, δ_{ji}, λ_j, η_{jr}, Π_r's are the model parameters. We assume the random disturbance terms associated with individual observations to have zero expectation and to be uncorrelated.

[1] Prasada Rao (2001) proposed a generalised CPD method in which a weighted residual sum of squares is minimised with each observation weighted according to the expenditure share of the item concerned in the given country.

It may be noted that equation (3) is essentially a representation of the *Quality Equation* in Prais and Houthakker's sense (Prais and Houthakker, 1955) in which the price/unit value paid by a household for a commodity is taken to be a measure of the quality of the commodity consumed and hence is a function of the household's level of living (measured typically as the per capita income/ total expenditure) and age-sex composition of the household. In the present context of inter-regional comparison of price levels, the population consists of households belonging to different regions with varying price levels. Hence, the specification of the quality equation is to be modified by (1) replacing the dependent variable price/ unit value by *real price/unit value* (i.e., nominal price deflated by the regional price level of the region to which the household belongs) and (2) measuring the level of living of the household by real income (i.e., nominal per capita income/total expenditure deflated by the regional price level).

The rationale for the inclusion of the household size-composition variables in equation (3) as additional explanatory variables is as follows: In the literature on measurement of the effects of household demographics on household consumption demand it is postulated that a change in household size/composition induces a *quasi-price substitution effect* on demand (Barten 1964). Therefore, in explaining price/unit value through a *quality equation* this quasi-price substitution effect of demographic change should be appropriately incorporated. Thus, for example, given other things, addition of a member of a specific age-sex category to a household is likely to cause a change in the price/ unit value paid for different commodities consumed and, more importantly, these changes will be category-specific.

2.2 Estimation

A CPD model is usually estimated on the basis of (R + 1)M data points. Here, however, we propose that the estimation of equation (3) be based on $M\sum_{r=0}^{R} N_r$ observations, N_r being the number of sample households in region r. In principle, Π_r may be interpreted as *the natural logarithm of the value of a reference basket of items purchased at the prices of region r*. Hence, $\Pi_r - \Pi_0$ is the natural logarithm of the price index number for the rth region with the price level of the *numeraire* region taken as the base. Equation (3) is basically the *quality equation* for the jth item/item group expressed in logarithmic form. It recognises three deterministic sources of the observed inter-household variations in the nominal price of an individual item/item group – viz., variation in regional price levels, inter-household variations in the level of living (measured by the natural logarithm of PCE) and household size and composition. As per the specification, $(\lambda_j + \eta_{jr})$ measures the *quality elasticity* of the jth item/item group in region r. To normalise these elasticities, let us set $\eta_{jr} = 0$ for every j for r = 0, so that λ_j denotes the quality elasticity of item j in the *numeraire* region.

The regression model specified in equation (3) above is nonlinear in parameters. To estimate the parameters of this model, we suggest the following two-stage method.

In the first stage, the item price equations are estimated, using OLS, on the pooled data set of all the states/regions. For this purpose, the price equation of an individual item may be expressed as the following linear regression equation:

$$p_{jrh} = \alpha_j^* + \sum_{i=1}^{4} \delta_{ji} n_{irh} + \sum_{p=1}^{R} \phi_{jp} S_p + \lambda_j y_{rh} + \sum_{p=1}^{R} \eta_{jp} y_{ph} S_p + \varepsilon_{jrh}\ , \qquad (4)$$

j = 1, 2,…,M
r = 0, 1,…,R
h = 1, 2, …, N_r,

S_p being a dummy variable, with $S_p = 1$ for p = r, and 0, otherwise.

To see the equivalence between equations (3) and (4), let us note that equation (3) can be rewritten as

$$p_{jrh} = \alpha_j + \Pi_r + \sum_i \delta_{ji} n_{irh} + (\lambda_j + \eta_{jr})(y_{rh} - \Pi_r) + \varepsilon_{jrh}$$

$$= \alpha_j + \Pi_r + \sum_i \delta_{ji} n_{irh} - (\lambda_j + \eta_{jr})\Pi_r + \lambda_j y_{rh} + \eta_{jr} y_{rh} + \varepsilon_{jrh}$$

$$= \alpha_j + (1 - \lambda_j)\Pi_0 + \sum_i \delta_{ji} n_{irh} + \{1 - (\lambda_j + \eta_{jr})\}\Pi_r - (1 - \lambda_j)\Pi_0 + \lambda_j y_{rh} + \eta_{jr} y_{rh} + \varepsilon_{jrh}\ .$$

Hence,

$$p_{jrh} = \{\alpha_j + (1 - \lambda_j)\Pi_0\} + \sum_i \delta_{ji} n_{irh} + [\{1 - (\lambda_j + \eta_{jr})\}\Pi_r - (1 - \lambda_j)\Pi_0] + \lambda_j y_{rh} + \eta_{jr} y_{rh} + \varepsilon_{jrh} \quad (5)$$

Comparing equations (4) and (5), we see that the two equations are identical with

$$\alpha_j^* = \alpha_j + (1 - \lambda_j)\Pi_0 \tag{6a}$$

$$\varphi_{jp} = \{1 - (\lambda_j + \eta_{jp})\}\Pi_p + \alpha_j - \alpha_j^* . \tag{6b}$$

Note that Π_0 denotes the parameter Π_r for the reference region (r = 0). Note also from equation (3) that $(\lambda_j + \eta_{jr})$ is the slope coefficient for state/region r (≠ 0), λ_j is that for the *numeraire* (ie., reference) state/region, while α_j^* is the intercept for the numeraire state/region and ϕ_{jr} is the differential intercept for the state/region r (≠ 0) of item/item group j. Thus, exp (ϕ_{jr}) is the price relative of item j for region r with the *numeraire* region taken as base. This model (ie., equation (4)) reduces to the standard CPD model when $\phi_{jp} = \phi_p$ for every j, $\eta_{jp} = 0$ for every j and p, and $\lambda_j = 0$ for every j. Thus equation (4) extends the CPD model to the present case of regional price variation in the context of a single country.

While the first stage estimation of equation (4) yields the estimated parameters, namely, $\hat{\alpha}_j^*, \hat{\delta}_{ji}, \hat{\phi}_{jp}, \hat{\lambda}_j, \hat{\eta}_{jp}$, the estimate of Π_p, namely, $\hat{\Pi}_p, p = 0,1,...,R$ may be obtained from the following second stage dummy variables regression equation

$$\hat{\phi}_{jr} = \sum_{p=1}^{R} \Pi_p \left(1 - (\hat{\lambda}_j + \hat{\eta}_{jp})\right) S_p - \Pi_0 (1 - \hat{\lambda}_j) + \varepsilon_{jr} \tag{7}$$

where $S_p = 1$ if p = r, and 0 otherwise. The parameters marked with hats are obtained from the price equations estimated at the first stage and ε_{jr} denotes the random disturbance term of the second stage regression equation[2].

[2] It may be noted that (7) actually is an alternative representation of

$$\hat{\phi}_{jr} = \Pi_r \left(1 - (\hat{\lambda}_j + \hat{\eta}_{jr})\right) - \Pi_0 (1 - \hat{\lambda}_j) + \varepsilon_{jr}$$

r =1, 2, …, R

which constitutes a system of R linear regression equations in each of which the term $-\Pi_0 (1 - \hat{\lambda}_j)$ appears. In other words, Π_0 in the present model is over-identified as R different estimates of this parameter may, in principle, be obtained by estimating this equation separately for r=1,2, …, R. To resolve this over-determinacy of Π_0, we propose estimation of the dummy variable regression equation (7) instead , which ensures that a

Note that equation (7) is derived from equation system (6a) – (6b) which is a system of 4 linear equations in 5 parameters, viz., $\Pi_0, \Pi_1, \Pi_2, \Pi_3$ and α_j. Thus each Π_r is a linear function of (every) α_j (which is unidentifiable and hence non-estimable, given the model). That is, the estimated Π's will have the α_j's confounded in them thus affecting the magnitude of the estimates. Actually, the Π's estimated for a given data set will contain an additive component which is some kind of an average of the non-estimable α_j's, say $\overline{\alpha}$. Thus, while the estimates of Π's *will not have any obvious interpretation*, their differences will unambiguously measure the logarithm of the price index number of one region with respect to another (as the $\overline{\alpha}$ will cancel out).

2.3 Features

The estimates of the region-specific price index numbers $\Pi_p - \Pi_0, p = 1,2,3$, are invariant to the choice of the default region due to the properties of the dummy variables regression model. This implies that the resulting price index numbers automatically fulfill the *circular* consistency required of a set of multilateral price index numbers.

The distinctive features of the proposed method are as follows:

1. Unlike most of the other methods of estimation of multilateral price index numbers, the present method does not require that price data on all items for all regions must be available for the method to work. The proposed method will work even if price data on some items are not available for some regions. As already described, for this method the first stage involves estimation of individual (logarithmic) price equations (based on item-specific price data for all the regions together). At the second stage the

---

single estimate of Π_0 is obtained. The number of observations used in the second stage estimation thus equals the number of items times the number of states/regions.

region-specific (logarithmic) price index numbers are estimated (based on linear regression equations with region-specific dummy variables) using region-specific item-wise intercept and slope differentials of the price equations estimated in the earlier stage. Therefore, if, say, for item j (j=1,2, .., M) price data for region p (p=0, 1, 2, …, R) are not available, estimation of the price equation for the jth item will not yield the estimate of ϕ_{jp}. This will, however, not hamper the estimation of the logarithmic price index numbers in the second stage (so long as the item is available in the other region) as the second stage estimation will be based on $\sum_p M_p$ observations, M_p being the number of items available in the pth region.

2. The methodology provided here is able to provide estimates of price level differences by regions, income groups or for any other grouping of households (poor and non poor , say). Thus, this method of estimation of regional price differentials is likely to be particularly useful in regional comparisons of poverty and inequality in large Federal countries with considerable regional heterogeneity in consumer preferences, quality of items and household characteristics.

3. DATA AND RESULTS

2.1.1 Data

The data base for this study is provided by the household level unit record data, in value and quantity, on consumer expenditure in the rural and urban areas collected for each of the States in India in the 50^{th} round of the National Sample Survey (NSS) (July, 1993 – June, 1994). As indicated earlier, the observation on unit price was obtained by dividing the value of expenditure on item j for household h residing in region r by the corresponding quantity. This meant that the empirical exercise was restricted to items for which the information on both value and quantity was available, namely, a subset of the food items covered by the enquiry. In the 50th round of the NSS, approximately 70,000 Indian households were surveyed in the rural areas, and 45,000 households in the urban, giving us a sample of over 1,15,000 households in one of the largest sampling exercises of its kind undertaken anywhere. The present study uses the original micro data from this survey. The sample size varies form State to State: while the number of observations (households) for a smaller State is often less than 500, that for a larger State is generally over 5,000.

Table 1 presents the list of 25 States used in this exercise. For the purpose of this study, these States are classified into 4 geographical regions, namely, North (r = 0, ie., the numeraire or reference region), South (r = 1), East (r = 2), and West (r = 3). Table 1 indicates the regional classification of the 25 States. The series on per capita total consumer expenditure (PCE), required in the estimation, was obtained by dividing total household expenditure (ie., the sum total of expenditure on food and non food items) by the household equivalence scale. In keeping with the spirit of this exercise, the equivalence scales, that were used, were estimated separately for each State – see Meenakshi and Ray (1999, Table 3) for the State specific equivalence scale estimates. These show considerable heterogeneity in the demographic effects across various States, thereby, pointing to the possibility of significant regional price differentials via the strong link between demographic and price effects often cited in the literature [e.g., see Barten (1964)].

Another feature of the empirical exercise, results of which are reported later, is that it is performed not only on all households but, also, separately on households above and below the poverty line[3]. The State specific poverty lines, taking account of size economies and

[3] It may, however, be mentioned that the procedure has been applied separately to poor and non-poor households and therefore, this does not offer comparisons across groups of households.

equivalence scale relativities, used in this study were taken from Dubey and Gangopadhyay (1998) who constructed these poverty lines separately for rural and urban areas in each State.

2.2 Results

Table 2 presents the OLS estimates of the parameters of equation (7)[4]. To show the stability of the parameters, we have also presented the jackknife coefficients[5]. The Table clearly reveals that all the estimated Π parameters are statistically significant and stable.

The relationship between the slope and intercept of equation (6b) are presented graphically in Figs. 1-18. These show the plot of $\hat{\phi}_{jp} + \hat{\Pi}_0(1-\hat{\lambda}_j)$ (on the y axis) against corresponding $1-(\hat{\lambda}_j + \hat{\eta}_{jp})$ (on the x axis). The graphs are shown separately for rural and urban sectors of the three regions for (i) all households (Figs. 1-3, 10-12), (ii) households above the poverty line (Figs. 4-6, 13-15), and (iii) households below the poverty line (Figs. 7-9, 16-18). These graphs provide a visual presentation of the (region wise) $\hat{\Pi}_r, r = 1,2,3$ estimates presented in Table 2. Each scatter diagram shows the corresponding estimated linear regression equation (without intercept). The close linear fit, underlined by a high R^2 in each case, provides strong supporting evidence for the relationships [eqns. (6a), (6b)] derived earlier (i.e., between the intercept and slope coefficients of equation (7) across items and regions).

---

4 The estimates of the parameters of equation (4) have not been presented here for reason of space. These may be available to interested readers.

5 The regressions have been run using SHAZAM. For the jackknife procedure, the regressions are run successively omitting a different observation [Judge, et.al. (1988)]. The jackknife coefficients are given by $\hat{\beta}_{(t)} = \hat{\beta} - (X'X)^{-1} X_t e_t^+$ where $e_t^+ = e_t/(1-K_{tt})$, K_{tt} being the t-th diagonal element of the matrix $X(X'X)^{-1}X'$, e_t the residuals and $\hat{\beta}$ the OLS estimate. A total of N (k + 1) coefficient vectors are generated each corresponding to a separate regression with the t-th observation dropped. The average of these N (k + 1) coefficient vectors is reported in the table.

Table 3 presents the region-specific price index numbers for each of the three different types of households separately for the rural and the urban sector with *North* taken as the reference region. The following features are worth noting:

(i) For the *all households* group, in both rural and urban areas, the price index number for the Eastern region is the highest among all the regions. The index numbers for 'South' and 'West' are lower than that of 'North', and Southern India is cheaper than Western India;

(ii) A similar pattern is observed for the group of *households above poverty line*;

(iii) For the group of *households below poverty line*, all price index numbers, except for Eastern India (urban), are greater than unity thus indicating a higher price level in Eastern India compared to that in Northern India. For rural India, *East* is the most expensive followed by *West* and then *East*;

(iv) An overall feature of the estimates presented in Table 3 is that the picture of regional differences in unit values/prices is not only sensitive to the rural-urban divide, but, perhaps more crucially, to whether a household lives below or above the poverty line.

3. CONCLUSION

In this paper we have proposed a method of estimating a set of regional price index numbers from a given household level data set on item-wise unit values/prices. The proposed method, being based on the linear regression technique, is quite simple and straightforward. To illustrate the method, we have used it to calculate regional consumer price index numbers for Eastern, Western and Southern India (taking Northern India as the reference region) separately for three categories of rural and urban households, viz., all households and those below and above the poverty line, using household level unit records of the NSS 50th round (1993 – 94) Consumer Expenditure Survey. Generally the results turn out to be robust and sensible.

The technical features of the proposed method are as follows: the method, (i) being based on household level data, is capable of bringing out the regional price differentials implicit in the given data set in a very robust manner and (ii) being based essentially on the CPD approach, which was originally devised to fill gaps in the available price data required for construction of multilateral price index numbers, will work even when all goods (and hence data on all prices) are not available in all the regions.

We may conclude by mentioning some of the potential uses of the proposed method of estimation of regional price index numbers to incorporate correction for regional price differentials in various studies on levels of living based on household level consumer expenditure data. The method is likely to be particularly useful in regional comparisons of poverty and inequality in large Federal countries such as India, Germany and U.S.A. with considerable regional heterogeneity in consumer preferences, quality of items and household characteristics. Yet another potential application of the method discussed here is in the area of optimal commodity taxes and tax reforms. Such tax calculations require reliable estimates of price elasticities which, in turn, are crucially dependent on the successful incorporation of regional variation in prices and behaviour in the tax analysis. Finally, the availability of regional price indices is useful in real income comparisons between different geographical areas within a country and in helping potential migrants with information on which to base their decision on their region of residence.

References

Barten, A.P. (1964), “Family Composition, Prices and Expenditure Patterns”, in P.E. Hart, G. Mills and J.K. Whitaker (eds.), *Econometric Analysis for National Economic Planning*, Butterworths, London, 277-292.

Bhattacharya, N., G. S. Chattejee and P. Pal (1988), “Variations in Level of Living Across Regions and Social Groups in India, 1963-64 and 1973-74”, in T.N. Srinivasan and P.K. Bardhan (eds.), *Rural Poverty in South Asia*, Oxford University Press.

Bhattacharyya, S.S., P.D. Joshi and A.B. Roychowdhury (1980), “Regional Price Indices Based on NSS 25th Round Consumer Expenditure Data”, *Sarvekshana, Journal of the NSS Organisation*, 3(4),107-21.

Coondoo, D. and S. Saha, (1990), “Between-State Differentials in Rural Consumer Prices in India: An Analysis of Intertemporal Variations”, *SANKHYA*, Series B, 52(3), 347-360.

Dubey, A. and S. Gangopadhyay (1998), “Counting the Poor: Where are the Poor in India?”, *SARVEKSHANA*, Analytical Report No 1, Department of Statistics, Government of India.

Geary, R.C. (1958), “A Note on Comparison of Exchange Rate and Purchasing Power Parities Between Countries”, *Journal of the Royal Statistical Society*, 121(Part 1), 97-99.

Hill, R. C., J. R. Knight and C. F. Sirmans (1997), “Estimating Capital Asset Pricing Indexes, *Review of Economics and Statistics*, 79, 226 –233.

Judge, G., R. Hill. W. Griffiths, H. Lutkepohl and T. Lee (1988), *Introduction to the Theory and Practice of Econometrics*, 2nd Edition, Wiley.

Khamis, S.H. (1972), “Properties and Conditions for the Existence of a New Type of Index Numbers”, *SANKHYA*, Series B, 32(Parts 1 & 2), 81-98.

Kokoski, M.F., B.R. Moulton and K.D. Zeischang (1999), “Interarea Price Comparisons for Heterogeneous Goods and Several Levels of Commodity Aggregation”, in R.E. Lipsey and A. Heston (eds.), *International and Interarea Comparisons of Prices, Income and Output*, National Bureau of Economic Research, Chicago University Press, 327 -364.

Koo, J., K. R. Phillips and F. D. Sigalla (2000), “Measuring Regional Cost of Living”, *Journal of Business and Economic Statistics,* 18(1), 127 – 136.

Kravis, I. B., A. Heston and R. Summers (1978), *International Comparison of Real Product and Purchasing Power*, John Hopkins University Press, Baltimore.

Meenakshi, J.V. and R. Ray, (1999), “Impact of Household Size, Family Composition and Socio Economic Characteristics on Poverty in Rural India”, Working Paper No. 68 of Centre for Development Economics, Delhi School of Economics, July 1999.

Moulton, B. R. (1995), “ Interarea indexes of the Cost of Shelter using Hedonic Quality adjustment Techniques”, *Journal of Econometrics*, 68 (1), 181 – 204.

Prais, S.J. and H.S. Houthakker (1955), *The Analysis of Family Budgets*, Cambridge University Press, Cambridge, (2nd Edition, 1971).

Prasada Rao, D.S. (1996), “On the Equivalence of the Generalised Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons”, Working Paper No. 5, Center for International Comparisons, University of Pennsylvania, Philadelphia.

Prasada Rao, D.S. (1997), “Aggregation Methods for International Comparison of Purchasing Power Parities and Real Income: Analytical Issues and Some Recent Developments”, *Proceedings of the International Statistical Institute*, 51st Session, 197-200.

Prasada Rao, D.S. (2001), “Weighted EKS and Generalised CPD Methods for Aggregation at Basic Heading Level and Above Basic Heading Level”, Joint World Bank- OECD Seminar on Purchasing Power Parities – Recent Advances in Methods and Applications, Washington D.C.

Summers, R. (1973), “International Price Comparisons Based Upon Incomplete Data”, *Review of Income and Wealth*, 19(1), 1-16.

Sen, A. (1976), “Real National Income”, *The Review of Economic Studies*, 43(1), 19-39.

Triplett, J. E., (2000), “ Hedonic Valuation of ‘Unpriced’ Banking Services: Application to National Accounts and Consumer Price Indexes”, Draft paper presented at NBER, Summer Institute.

Table 1: List of States Covered

North (r = 0: Reference Region)	South (r = 1)	East (r = 2)	West (r = 3)
Haryana	Andhra Pradesh	Arunachal Pradesh	Goa
Himachal Pradesh	Karnataka	Assam	Gujarat
Jammu & Kashmir	Kerala	Bihar	Maharashtra
Madhya Pradesh	Tamil Nadu	Manipur	Rajasthan
Punjab		Meghalaya	
Uttar Pradesh		Mizoram	
		Nagaland	
		Orissa	
		Sikkim	
		Tripura	
		West Bengal	

Table 2: Estimated ∏ Coefficients for Different Regions and Household Groups: Rural and Urban India

Household Group	No. of observations	OLS coefficients North ($\hat{\pi}_0$)	OLS coefficients South ($\hat{\pi}_1$)	OLS coefficients East ($\hat{\pi}_2$)	OLS coefficients West ($\hat{\pi}_3$)	R^2	Jackknife coefficients North ($\hat{\pi}_0$)	Jackknife coefficients South ($\hat{\pi}_1$)	Jackknife coefficients East ($\hat{\pi}_2$)	Jackknife coefficients West ($\hat{\pi}_3$)
RURAL INDIA										
All households	123	11.125 (0.229)	11.115 (0.238)	11.216 (0.251)	11.121 (0.242)	0.9489	11.208 (0.522)	11.199 (0.520)	11.302 (0.545)	11.208 (0.526)
Households above poverty line	123	11.277 (0.181)	11.251 (0.190)	11.358 (0.196)	11.275 (0.195)	0.9683	11.316 (0.352)	11.287 (0.352)	11.397 (0.364)	11.315 (0.359)
Households below poverty line	120	9.598 (0.051)	9.677 (0.012)	9.694 (0.083)	9.687 (0.070)	0.9998	9.502 (0.128)	9.512 (0.167)	9.593 (0.152)	9.623 (0.137)
URBAN INDIA										
All households	123	11.139 (0.227)	11.068 (0.235)	11.179 (0.248)	11.099 (0.238)	0.9508	11.194 (0.350)	11.123 (0.351)	11.237 (0.369)	11.155 (0.364)
Households above poverty line	123	11.163 (0.188)	11.085 (0.199)	11.205 (0.206)	11.114 (0.200)	0.9646	11.152 (0.219)	11.073 (0.224)	11.193 (0.233)	11.102 (0.231)
Households below poverty line	117	9.982 (0.052)	10.083 (0.073)	9.976 (0.077)	9.990 (0.047)	0.9979	9.936 (0.091)	9.998 (0.119)	9.925 (0.108)	9.924 (0.098)

∗ Figures in parentheses are the standard errors.

Table 3: Estimated Price Index Numbers for Different Regions and Household Groups: Rural and Urban India (Base: North =1.0)

Household Group	Rural India			Urban India		
	South	East	West	South	East	West
	$(e^{(\hat{\pi}_1-\hat{\pi}_0)})$	$(e^{(\hat{\pi}_2-\hat{\pi}_0)})$	$(e^{(\hat{\pi}_3-\hat{\pi}_0)})$	$(e^{(\hat{\pi}_1-\hat{\pi}_0)})$	$(e^{(\hat{\pi}_2-\hat{\pi}_0)})$	$(e^{(\hat{\pi}_3-\hat{\pi}_0)})$
All households	0.990	1.095	0.996	0.931	1.041	0.961
Households above poverty line	0.974	1.084	0.998	0.925	1.043	0.952
Households below poverty line	1.082	1.101	1.092	1.058	0.994	1.009

Figure 1: Relationship between slope and intercept parameters: Rural all households: South

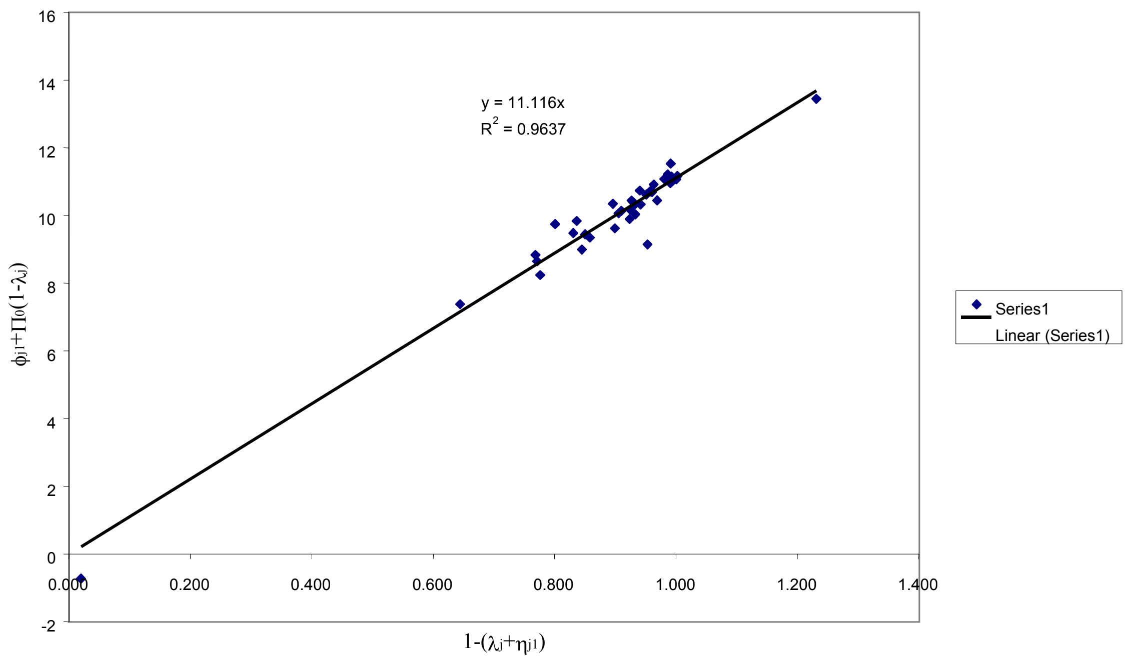

Figure 2: Relationship between slope and intercept parameters: Rural all households:East

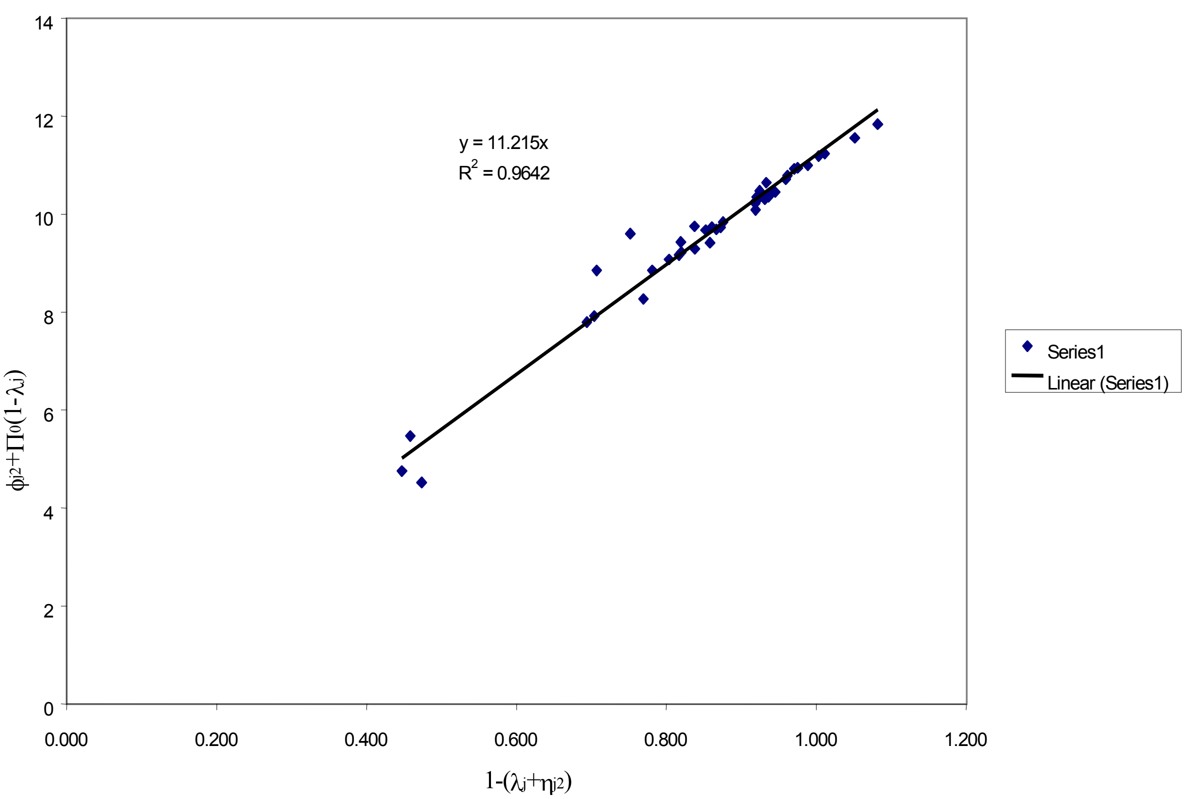

Figure 3: Relationship between slope and intercept parameters: Rural all households: West

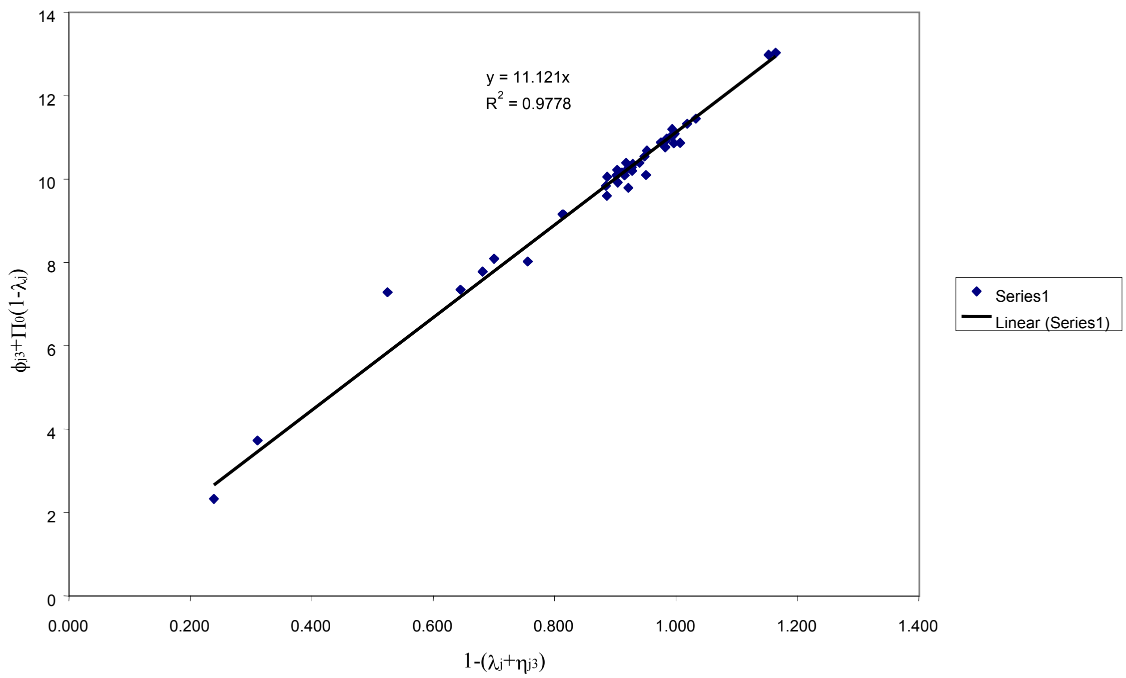

Figure 4: Relationship between slope and intercept parameters: Rural non-poor households: South

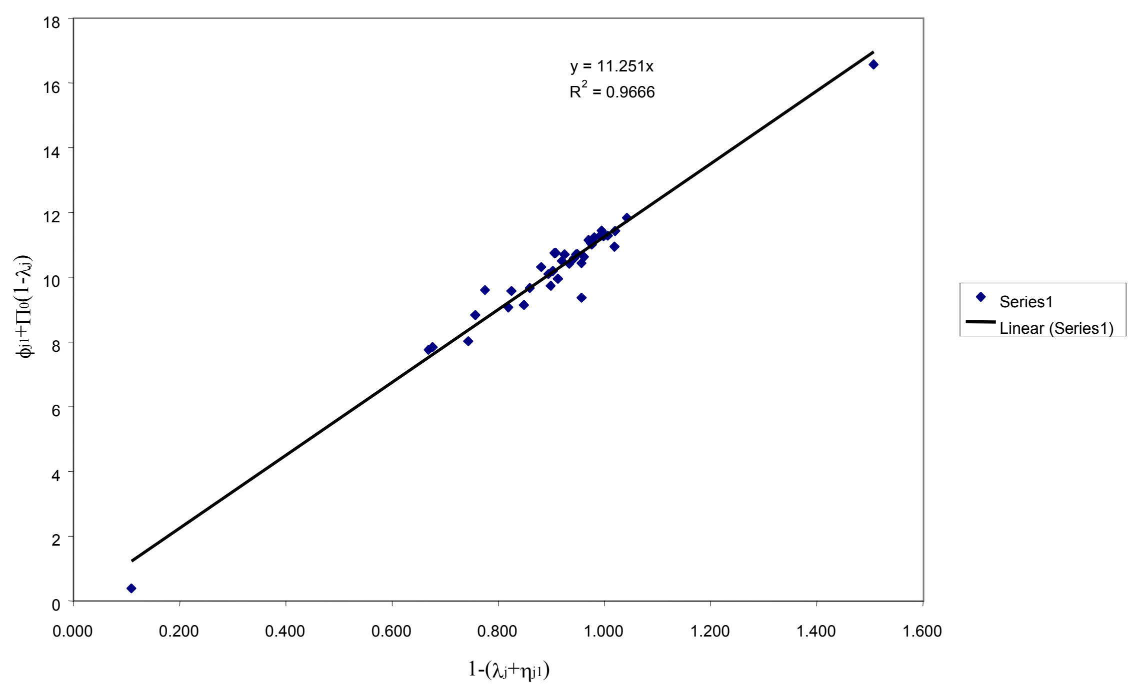

Figure 5: Relationship between slope and intercept parameters: Rural non-poor households: East

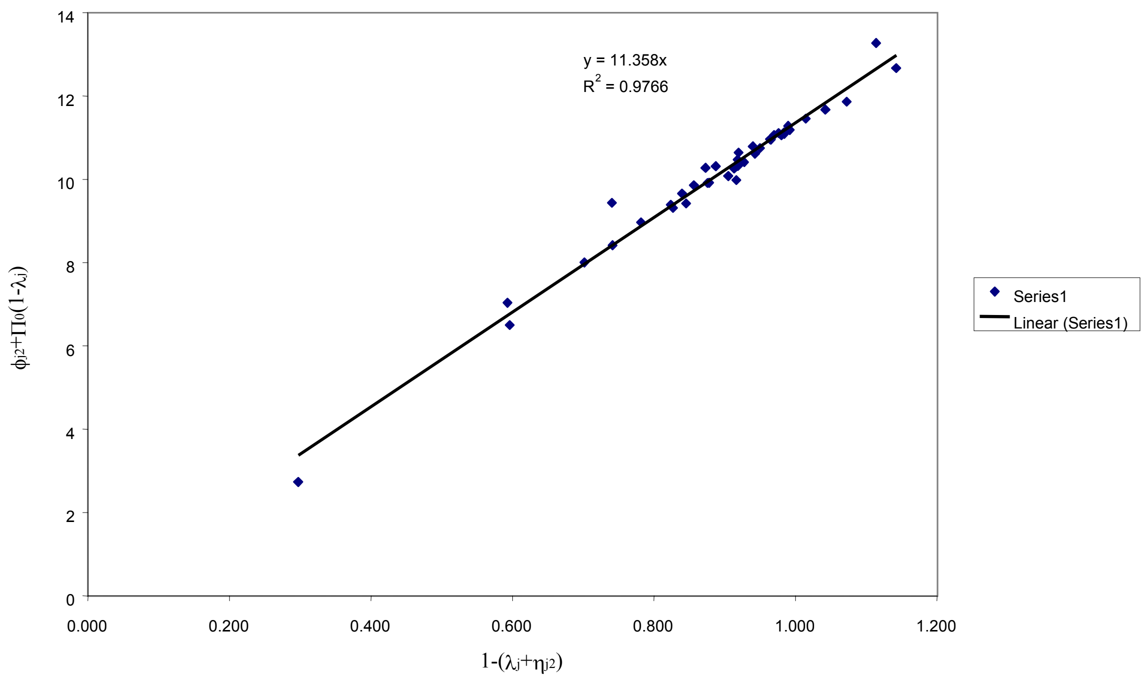

Figure 6: Relationship between slope and intercept parameters: Rural non-poor households: West

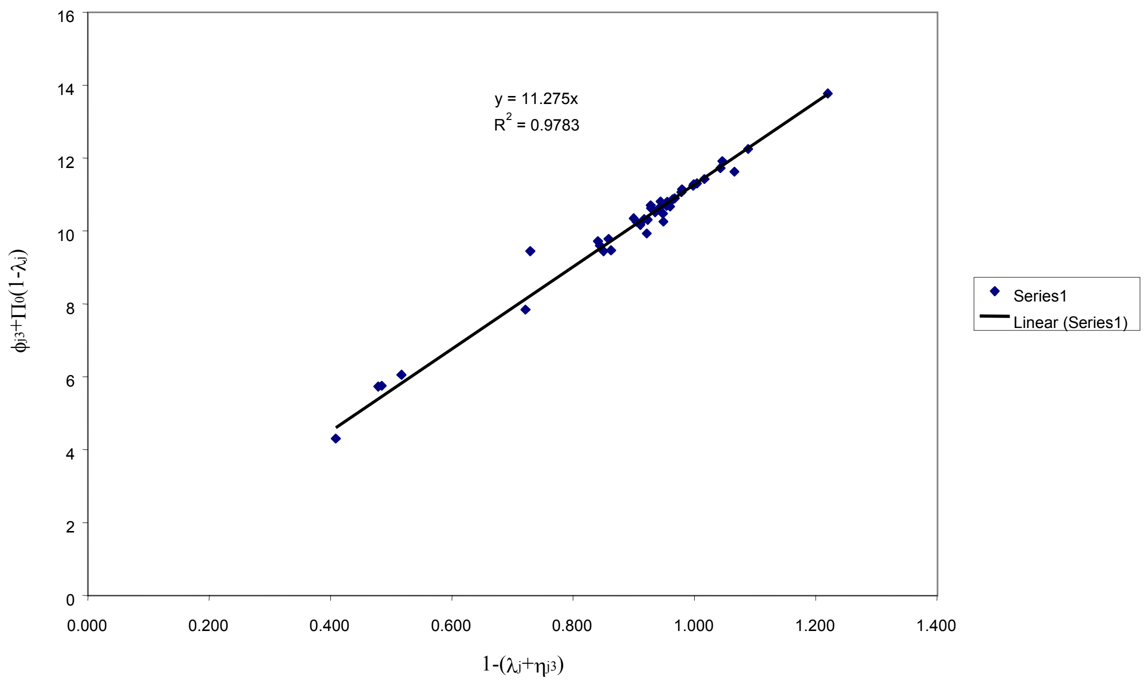

Figure 7 : Relationship between slope and intercept parameters: Rural poor households: South

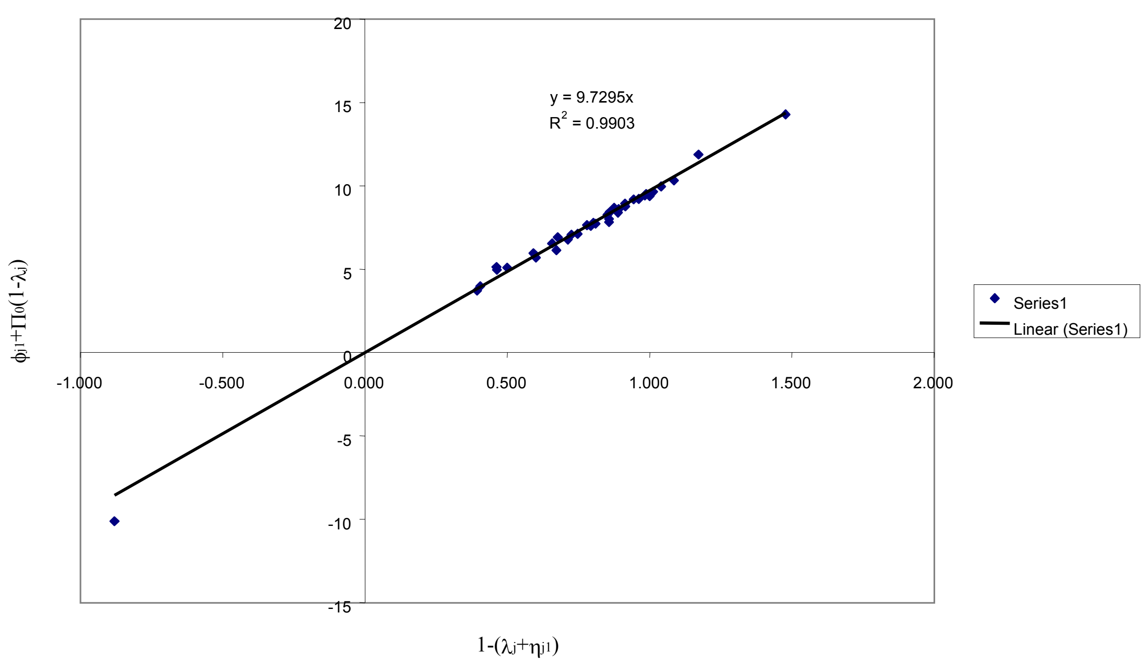

Figure 8: Relationship between slope and intercept parameters: Rural poor households: East

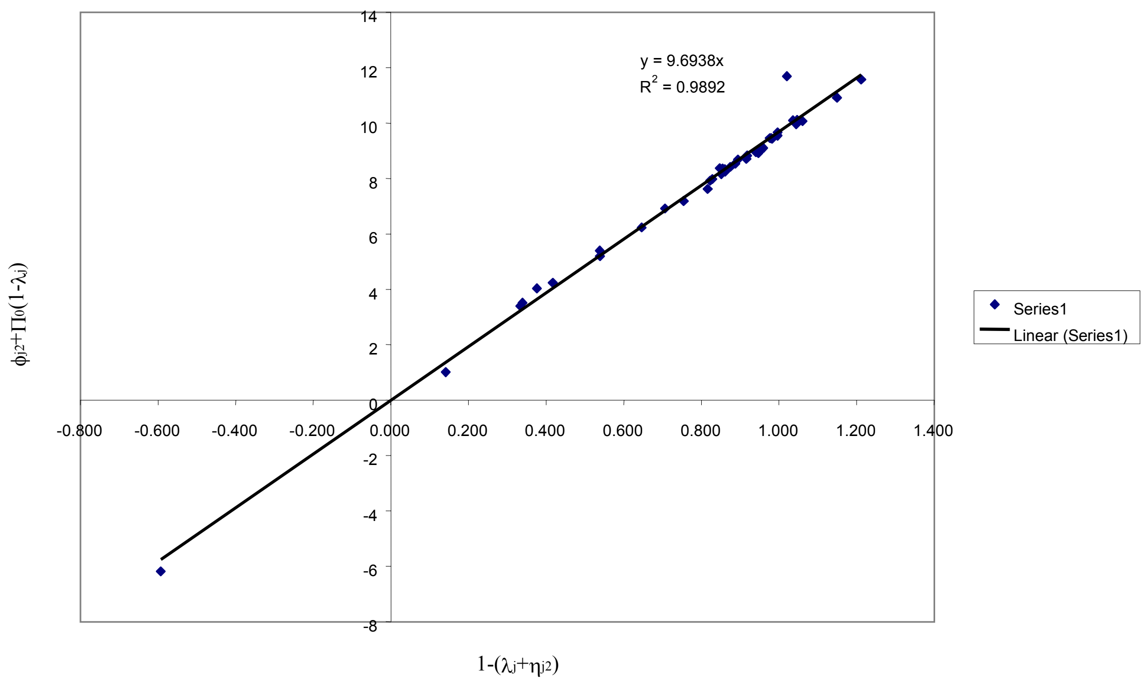

Figure 9: Relationship between slope and intercept parameters: Rural poor households: West

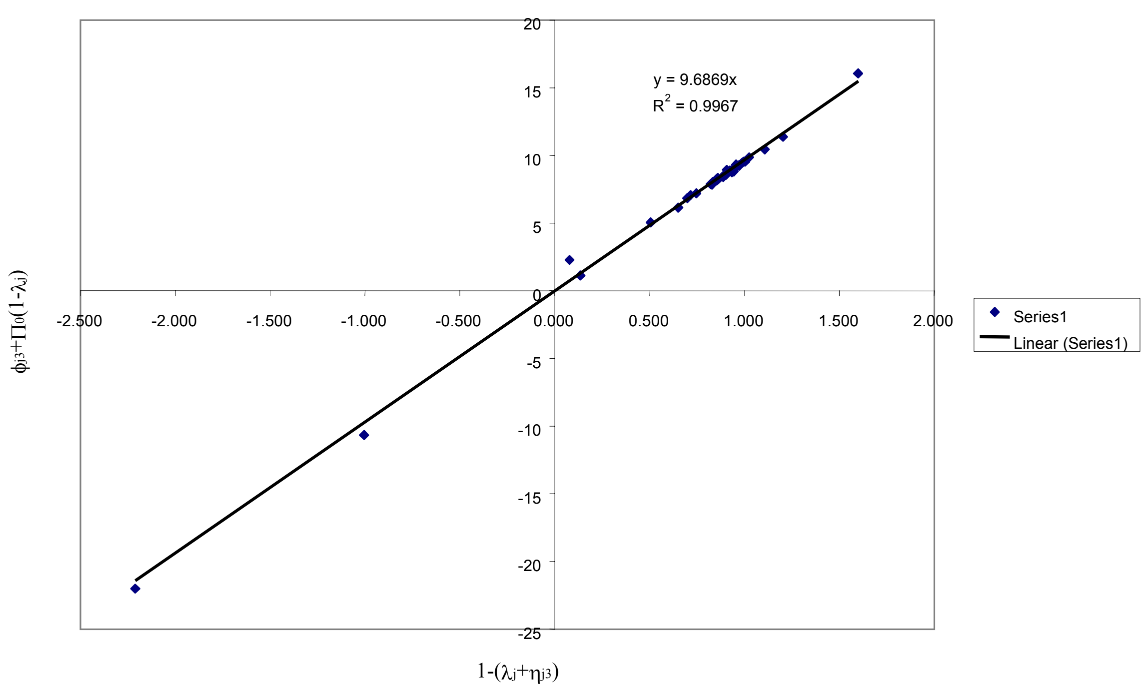

Figure 10: Relationship between slope and intercept:Urban all households:South

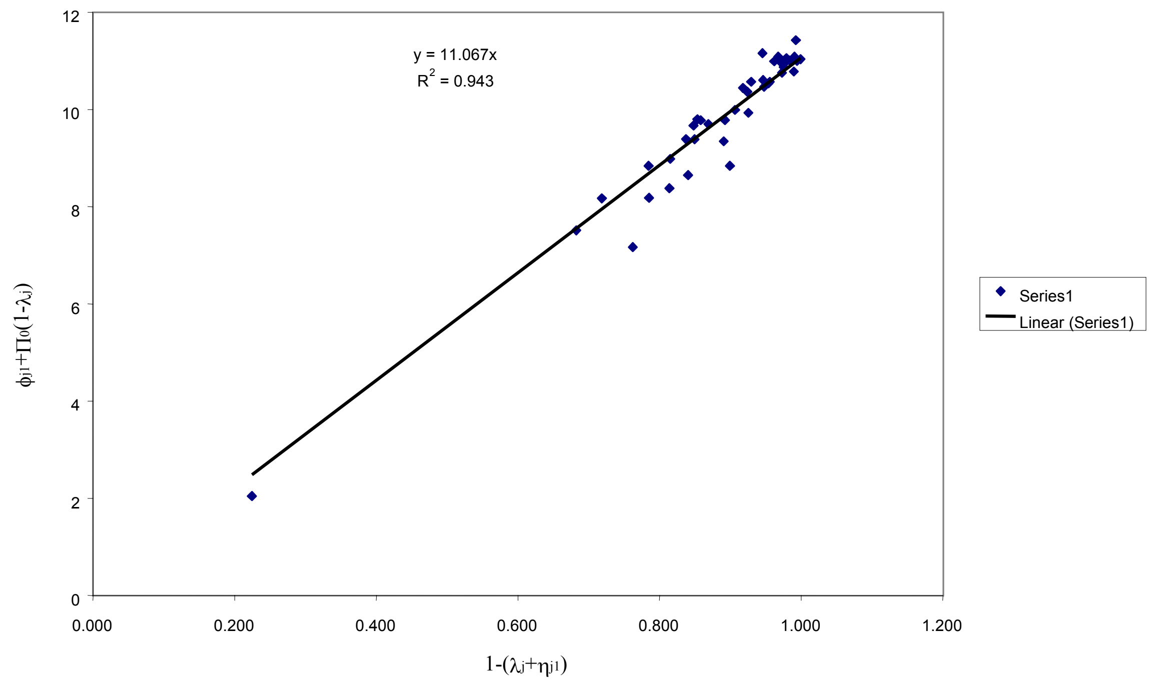

Figure 11: Relationship between slope and intercept:urban all households:East

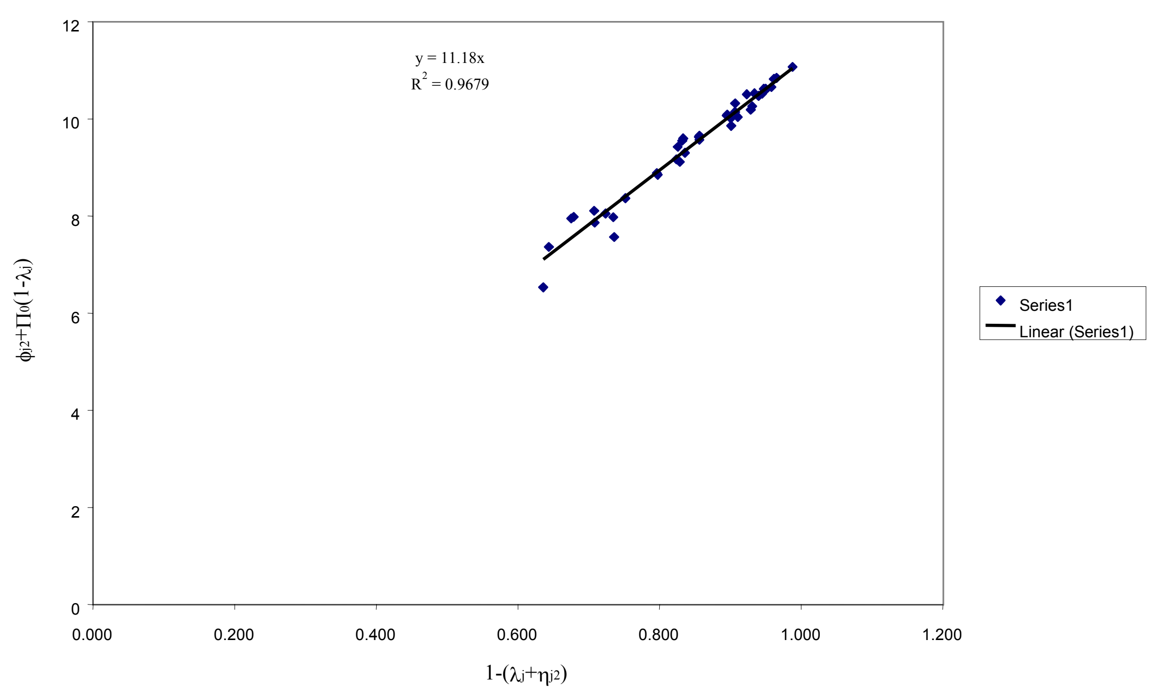

Figure 12: Relationship between slope and intercept:Urban all households:West

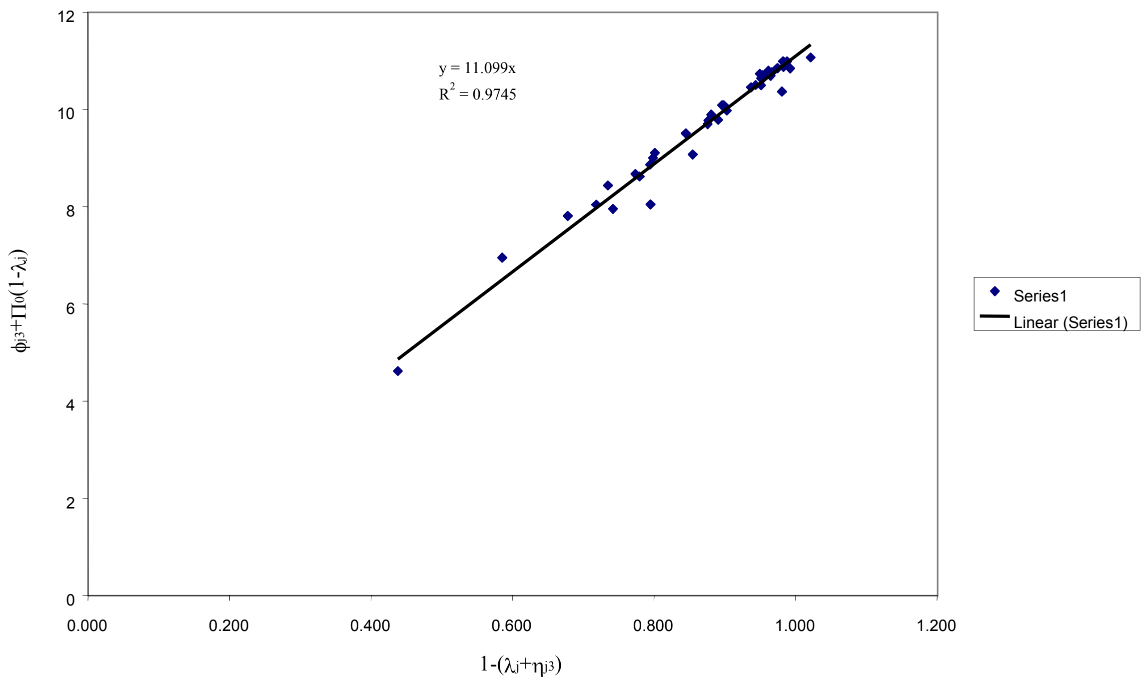

Figure 13: Relationship between slope and intercept: urban non-poor households:South

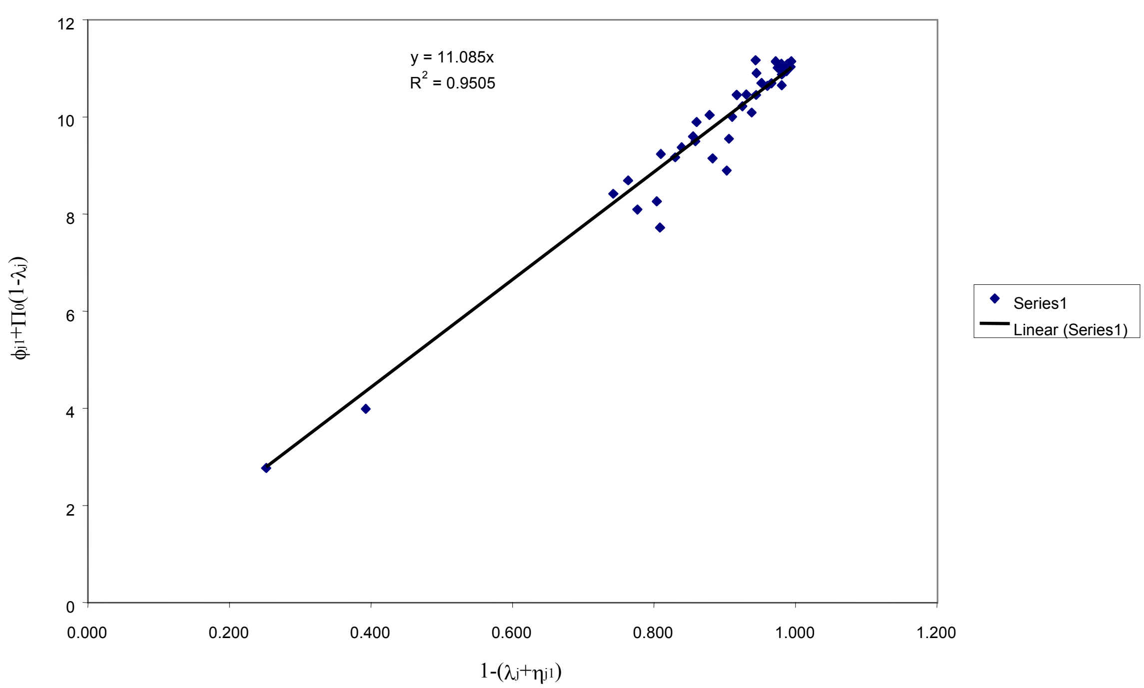

Figure 14: Relationship between slope and intercept:urban non-poor households:East

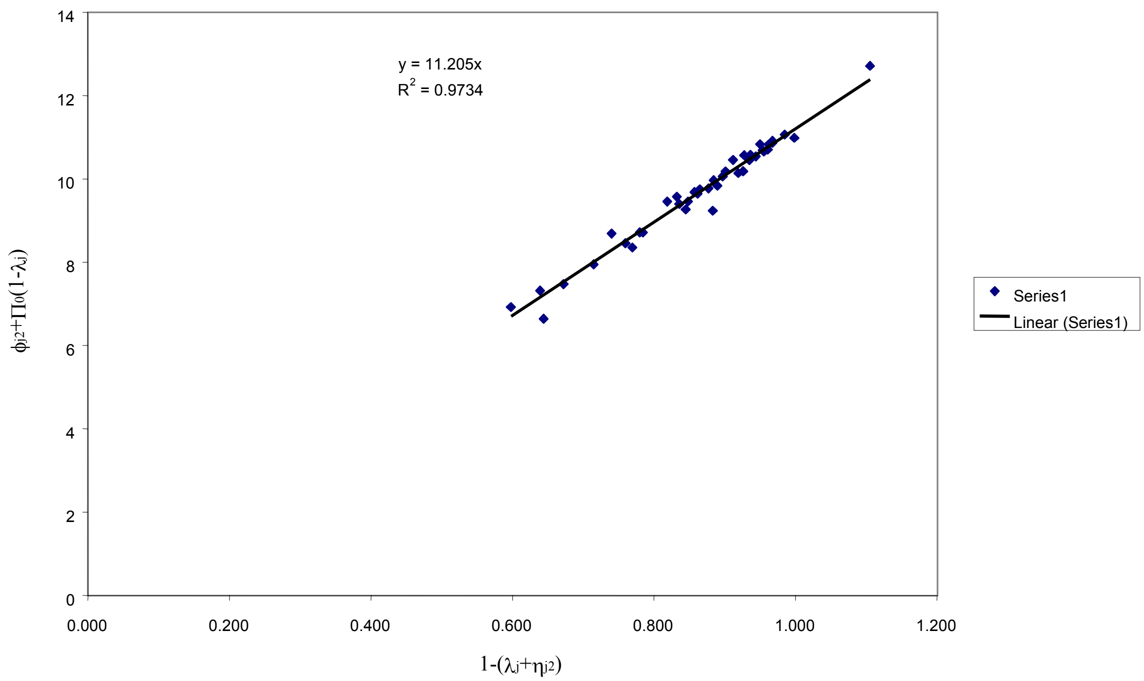

Figure 15: Relationship between slope and intercept:Urban non-poor households:West

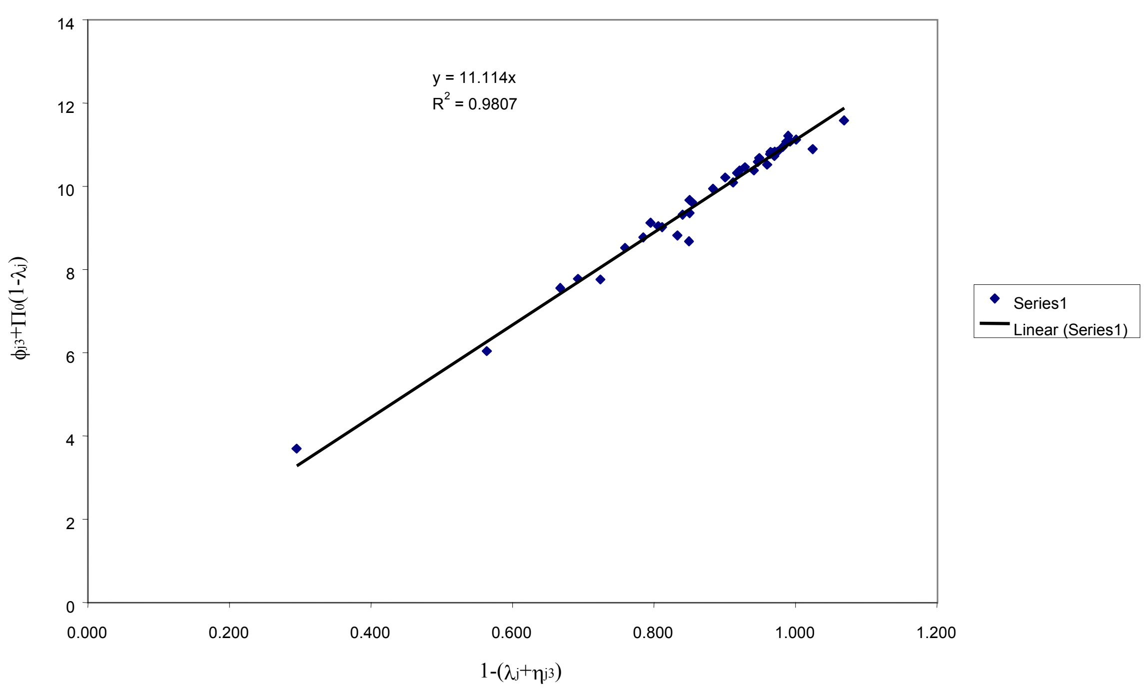

Figure 16: Relationship between slope and intercept:Urban poor households:South

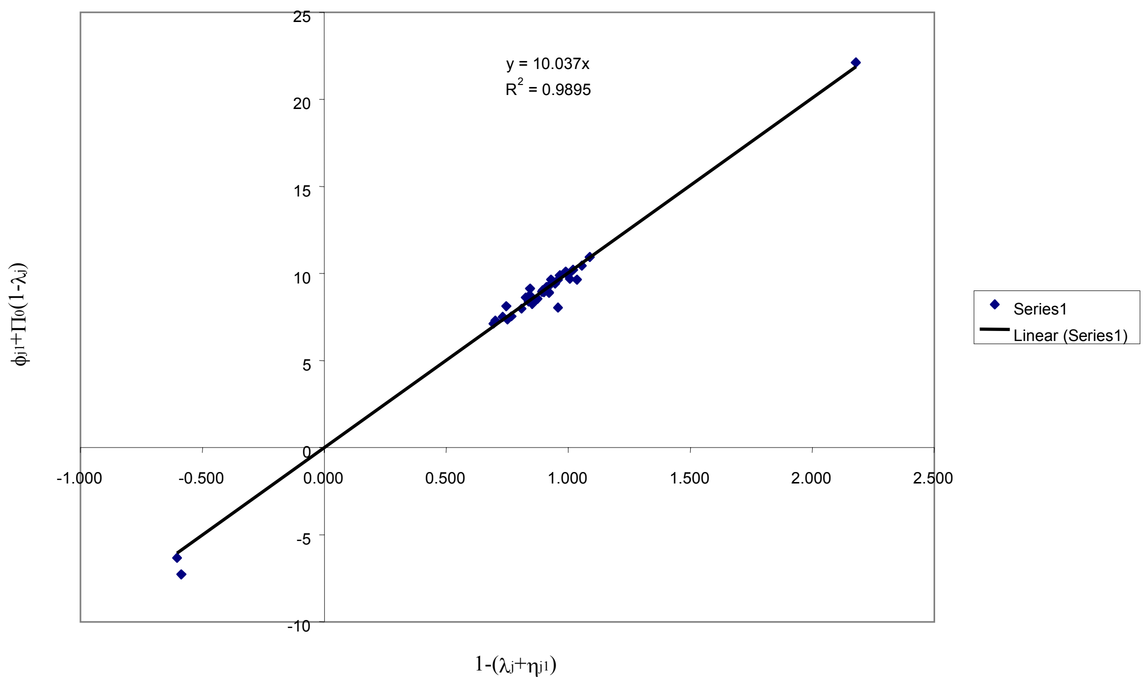

Figure 17: Relationship between slope and intercept:Urban poor households:East

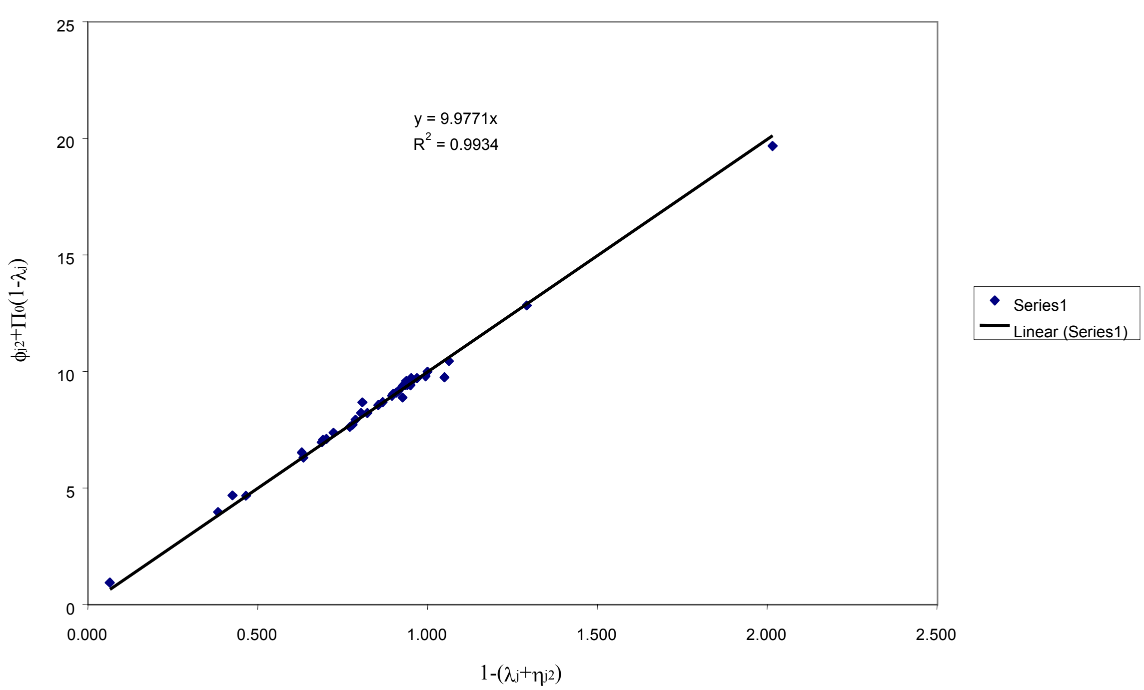

Figure 18: Relationship between slope and intercept:Urban poor households:West