Contructing a Broader Measure of Welfare Incorporating the Access to Public Goods

Felipe Diniz
University of Chicago

July 18, 2005

Abstract

This paper attempts to construct a broader measure of welfare that takes in account the access people have to some public goods. If the data on household access to public goods and private assets is assumed to be the result of a maximization problem, a latent indirect utility level may be estimated by some factor model. In this paper the individual measure of welfare is constructed using Principal Component Analysis (PCA) in the ownership of private assets and the existence of public goods in the neighborhood the agent lives in. The resulting welfare distributions are used in different analysis: Calculate the treatment effect of having access to certain public goods; investigate the effects of public goods in poverty and inequality alleviation; development of an algorithm to locate public goods in order to maximize some social welfare function.

Abstract

No artigo tenta-se construir uma medida para o bem estar social que leve em conta o acesso que os agentes tem a bens públicos. Se os dados domiciliares de consumo de bens duráveis e o local do domicílio são resultado de uma escolha ótima dos agentes, pode-se usar um model de fatores para se estimar o nível latente de bem estar associado a essa escolha. Neste artigo usa-se o o método de Componentes Principais sobre dados de consumo de bens duráveis e informaçoes sobre a existencia de bens públicos na vizinhança onde os domicílios estão localizados para se construir uma medida mais ampla de bem estar. A distribuição de bem estar obtida é utilizada em diferentes exercícios: Cálculo do "Treatmente Effect" do acesso a certos bens públicos; Investigação dos efeitos do acesso a bens publicos sobre medidas de pobreza e desigualdade; Desenvolvimento de um algorítmo para se instalar um novo bem publico de forma a maximizar alguma fução do bem estar social.

Keywords: Bens Públicos, Pobreza, Desigualdade, Bem Estar
Palavras Chaves: Public Goods, Poverty, Inequality, Social Welfare
JEL: H41, I31, I38

1 Motivations

The concept of welfare is related to some weighting to a multiplicity of features that determines the living standards of agents in an economy. In a more concrete way, we could say that the welfare of a society is derived from a bundle of goods and services consumed by its individuals. In Economics the welfare that an individual enjoy is summarized by an abstract measure called utility.

When people do welfare analysis, for example measuring poverty, inequality, they use measures of household income or consumption as a proxy to the unobservable utility distribution.

Unfortunately, in this kind of analysis the utility people derive form the utilization of public goods is not captured. That is, having access to public goods like schools, hospitals, roads etc. has an impact on the utilities of the agents, and a measure of welfare should capture that. So, in order to have better and more robust estimators to poverty and inequality we should try to incorporate the information on the individual access to public goods in our analysis.

Another topic very stressed in the literature is the measurement of the value each individual attribute to a given public good. As public goods are in general financed by taxes, we do not have a clear measure of public good's price. In the literature many strategies have been tried to estimate the individual willingness to pay for a public good: use of surveys that ask households directly about their the valuation of the public good; hedonic methods that assume that the price of land or housing is a function of the amenities of the neighborhood.

The goal of this paper is to construct a measure of the individual welfare that incorporate in some sense the utility derived from the consumption of public goods. We argue that such a measure of welfare would allow us to perform a more accurate welfare analysis. Once this broader measure of welfare is constructed it would be possible to estimate the individual gain in terms of welfare of having access to a given set of public goods. In this sense, the objectives of this paper are the construction of a broader measure of welfare that incorporates in some way the utility derived by the consumption of public goods. After such a measure is created the following exercises may be performed: Analyze the impact of public goods in poverty and inequality dynamics; calculate the "treatment effect" of having access to certain public goods; find the optimal geographical position of a public good according to some social welfare function.

We argue that the development of such techniques may be very useful for policy evaluation. In this sense , it is pointed out here that the paper raises some interesting questions and our hope is that we can provide some good answers.

1.1 Theoretical Background

In order to perform empirical studies about welfare, poverty and inequality, it is necessary to have a welfare indicator for each household. That is, it is necessary to measure the utility of the agents in some way. The way it is usually done in the economic literature is the construction of a money metric indicator of welfare. There are two major money metric indicators used empirically: household income or its consumption expenditure. The use of income (or consumption expenditure) as an approximation of utility is justified as follow: Consumers maximize their utility function subject to some budget constraint. From the consumers' maximization problem we can derive his expenditure function as

$$e(p,u) = Min\ p.x \qquad st\ u(x) > u^*$$

where u stands for the utility level, x is a bundle of consumption goods, and p is its respective price vector.

In this sense, $e(p,u)$ is the minimal cost necessary to obtain some level of utility, for a given price vector p. But from the same maximization problem, we can show that for a given vector of goods x we can find an associated level of utility $u(x)$. Then, we can construct a money metric utility function from the prices and quantities consumed by the agent in the following way

$$m(p,x) = e(p,u(x))$$

Note that $m(.)$ has the same properties that the expenditure function, and for a given price vector it is a monotonic transformation of the utility function: the bigger the $m(.)$ the larger the utility. In this context, income (or consumer expenditure) are justified as a valid welfare indicator.

Note that in the previous analysis no explicitly assumption was made about the role of public goods in the welfare indicator. So if one uses measures of income or expenditure in market goods as proxy of welfare the obtained measure may not be correctly capturing the well-being of the individuals.

Different models used in the literature try to account to public goods in the welfare of agents. If data on the use and the cost of usage of public goods is available, we could then mimic the money metric utility model. In this set up agents consume market goods x and use some public goods g. The utilization of the public good has a cost denoted by d. Then the consumer's expenditure function would be given by

$$e(p,u) = Min\ p.x + d.g \qquad st\ u(x,g) > u^*$$

and the correspondent money metric utility would be

$$m(p,d,x,g) = e(p,d,u(x,g))$$

In this framework it is easy to see that in constructing an empirical proxy of the money metric utility one should try to input the value of the utilization of public goods for each individual.

Another possibility is to measure the welfare derived by the consumption of public goods using an hedonic approach. In this case the assumption is that the prices of some goods like land or housing reflect some quality these goods have. In particular, the price of land (or housing) could be reflecting the proximity to different public goods. So one would expect to observe a higher price of land in a neighborhood where there are plenty of public goods available. On the other hand, everything else constant, the price of the land would be lower in a neighborhood or village located far from public goods. Using such a model it could be possible to capture in a money metric welfare indicator the utility derived by the consumption of public goods. At the same time it could be possible to estimate the value each individual or household attribute to a given set of public goods.

Both the models described above have very appealing qualities, however they require a very detailed data base to allow for their implementation. This paper tries to develop a different model to construct a broader measure of welfare including the access people have to public good. The idea here is to extend the models that try to overcome the absence of expenditure data using information on the ownership of assets by households. Such models as Filmer and Pritchett (1998) rank households using data only on the assets they own. Filmer and Pritchett main assumption is that the ownership of a given asset is a function of a set of parameters and what they call the "long term wealth". In this sense, it would be possible to estimate this latent "long term wealth" in the absence of expenditure data by using some factor model in the asset ownership data. The model developed in the next section tries to justify this approach and allow for a more general set of assets.

2 A Simple Model

Suppose the researcher observes data on the ownership of individual assets. Suppose also that there is data for the existence of public goods in the neighborhood (or village) the household lives in. Let's assume that the bundle of assets we observe is the result of the household's optimal choice. Assume the bundle household h living in village v consumes has the following form

$$X_{iv} = (x_1, ..., x_k, G_v)$$

where $(x_1, ..., x_k)$ are the assets (like TV, VCR, automobile, etc.) and $G_1 = (g_{v1}, ..., g_{vL})$ are the village v characteristics (such as schools, police station, post office etc.). Note that observed bundle X_{iv} is a vector with coordinates that assume values of zeros or ones. In this set up the space of commodities the household can chose within is given by

$$S = \{X_{iv} \in R^{k+L}; X_{iv} \text{ is any possible combination assets } x_1, ..., x_k \text{ and } G_l\}$$

Then we see that in this model the consumer chooses the preferred bundle in a set of $M = 2^K L$ possible bundles.

The main assumption here is that the household is choosing the bundle X_{iv} in order to maximize his welfare. That is, given a set of parameters representing prices and preferences the household is choosing optimally his portfolio of assets and the village he is living in. So let $Y_i \equiv V_i(X_{iv}; \theta)$ be the indirect utility of consuming the bundle X_{iv}, where θ is a vector of parameters representing prices and preferences. Then household i will chose the bundle X^*_{iv} such that

$$V_i(X^*_{iv}; \theta) \geq Max\{V_i(X^*_{1v}; \theta), ..., V_i(X^*_{MV}; \theta)\}$$

Our problem then is to estimate the latent level of utility $Y^*_i = V_i(X^*_{iv}; \theta)$. But as stated before we just have data on X^*_{iv}. So assuming a linear relationship between that Y^*_i and X^*_{iv} we obtain $Y^*_i = X^*_{iv}\theta$, and then we can use Principal Component Analysis (PCA) or Factor Analysis to estimate Y^*_i. In the present paper due to time restrictions just PCA was employed,but the same kind of analysis could be carried on using a more general factor model.

2.1 PCA

In this section we try to justify the use of the PCA as a reasonable technique to be employed in assets and public goods index construction. Principal Component analysis is a useful statistic technique that allows us to "extract from a large number of variables those few orthogonal linear combinations of the variables that best capture the common information" . The idea is, as we have data on the household ownership of a set of assets, and we also have information about the existence of some public goods in the village where each household is located, we can construct a welfare index using the PCA. Suppose we have data on k assets households hold. Then $X' = [X'_1, ...X'_k]$, where the n-dimensional vector X'_k represents the the ownership of asset k by the households in the population. The main assumption here is that the ownership of the assets is a function of a set of parameters and the vector of orthogonal latent variables $Y' = [Y'_1, .., Y'_k]$. That is

$$X = f(Y; \phi) \tag{1}$$

and

$$Cov(Y_i, Y_j) = 0 \quad i \neq j$$

In particular we assume that the latent variable Y_1, that accounts for the greater source of variance in X, is a proxy for the long term welfare households experience. In this context if we are interested in measuring the latent variables we would like to find

$$Y = V(X, \theta) \tag{2}$$

Let' assume that the relation between X and Y is the following

$$X = \Phi Y$$

where Φ is a (kxk) non singular matrix. Then we can write

$$Y = \Theta X$$

where $\Theta = \Phi^{-1}$. The same expression can be written as a system of linear equations

$$\left\{ \begin{array}{c} Y_1 = \theta_{11} X1 + ... + \theta_{k1} X_k = \theta_1' X \\ ... \\ Y_p = \theta_{1k} X1 + ... + \theta_{kk} X_k = \theta_k' X \end{array} \right.$$

Assuming without loss of generality (WLOG) that X has mean zero we define

$$\Sigma = E[XX'] = \left[\begin{array}{ccc} \sigma_{11} & ... & \sigma_{1k} \\ ... & ... & ... \\ \sigma_{k1} & ... & \sigma_{k1} \end{array} \right]$$

So given this setup we want to find the matrix Θ such that

$$\left\{ \begin{array}{c} Var(Y_j) = \theta_j \Sigma \theta_j' = \lambda_j \\ Cov(Y_i, Y_j) = \theta_i \Sigma \theta_j' = 0 \\ \theta_j \theta_j' = 1 \end{array} \right.$$

where the last restriction is necessary to guarantee that the variance of Yj is not arbitrarily big. Note that without this last restriction one could make the variance of Y_j arbitrarily big by multiplying θ_j by a constant.

Claim 1 *Let $\Sigma = E[XX']$. Let $(\lambda_1, e_1), ...(\lambda_k, e_k)$ be the eigenvalue and eigenvector pairs associated with Σ, and $\lambda_1 \geq ... \geq \lambda_k$. Then*

$$\left\{ \begin{array}{c} Var(Y_j) = e_j' \Sigma e_j = \lambda_j \\ Cov(Y_i, Y_j) = e_i' \Sigma e_j = 0 \end{array} \right.$$

and

$$Y_j = e_{1j} X1 + ... + e_{k1j} X_k = e_j' X$$

Proof. Let $(\lambda 1, e1), ...(\lambda p, ep)$ be the eigenvalue and eigenvector pairs associated with Σ. Then for all j we have by definition

$$\Sigma e_j = \lambda e_j \Rightarrow e_j' \Sigma e_j = \lambda e_j' e_j = \lambda_j$$

and

$$Cov(Y_i, Y_j) = e_i' \Sigma e_j = 0$$

■

Corollary 2 *From the previous result we have*

$$V \equiv \sigma_{11} + ... + \sigma_{kk} = \sum_{i=i}^{k} Var(X_i) = \lambda_1 + ... + \lambda_k = \sum_{i=i}^{k} Var(Y_i)$$

Proof. It was shown that $Var(Xi) = \sigma_{ii}$. Then $\sum_{i=i}^{k} Var(X_i) = \sigma_{11} + ... + \sigma_{kk}$. At the same time

$$Var(Y_j) = e_j' \Sigma e_j = \sum_{p}^{i=1} \sum_{k=1}^{p} e_{ij} e_{kj} \sigma_{ik} = e_{1j}^2 \sigma_{11} + ... + 2 e_{1j} e_{kj} \sigma_{kj} + ... + e_{kj}^2 \sigma_{kk}$$

But as $e_i'e_j = 1$ and $e_i'e_k = 0$, as we calculate the total variance for Y we find

$$\sum_{i=i}^{p} Var(Y_i) = \sum_{j=1}^{p} e_{1j}^2\sigma_{11} + ... + 2e_{1j}e_{kj}\sigma_{kj} + ... + e_{kj}^2\sigma_{kk}$$

$$\begin{aligned}\sum_{i=i}^{p} Var(Y_i) &= \left(e_{11}^2 + ... + e_{1k}^2\right)\sigma_{11} + ... + 2\left(e_{11}e_{k1} + ... + e_{k1}e_{kk}\right)\sigma_{p1} \\ &\quad + ... + \left(e_{k1}^2 + ... + e_{kk}^2\right)\sigma_{kk}\end{aligned}$$

$$\sum_{i=i}^{p} Var(Y_i) = (1)\,\sigma_{11} + ... + 2\,(0)\,\sigma_{k1} + ... + (1)\,\sigma_{kk} = \sigma_{11} + ... + \sigma_{kk}$$

■

So we see that the first component accounts for the biggest share of the variance of the population $\frac{\lambda_1}{V}$. In this sense we can reinterpret the PCA problem. We can say that the principal component is the single linear combination of the assets $[X_1', ...X_k']$ that accounts for the biggest share of the variance of the population.

Finally, in matricial terms we have

$$\Lambda = \begin{bmatrix} \lambda_1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & \lambda_k \end{bmatrix}$$

and

$$\Theta = \begin{bmatrix} e_{11} & ... & e_{k1} \\ ... & ... & ... \\ e_{1k} & ... & e_{kk} \end{bmatrix}$$

with $\Theta'\Theta = \Theta\Theta' = I$. Then we have

$$Y = \Theta X \Leftrightarrow X = \Theta' Y$$

So if we get back to our model we see that the latent utility level is given by

$$Y_i^* = V_i(X_{iv}^*; \theta) = X_{iv}^*\theta$$

where $X_{iv} = (x_{i1}, ..., x_{ik}, g_{v1}, ..., g_{vL})$. So using the PCA approach described above we construct a welfare index such that

$$Y_{iv} = b_1 x_{i1} + ... + b_k x_{ik} + \gamma_1 g_{v1} + ... + \gamma_l g_{vl}$$

where $\theta = (b, \gamma)$

Note that in this simple model it was assumed that the household is choosing optimally its assets and the public goods it consumes. Then the latent utility level given by $Y_i^* = X_{iv}^*\theta$ is estimated as linear combination of assets that accounts for the biggest share of the variance in the data.

2.2 Exercises

Once we have calculated the welfare measure for each individual we will use this distribution to do different analysis. First we can use the welfare distribution to calculate poverty and inequality index. Using the model described above we can compare poverty and inequality index if we account or not for public goods. In this sense, we can estimate the impact of a given public good in some aggregate welfare measure. Other possible exercise is to decompose poverty and inequality by village and by access to public goods. Following this approach we could be able to identify if the existence or the lack of a given set of public goods is a source of poverty and inequality.

This model also allows us to calculate the treatment effect of having access to public goods. The idea is that we could model the choice of the village the household is located as a discrete choice problem. In this context, if we observe a household that is located in a village not endowed with public good k we could try to identify the welfare gain this household obtains from having this public good in his village. To do this we need to find the counterfactual utility level for this household. That is, we have to estimate the utility the household would have if in his village there were public good k.

Finally, as we constructed a welfare distribution that is a function of the public goods villages have, we can construct counterfactual distributions of welfare by placing a given public good in a village that does not have this public good. For each of these counterfactual distributions we can calculate the associated aggregate welfare measured by a social welfare function. Following this method we chose the village to place an extra public good in order to maximize the aggregate welfare.

2.2.1 Welfare Analysis

1. Poverty: For a given poverty line z we can calculate different measures of poverty using the formula shown bellow. High values of α imply that the poverty gaps on the left tail of the welfare distributions are more important.

$$FGT(\alpha) = P_\alpha = \frac{1}{N}\sum_{i=1}^{N}\left[Max\ \left\{1-\frac{y_i}{z},0\right\}\right]^\alpha$$

   We can also decompose the total measure of poverty in a weight sum of the contribution of poverty of different groups

$$P_\alpha = \frac{1}{N}\sum_{i=1}^{k} n_k P_\alpha^k$$

2. Inequality: Using the data we constructed on welfare Y_i we can also construct different inequality measures. The Generalized Entrophy is given by

$$GE(\alpha) = \frac{1}{\alpha 2-\alpha}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\overline{y}}\right)^\alpha - 1\right]$$

   where the greater the parameter α the bigger is the weight given to income differences in the inferior tail of the distribution. The Gini coeficient is given by

$$Gini = \frac{1}{2N^2\overline{y}}\sum_{i=1}^{N}\sum_{j=1}^{N}|y_i - y_j|$$

Using our estimated distribution of welfare we are going to decompose the inequality. The formula for the inequality decomposition for a given inequality measure I is

$$I = WI + BI$$

where the within-group inequality WI is the sum of subgroup inequality levels (weighted by the population shares of subgroups). BI stands for the inequality between groups.

3. Finally, can test the stochastic dominance for different distributions of welfare. That is, we can rank distributions of welfare that take or not in account individual access to public goods.

2.2.2 Calculate "The Treatment Effect"

Using the welfare distribution we will calculate the "treatment effect" of accessing a public good. Following the treatment effect literature we could write the Average Treatment effect of having access to a given public good as

$$ATE(X, g_k) = E[Y_{iv} - Y_{i\tilde{v}}]$$

where Y_{iv} is the outcome of someone that has access the public good g_k, and $Y_{i\tilde{v}}$ is the outcome of an agent that has no access to the public good g_k. In this multiple outcome discrete model the expression for the treatment on the treated is given by

$$TT(X, g_k) = E[Y_{iv} - Y_{i\tilde{v}}/D_k = 1]$$

where $D_k = 1$ implies that treatment k was chosen. The usual problem in this literature is that we do not observe the same household in two different states, or with two different treatments. In this particular model we are going to use two different approaches to try to identify the treatment effect. The first one is Matching. The idea here is that if we want to measure the treatment effect of public good k to household i, we could use as a counterfactual an household that "looks like" household i. In this case we would compare the welfare of households that are very similar in the ownership of private assets and have access to the same public goods but one. As the index of welfare given by our model is

$$Y_{iv} = b_1 x_{i1} + ... + b_k x_{ik} + \gamma_1 g_{v1} + ... + \gamma_l g_{vl}$$

theTreatment on the Treated (TT) would be given by

$$TT(X_i, g_k) = \gamma_k g_{ik}$$

If we are interested in the in the effect for all the population, we just have to sum this for all the individuals

$$TT(X, g_k) = \sum_i \gamma_k g_{ik}$$

Other strategy that can be used to calculate the treatment effect of a public good is the directly construction of the counterfactual distribution of welfare assuming a different distribution of public goods. That is, we generate a fake data set assuming that that village v has the public good k (in the real data village v has no public good k). Then we run PCA in this fake data set and calculate the welfare index for each individual. So bu simply inputting a counterfactual distribution of the public good we calculate the resulting counterfactual distribution of welfare

$$Y_i^{vC} = b_1^c X_{1i} + ... + b_n^c X_{ni} + \gamma_1^c g_{1v}^c + ... + \gamma_k^c g_{kv}^c$$

Then, given the original data on assets and public good and the fake data on public good we can calcutate the TT in the following way

$$TT(X_i, g_v, g_v^c) = Y_i^v - Y_i^{vC}$$

We can aslo aggregate this treatement effect in differente ways. Aggreagating for the entire population we get

$$TT(X, g_v, g_v^c) = \sum_i \left[Y_i^v - Y_i^{vC}\right] \equiv \Psi^{vc}$$

We could also aggregate the treatment effect by village.

2.2.3 Placing The Public Good

Take a Social Welfare Function (SWF) that aggregates the welfare of all the indivudals in the economy.

$$W = \frac{1}{n}\sum_{1=1}^{n} f\left(Y_i\right)$$

where $f(.)$ is some concave function. We we could also define our function in terms of the villages.

$$W^v = \frac{1}{n_v}\sum_{1=1}^{n_v} f\left(Y_i^v\right)$$

Then the welfare in the society would be defined as an aggregation of the welfare in each village

$$\bar{W} = \frac{1}{V}\sum_{v=1}^{V} g\left(W^v\right) = \frac{1}{V}\sum_{v=1}^{V} g\left(\frac{1}{n_v}\sum_{1=1}^{n_v} f\left(Y_i^v\right)\right)$$

Note that the choice of an especifc functional form for the SWF will be always arbitrary. This choice will depend on the society's degree of the aversion to inquality. That is, the more concave the functions $f()$ and $g(.)$ the biggest will be the lost of welfare due to inequality.

Once we have a SWF we are able to write an algorithm to place the public good in the following way:

Suppose we want tho chose the location of the public good g_k. Note that some villages have already access to this public good. So WLOG assume that our problem is choose a village in the sub set of $V^* = \{v_1, ...v_L\}$ where there is no public good g_k available. Then the algorithm is given by the following steps

1. Place the public good g_k in village 1. Run PCA for the entire economy.
2. Find the new contrafacutual welfare distribution
3. Using the resulting contrafacutual welfare distribution calculate the $\bar{W}(1)$, that is the value of the SWF if the public good is placed in village 1
4. Repeat steps 1,2 and 3 for all villages
5. Chose to place the public good in the village where the SWF values is higher, that is

$$\text{is such that } W(v^*\ ) = Max\{W(v_1), ...W(v_L)\}$$

The algorithm developed above chooses to place an extra public good k in village v^* because the welfare of the society, measured by an aggregation of Y_{iv} will be maximized if the extra public good is placed there. But note that in doing that we are not taking in account any information about the spatial distributions of the pubic goods. We can then add more structure to this problem to obtain more interesting results.

Suppose when placing public goods in villages we also want to take in account geographic conditions. So, we could take in account variables like the number of people would be able to use the public good; distances from other villages, cost of transportation between villages etc. In this sense, we would be assuming that the "real" social welfare of placing a public good in village v_l would be a function of the the social welfare $W(v_l)$ described above, and also a function of some geographical conditions $Z(v_l)$. Then the aggregate welfare of placing public good k in village v would be giving by

$$F(v_l) = F(Z(v_l), W(v_l))$$

In this context, we would chose the village to place the public good in order to maximize this function, that is,

$$v^* \text{ is such that } F(v^*) = Max\{F(v_1), ...F(v_L)\}$$

It should be note that the biggest problem in implementing this algorithm is to choose a reasonable functional form to the function $F(.)$ and pick sensible variables Z.

3 Data

In order to implement the methods described above we are going to use data on household and villages in Thailand. Three data bases were combined in this effort to incorporate the consumption of public goods in the household welfare indicator: CDD, Townsend Thai Data and Felkner Distances Data.

The Community Development Dept (CDD) data set was collected by Rural Development Committee (RDC) at the village level. CDD is a section of the Thai Ministry of the Interior, and it is responsible for the health and well being of Thai citizens. This data set consists of several questions answered by CDD officials and village headmen concerning household characteristics, education, health and sanitation, the environment and public equipment. This 1996 data base was used to access the existence of public goods as schools, health care center, and police station in each village

The Townsend Thai is a very detailed household level survey. The data was collected in two provinces in the Northeast part of the country and two provinces in the more industrialized central corridor, around Bangkok. In each of those provinces, 48 villages were sampled, and in each village 15 households were surveyed. The sample excludes urban households. The questionnaire includes detailed information of household characteristics such as sources of income and expenditure, education, asset ownership, occupation etc. This 1997 survey was used to construct measures of household welfare using asset ownership

The so called Felkner Distances data consists in the geo-location of the villages in the space. This data also contains information on distances (measured in a straight line, and timed in minutes when using the road system) from the villages to the district centers. At the district centers we can assume that there are a set of public goods such as schools, hospitals and so on. Using this data it is also possible to calculate the distances any household is from any public good. So this data base is used to construct a measure of the cost of accessing public goods as detailed above.

The process of combining the data sets is the following: For each household in Townsend Thai data base we have the information about the village he lives in. Using this information we can attribute to each household a bundle of public goods consumed, or measures of distances to the different public goods (these distances will be zero if the public goods are located in the village the household is located in). In the resulting data base we have for each household data on his consumption, education, asset ownership, public goods in his villages, and distances to public goods and so on.

4 Some Empirical Results

The first problem we face when we try to construct an individual welfare measure based on private assets and public goods the choice of assets and the public goods that will compose the index. The problem is that we don't have a model to base our choice: the choice will be arbitrary. It is important to notice that following our model, Yiv is the latent indirect utility of household i locate at village v. And it is estimated by PCA as the single component of the matrix of assets and public goods that accounts for the biggest share of the variance of data. So, we have a trade off when choosing the set of assets and public goods used in the model. In one hand ,we want a sufficiently big number of "goods" (assets and public goods) in order to have representative proxy of the individual welfare. But on the other hand we want that the welfare index accounts for a reasonable share of the variance of the data. But as the number of goods is increased the share of variance of the index falls. So our strategy here was to try different combinations of "goods". The final combination was chosen based on both the variance share criterion and also in the economic appealing of the "goods". In the appendix we show the different models we tried.

The final choice assets and public goods used through the paper is given in the table bellow.

List of Assets and Public Goods

Assets	Public Goods
TV - Color	Community public health service
VCR	Basic medical treatment and reference
Air Conditioner	Kindergarten school
Regular Telephone	Primary school within 3km distance
Cellular Telephone	Secondary school
Refrigerator	Vocation training center
Motorcycle	Postal service center
Car	Telephone service center
Pick-up Truck	Police station
Long Tail Boat w/ Motor	
Washing Machine	
Large Fishing Boat	
Gas Stove	
Bicycle	
Stero	

4.1 Welfare Analysis

The goal of this section is to present the results of the welfare analysis detailed above for both the so called "Asset Index" and for the so called "Welfare Index" that augments the asset index by the inclusion of public goods.

Once the asset and welfare indexes have been constructed, we can start comparing them. Our first problem is that both indexes have mean zero by construction. As we are interested in measuring inequality we need a distribution with positive support. To overcome this problem we just shift both distributions to the right by adding to each component the minimum value of its respective distribution. Note also that both index have no specific unit of measure. So from now on we assume that that both index are measured in utility units.

Inequality in Both Models

Inequality Measures	Asset Index	Welfare Index
Relative mean deviation	0.314	0.270
Coefficient of variation	0.803	0.689
Standard deviation of logs	1.253	0.925
Gini coefficient	0.437	0.376
Mehran measure	0.616	0.533
Piesch measure	0.347	0.298
Kakwani measure	0.175	0.129
Theil index (GE(a), a = 1)	0.337	0.235
Mean Log Deviation (GE(a), a = 0)	0.371	0.312
Entropy index (GE(a), a = -1)	2.043	0.850
(Coeff.Var. squared) (GE(a), a = 2)	0.322	0.237

As we start to perform some welfare analysis with the two distributions, it is easy to see that the "welfare distribution" displays less inequality than the "asset distribution" for many different inequality measures. From Figure 1 we can see for example, that the Gini coefficient associated with the asset index is 0.43, much higher than the 0.37 calculated for the welfare index.

Note that this moderate fall in inequality in the "living standards" distribution when we add the access to public goods could be driven by the methodology applied. Remember that to construct the "welfare" index we are combining household level data with village level data. That is, for each household we have data on its ownership of TV, VCR, motorcycle and so on. And for each village we have data on the existence of schools, police station, post office etc. Then, to all the households located in the same village we are attributing the same access to public goods. This could be generating the observed decrease in inequality. In order to take this problem into account we ranked all households by his relative position in its village and compared inequality levels. Therefore, we calculated the Gini coefficient among the "ranked one" households in each village, and repeated the same procedure to those "ranked two", "ranked three" and so on. In doing that we cannot attribute the decrease in inequality to the fact we are assigning the same level of public good access to each household in its respective village. The results presented in the table 2 show that inequality in the asset index is greater for each household rank. Another important feature of Figure 2 is that as we account for the public good access, inequality falls more among the better ranked households in each village. Those results seem to imply that the access to public goods leads to a reduction in the inequality level.

Using the asset and welfare indexes we constructed we can also calculate poverty measures. We can show that for different poverty lines the number of poor people in the asset distribution is greater than in the welfare distribution. The key feature of measuring poverty in those two distributions is not to say that the level of poverty in rural areas in Thailand is over estimated when one uses a living standard distribution that doesn't account for public goods. In contrast, the point here is to call attention to the fact that without a living standard distribution that takes in account public goods access, one cannot recognize the effects of public good provision in terms of poverty alleviation. Table 3 allows us to see that using a distribution such the welfare index could improve our capacity to analyze public policies. Another interesting feature shown in table 3 is that the effect of the public goods seem to bigger to the poorer individuals. We see that the difference in the in the poverty index is $FGT(\alpha)$ in the assets and welfare index is increasing with α. This fact suggests that welfare distribution displays less poverty gaps in the

Gine Coefficients - Asset Index vs Welfare Index

Household Rank in its Village	Gini Coeficient	
	Asset Index	Welfare Index
1	0.475	0.323
3	0.476	0.353
5	0.312	0.276
7	0.303	0.266
9	0.282	0.254
11	0.260	0.247
13	0.237	0.226

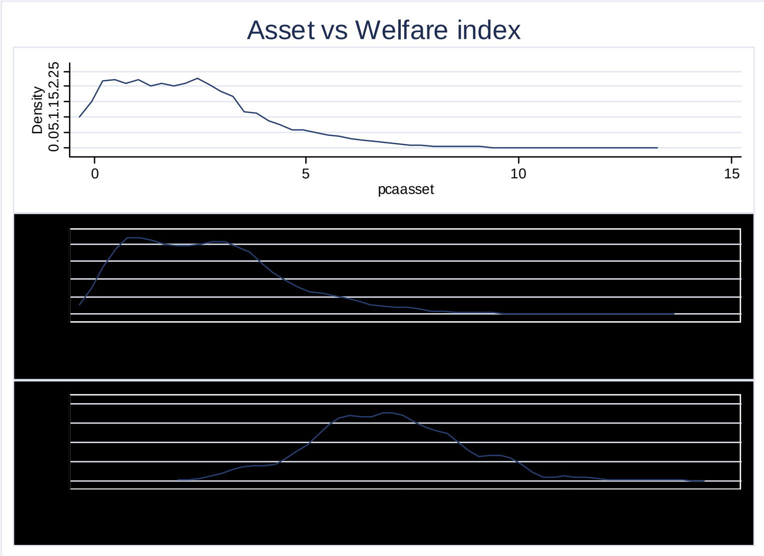
left tail for when we account public goods.

Poverty in Both Models

Poverty Line	FGT(0)		FGT(1)		FGT(2)	
	Asset	Welfare	Asset	Welfare	Asset	Welfare
Z=Mean	0.578	0.552	0.314	0.270	0.236	0.176
Z=0.8*Mean	0.434	0.439	0.268	0.216	0.205	0.138
Z=0.6*Mean	0.369	0.326	0.218	0.158	0.173	0.100
Z=0.5*Mean	0.287	0.276	0.196	0.129	0.158	0.080
Z=0.4*Mean	0.283	0.226	0.174	0.099	0.143	0.060

Tables 4 and 5 show some inequality decompositions by villages using both the assets and the welfare index. From the tables it is easy to see that the greater source of inequality in the data is associated with the inequality within the villages. That means that inside each village the household living standards are relatively more unequal distributed than the living standards between different villages. However, it should be noted that the share of inequality between villages gets bigger when we account for public goods. this can be a signal that the public goods an unevenly distributed through the villages.

Inequality Decomposition by Village - Asset Index	**GE(-1)**	**GE(0)**	**GE(1)**	**GE(2)**
Both Groups	2.373	0.456	0.276	0.274
Within Groups	94.9%	77.8%	66.4%	66.5%
Between Groups	5.1%	22.2%	33.6%	33.5%



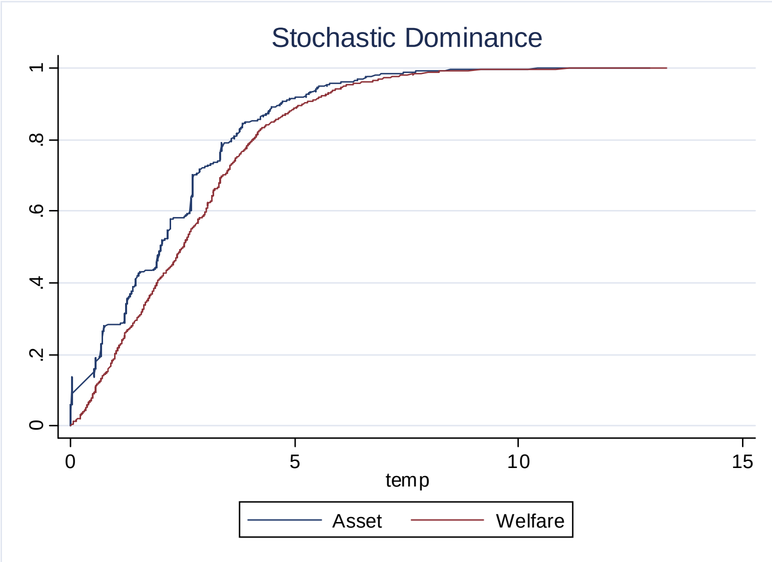
Inequality Decomposition by Village - Welfare Index	GE(-1)	GE(0)	GE(1)	GE(2)
Both Groups	0.853	0.313	0.234	0.237
Within Groups	84.6%	65.3%	57.8%	58.8%
Between Groups	15.4%	34.7%	42.2%	41.2%

In the figures displayed in appendix we show the spatial distribution of poverty and inequality in the Thai Tambons using our two models. The maps show how poverty and inequality vary in each village when we account for public goods in the welfare indicator. The main result shown in these maps the villages that have more poverty when we use the asset index are still the villages with more poverty and inequality when we use the welfare index. However, in all villages the level of poverty and inequality fall when we account for public goods.

Finally, by plotting the accumulated distributions of the asset and welfare indexes we can see that the latter stochastic dominates in first order the former. The Saposnik theorem (1983) shows that any increasing social welfare function will display a higher welfare in a distribution that stochastic dominate in first order other distribution. We can then use this theorem to argue that the "welfare index" distribution displays more welfare than the 'asset index" one.

4.2 The Treatment Effect

As described above one objective of this paper is to use the broader distribution of welfare estimated above to compute the treatment effect for the individuals of having access to public goods. As described above we used two strategies to identify the treatment effect. We first used matching. Following the strategy discussed above, in order to measure the gain a given agent obtain from having a given public good in his village we compared his welfare with the welfare of other agents with similar consumption of assets and public goods. The table bellow shows the estimated average treatment effect of having access to given public good.



Treatment Effects - Matching Results

Public Good	Community public health service	Basic medical treatment	Kindergarten school	Primary school within 3km	Secondary school	Vocation training center	Postal service center	Telephone service center	Police station
ATE	0.149	0.184	-0.208	-0.253	-0.279	0.148	-0.302	0.417	-0.350
ATE (%)	5.49%	6.80%	-7.68%	-9.33%	-10.30%	5.47%	-11.16%	15.38%	-12.91%

Some results obtained here a counterintuitive. We found negative average treatment effect for some public goods. This result may be related to the way PCA assigns weights to the welfare index. As our welfare index is the liner combination of assets and public goods that accounts for the biggest variance in the data, it is the case that the weights on public goods that are relatively common to all agents are negative. That means that such public goods are not explaining much of the variation on the data. Looking to the public goods that have a positive average treatment effect, we note that the biggest impact is associated with the Telephone Service Center (TSC). According to the model, the average treatment effect of having access to a TCS is about 0.417 utility units, which corresponds to an increase of 15% on the average utility level. It is important to remark that TSC is the rarest public good in our sample. About 95% of the households on the sample have no access to this public good. This is the main reason why having access to TSC has a so big impact on the individual welfare.

We also estimated the treatment effect using a different approach. As stressed above we created fake data bases where we assumed that a given village had certain public good, as in reality it was not the case. In this sense, we constructed for each household a counterfactual distribution of welfare where we assume we are observing the household in a different world. So we can directly compute the treatment effect as the difference of the actual and counterfactual distributions. This resulting treatment effect in different ways. In the table bellow we display some results. The table bellow shows for example the treatment effect of having a Postal Service Center installed in village 160. We see that households that are located in this village benefit much more than that rest of population. But, this is not always the case. Again some counterintuitive results appear. For example, a police station in village 155 will imply a positive treatment effect to the entire society, but at the same time that implies a negative treatment effect to the

Treatment Effect - Contrafactual Results

Public Good/Village	ATT (%) - Entire Population	ATT (%) - Village Population
Telephone Service in Village 160	2.35%	12.54%
Basic Medical Tretment in Village 101	0.33%	3.88%
Vocational Training Center in Village 36	0.05%	5.85%
Police Station in Village 155	3.43%	-14.58%

population of village 155. The same kind of exercise can be carried out for any village and any public good.

A final but important remark has to be made concerning the computation of the treatment effects displayed above. As we construct fake data and assume we are observing the agent in a different world, where some public good k is now located in certain village v, we are making a huge assumption. The point is that we can not be sure that the agent would choose to consume the same bundle of assets, or even if he would choose to live in the same village if the distribution of public goods were different from the observed. It may very well be the case that a given public good is a substitute to certain set o assets or vice versa. So, we can not be sure that as we place exogenously a public good in a given village the individual choices would be the same. In contrast, we can try to understand the treatment effects derived here as short run effects. That is, assuming that a public good is placed in a giving village, the treatment effect would be measuring a short run effect, the effect before any migration or any portfolio change. In order to obtain more robust results we should we a better model of individual choice where we model explicitly the simultaneous choice of consumption of private and public goods.

4.3 Placing The Public Good

Using the broader welfare distribution derived above we developed an algorithm to place public goods in villages in order to maximize a given social welfare function. For simplicity we chose to use an "utilitarist" SWF given by

$$W = \frac{1}{n}\sum_{1=1}^{n} Y_i$$

This implies as W is maximized the average utility level of the economy is maximized. The table bellow presents some results derived using the algorithm describe above. We see that given our choice of W, the best village to place an extra police station would be village 155. The aggregate welfare gain of placing a police station in this village would be 3.44%. Note that welfare gain is calculated as the perceptual difference in the value of SWF applied to the fake data (where we place a police station in village 155) to the value of the SWF applied to the original data. We see that the welfare gain vary from 0.67% when a "Basic Medical Treatment" facility is installed in village 107 to 7.08% when a "Secondary School" is placed in village 151.

Maximum SWF - Utilitarian

Public Good	Community public health service	Basic medical treatment	Kindergarten school	Primary school within 3km	Secondary school	Vocation training center	Postal service center	Telephone service center	Police station
Village	15	107	90	149	151	104	52	94	155
Welare Gain	0.70%	0.67%	3.47%	0.38%	7.08%	0.41%	4.17%	0.71%	3.44%

It is important to mention, that the public good placement algorithm implemented here just takes in account the individual welfare calculate by PCA. We would expect to obtain different results with we used a more general model allowing for geographical variables, distance for close public goods, transportation costs among villages and so on. In the appendix we present a map showing the spatial location of the villages chosen to receive the extra public goods.

Again we have to remark that the analysis developed above have to be interpreted as short run effects. That is, we have to read the results above as: in the short run, before any migration or change in individual's portfolios, social welfare would be maximized if a public good k were placed in village v.

5 Conclusions and Possible Extentions

In this paper we tried to construct a broader measure of welfare taking in account the access households have to public goods. A simple discrete choice model was developed to justify the use of data on the ownership of assets and access to public goods in the construction of a broader welfare index. It was argued, that the utilization of factor methods are suitable in the estimation of a latent utility model derived form such simple discrete choice model. For simplicity we chose to use PCA as the estimation method. But the same analysis could have been carried out using a more general factor model.

In the paper we make an attempt to justify theoretically the use of Principal Component Analysis in the construction of a wealth or welfare index (this technique has already been used for this purpose in the literature without much justification). The technique employed has naturally pros and cons. One of the majors advantages of the PCA approach is it simplicity. It is relatively easy to find reliable data on asset ownership and it is still easy to find data on the existence of public goods in neighborhoods where households are located. In contrast, this approach relies heavily on the assumption that PCA on the ownership of assets and the access to public goods is able to capture the variation of a latent variable identified as the long run welfare level. Even though this assumption has been heavily used in the literature, it is still a pretty strong assumption.

Using the techniques developed above we found some interesting results. It was shown that distributions of "utility" that includes information on the access households have to public goods display less poverty and inequality and bigger welfare relatively to distributions of utility derivated form the ownership of assets. We have also shown that using a broader measure of welfare the inequality between villages increases. This fact may imply that the uneven distribution of public goods across villages is an important source of inequality in this economy.

The broader distribution of welfare was also used in the estimation of the treatment effect of having access to public goods. Unfortunately in this session results were not so satisfactory. Two strategies were employed to identify the treatment effect: matching and direct construction of counterfactual distribution of welfare. In both cases we found results hard to explain, such as negative treatment effect of having access to certain public goods. One important remark regarding this session is that when we do matching or when we construct counterfactual distribution of welfare we are making some very strong assumptions. In this sense, all the results have to be understood in a short run context. The treatment effect we obtain by comparing the original and a counterfactual distribution of welfare is assumed to hold only in the short run, before any kind of migration of portfolio change. In this sense, in order to obtain more robust estimates of treatment effects it may be necessary the construction of a more general choice model where the simultaneous choice of private and public goods is contemplated. The use of a structural hedonic model, where the price of housing or land is a function of the distance to public goods may solve this problem.

The last exercise developed in the paper was the placement of public goods in order to maximize a social welfare function. Here again, the results obtained have to be understood in a shot run context, as stressed above. The simple algorithm developed above can be used in a more general model where individual welfare is not the single relevant variable for public good placement. The analysis could be more complete if in placing public goods in villages we could also take in account geographical and demographic variables.

But apart from the implementation problems we faced, the idea of constructing a broader measure of welfare still seems very promising from both an empirical and theoretical point of view. Note that the model developed here could be augmented in different ways. The first and obvious one would be to estimate the latent utility level using a factor model that allows for heterogeneity among the agents. The model could be even more sophisticated if the heterogeneous agents are spatially correlated. That is, if we make some assumption that the parameters of the model (that represent prices end preferences) are function of the village household is located in. The main idea is that such a model could be enlarged in some fashion to capture effects of the spatial distribution of public goods. So far we were not able to treat space properly. As mention above, in the exercise of the placement of the public good, we did not take in account any measure of distances between villages. So we think the study developed here in a kind of explanatory way, could be improved with better theoretical models and more sophisticated empirical analysis

The important point to remark is that the development of such techniques could be very helpful to improve public policy decisions. Once we can consistently estimate this broader measure of welfare we will be able analyze the effects on poverty, inequality of the provision of different public goods. Such a technique will also allow policy makers to think about the geographical or spatial distribution of public goods in order to maximize some social welfare function. At the same time, as we obtain a proxy for the welfare distribution with and without taking in account public goods, one can think about the development of a model to attribute social value for public goods based on the difference of those two distributions.

In summary, the exercises developed above have shown, still imperfectly, that the account of public goods in the households' living standards may have an important effect on their welfare. We are sure this is an interesting research topic, and we are quite confident that, with better techniques, we would be able to use the data in a more efficient way, and improve the results displayed above in future work.

References

[1] Aaron, H., McGuire, M (1970) "Public Goods and the Income Distribution". Econometrica Vol 38 no 6

[2] Anand, S. Ravalllion, M. (1993) "Human Development in Poor Countries: The Role of Privite Income and Public Services". Journal of Economic Perspectives, Vol 7 No 1

[3] Carneiro, P. Hansen, K. Heckman, J. (2003) Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice. IZA DP No. 767

[4] Deaton, A., (1997), The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. The John Hopkins University Press, Baltimore.

[5] Filmer, D. Pritchett, L. (1998) "Estimating Wealth Effects without Expenditure Data or Tears: With an Application to Educational Enrollments in States of India" World Bank Policy Research Working Paper No. 1994

[6] Jeong, H. Townsend, R. (2003) "Growth and Inequality: Model Evaluation Based on an Estimation-Calibration Strategy". Working paper University o Chicago

[7] Mailtal, S. (1973) “Public Goods and the Income Distribution: Some Further Results”. Econometrica Vol 41 no 3

[8] Saposnik, R., (1983), “On Evaluating Income Distributions: Rank Dominance, the Suppes-Sen Grading Principle of Justice and Pareto Optimality“, Public Choice, 40: 329-36.

[9] Tucker, L. MacCallun, R. (1997) "Explanatoty Factor Analysis"