UNU-WIDER | 30 YEARS OF RESEARCH FOR DEVELOPMENT

WIDER Working Paper 2015/102

Poverty decomposition by regression

An application to Tanzania

Tomoki Fujii*

October 2015

Abstract: We develop a poverty decomposition method that is based on a consumption regression model. Because this method uses an integral of the partial derivatives of a poverty measure with respect to time, the resulting poverty decomposition satisfies time-reversion consistency and sub-period additivity. Unlike the existing poverty decomposition methods, it allows us to ascribe the observed change in poverty to various covariates of interest collected at a disaggregate level. This method is applied to two datasets from Tanzania to assess, among others, the short- and long-term impacts of infrastructure and market access on poverty.

Keywords: FGT measure, Watts measure, market access, infrastructure
JEL classification: I32, O10

Acknowledgements: I thank Daniel H. Vedia Jerez for useful comments. A preliminary version of this paper was presented at the UNU-WIDER conference on 'Inclusive Growth in Africa: Measurement, Causes, and Consequences', in Helsinki, Finland. The author gratefully acknowledges the support of Tokyo Center for Economic Research for this research.

*School of Economics, Singapore Management University; tfujii@smu.edu.sg

An earlier version of this paper has been presented at the UNU-WIDER 'Inclusive Growth in Africa: Measurement, Causes, and Consequences' conference, held 20–21 September 2013 in Helsinki, Finland.

ISSN 1798-7237 ISBN 978-92-9230-991-6 https://doi.org/10.35188/UNU-WIDER/2015/991-6

Typescript prepared by the author.

UNU-WIDER gratefully acknowledges the financial contributions to the research programme from the governments of Denmark, Finland, Sweden, and the United Kingdom.

The World Institute for Development Economics Research (WIDER) was established by the United Nations University (UNU) as its first research and training centre and started work in Helsinki, Finland in 1985. The Institute undertakes applied research and policy analysis on structural changes affecting the developing and transitional economies, provides a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and promotes capacity strengthening and training in the field of economic and social policy-making. Work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

UNU-WIDER, Katajanokanlaituri 6 B, 00160 Helsinki, Finland, wider.unu.edu

1 Introduction

Poverty decomposition technique has been used to describe the main driving forces of poverty change. However, existing methods typically rely on the population-level characteristics such as mean income and income distribution, which makes it difficult for users to clearly see the relationship between poverty and the changes in the distribution of covariates observed at a disaggregate level such as individual, household, and community levels.

To address this issue, we propose a new decomposition method using regressions to ascribe observed poverty changes to variables collected at a disaggregate level while retaining some desirable characteristics such as time-reversion consistency—the property that when the initial and terminal time periods are swapped the contribution of each component in the decomposition to the observed poverty change is exactly the same as the original decomposition except that the sign is opposite—and the sub-period additivity—the property that the contribution of each component for a given period can be expressed as the sum of its contributions for all the sub-periods within that period.

As an illustration, we apply the proposed method to panel datasets from Tanzania to assess the short- and long-term impacts of infrastructure development and market access. This application is of interest for two reasons. First, an analysis of this sort is difficult to carry out with the existing decomposition methods. This point is particularly relevant when the variable of interest is continuous (e.g., distance to the main road). Therefore, this application showcases the strength of our method.

Second, while the importance of infrastructure development and market access is well recognised in economics, existing studies typically focus on one type of infrastructure to obtain a clean identification. This, however, comes at a cost of not being able to compare the relative importance of various types of infrastructure. As a result, it is difficult for users of decomposition analysis to determine what types of infrastructure may be most relevant to poverty reduction. While our decomposition is essentially an accounting exercise and does not describe the causality, it provides policy-makers with useful information by enabling simultaneous comparison of multiple potential sources of poverty change.

This paper is organised as follows. In Section 2, we provide a brief review of related literature. Section 3 develops the method in a general framework based on a class of additively decomposable measures of poverty that are continuous at the poverty line. We also provide a graphical explanation of our decomposition method to facilitate intuitive understanding. Section 4 describes the data, followed by the presentation of decomposition results in Section 5. Finally, we provide some discussion in Section 6.

2 Review of related literature

This paper is related to two strands of literature. The first strand is the literature on poverty decomposition, which has grown steadily over the last two decades or so. The standard approach to poverty decomposition in this literature has been to allow only one factor to change while keeping all the remaining factors fixed to compute the contribution of the moving factor to the overall poverty change. For example, a seminal work by Datt and Ravallion (1992) takes as the growth [redistribution] component the poverty change that can be explained by the change in the mean consumption [consumption distribution] while keeping the consumption distribution [mean consumption] fixed at the initial period. Their method has been widely used in the literature, including Ravallion and Huppi (1991), Grootaert (1995), and Sahn and Stifel (2000).

One major drawback of the decomposition by Datt and Ravallion (1992) is that their decomposition method comes with a residual term that does not have a straightforward interpretation. One way to deal with this issue is to consider the change sequentially. For example, Kakwani and Subbarao (1990) implicitly assumed that the growth takes place first and the redistribution second, whereas Jain and Tendulkar (1990) consider a decomposition in which redistribution precedes growth. Instead of relying on this arbitrary choice of sequence, it is also possible to take the average of decompositions based on these alternative sequences. This decomposition is called Shapley decomposition because it is characterised by the marginal contribution of each component for all the possible sequences, which is similar to the Shapley solution in cooperative games (Kolenikov and Shorrocks, 2005; Shorrocks, 2013).

All of these decompositions suffer from the lack of sub-period additivity. That is, the contribution of a particular component to poverty change between t_0 and t_2 cannot be expressed simply as a sum of contributions between t_0 and t_1 and between t_1 and t_2. Furthermore, with an exception of the Shapley decomposition, none of these decompositions satisfy the time-reversion consistency. In other words, the contribution of each component for the change in poverty from time t_0 to time t_1 is not the same as that component for the change from time t_1 to time t_0 with the opposite sign as pointed out by Fujii (2014).

To address these issues, Fujii (2014) proposed a dynamic poverty decomposition method based on integration, thereby effectively internalizing the reference period. He then demonstrated that his decomposition method can be extended to include additional components such as within-region inequality and between-region inequality components, which was not possible or practical before.

All the methods discussed above essentially rely on the population-level characteristics such as mean consumption and consumption distribution. However, it is difficult to see from poverty decomposition analysis based on these methods how unit-level characteristics affect poverty. Therefore, we propose a new decomposition method using regressions such that observed poverty changes can be ascribed to unit-level variables. As with Fujii (2014), our method is theoretically founded on integration and thus retains sub-period additivity and time-reversion consistency.

The second strand of literature that this paper relates to is the growing body of economic literature on infrastructure. Various researchers have investigated the social and economic impacts of specific types of infrastructure such as electricity (Peters et al., 2011; Rud, 2012; Lipscomb et al., 2013; Dinkelman, 2011; Grogan and Sadanand, 2013), dams (Duflo and Pande, 2007), transportation infrastructure (Fernald, 1999; Banerjee et al., 2012), and telecommunications infrastructure (Röller and Waverman, 2001) among others (See also Gramlich (1994) and Straub (2008) for a review of literature). Because these studies focus on a particular type of infrastructure, we are unable to compare the relative importance of different types of infrastructure in reducing poverty and achieving other important policy objectives. As a result, it is difficult for policy-makers to determine what types of infrastructure may be most effective for poverty reduction. Therefore, our decomposition complements the literature on infrastructure by enabling the comparison of poverty impacts across different kinds of infrastructure.

3 Methodology

General Framework

To formally introduce our decomposition method, it is necessary to introduce some notations. Suppose that there is a continuum of households indexed by h on a unit interval. The consumption of household h at time t is given by $y_{ht}(>0)$. We start with a household-level poverty measure p_{ht} that has the following form:

$$p_{ht} = g\left(\frac{y_{ht}}{z}\right)\mathbf{1}(y_h \leq z) = g\left(\tilde{y}_{ht}\right)\mathbf{1}(\tilde{y}_{ht} \leq 1), \tag{1}$$

where z, $\tilde{y}_{ht} \equiv y_{ht}/z$, and $\mathbf{1}(\cdot)$ are respectively the poverty line, consumption per capita normalised by the poverty line, and an indicator function, which takes one if the argument is true and zero otherwise. We assume that the function $g(\tilde{y}_{ht})$ is once-differentiable for all $\tilde{y} > 0$. Further, we assume that p_{ht} to be continuous in $\tilde{y}_{ht}$, which requires $g(1) = 0$.

We consider a class of additively decomposable measures P_t, which can be written as:

$$P_t = \int_0^1 g\left(\frac{y_{ht}}{z}\right)\mathbf{1}(y_h \leq z)\mathrm{d}h, \tag{2}$$

The class of poverty measures given in eq. (2) is not restrictive because it includes commonly used poverty measures. For example, we obtain the Foster-Greer-Thorbecke (FGT) measure of poverty (Foster et al., 1984) by setting $g(\tilde{y}) = (1-\tilde{y})^{\alpha}$. In particular, FGT measures with $\alpha = 0$, $\alpha = 1$, and $\alpha = 2$ are commonly called the head count index (or poverty rate), poverty gap index, and poverty severity index, respectively, and widely used in the poverty literature. Because the head count index does not satisfy $g(1) = 0$, we do not use it in our application. We also use the Watts poverty measure (Watts, 1968), which can be obtained by setting $g(\tilde{y}) = -\ln \tilde{y}$. While this measure is less frequently used than the FGT poverty measures, we include them because it possesses some desirable properties (Zheng, 1993) and its graphical representation is particularly helpful for intuitive understanding of our decomposition method as elaborated later in this section.

To assess the impacts of covariates on poverty, we consider the following model of logarithmic consumption:

$$\ln \tilde{y}_{ht} = x_{ht}^T \beta_t + \epsilon_{ht}, \tag{3}$$

where x_{ht} is a vector of covariates at time t for household h, β_t is a (potentially) time-varying coefficient and ϵ_{ht} is a residual term. The jth component of x_{ht} and β_t are denoted by $x_{ht,j}$ and $\beta_{t,j}$, respectively. We will consider the cases where some additional restrictions are added in our empirical applications but the following general discussion is still applicable.

For simplicity of presentation, we hereafter use the dot notation for time derivatives (e.g., $\dot{P}_t \equiv \mathrm{d}P_t/\mathrm{d}t$). By taking the time derivative of eq. (2) under the assumption of eq. (3) and applying the chain rule, we have:

$$\begin{aligned} \dot{P}_t &= \int_0^1 g'(\tilde{y}_{ht})\, \dot{\tilde{y}}_{ht} \mathbf{1}(\tilde{y}_{ht} \leq 1) \mathrm{d}h + \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_0^1 \mathbf{1}(\tilde{y}_{ht} \leq 1) g(1) \mathrm{d}h\right] \\ &= \int_0^1 g'(\tilde{y}_{ht})\, \tilde{y}_{ht} (\dot{x}_{ht}^T \beta_t + x_{ht}^T \dot{\beta}_t + \dot{\epsilon}_{ht}) \mathbf{1}(\tilde{y}_{ht} \leq 1) \mathrm{d}h. \end{aligned} \tag{4}$$

where $g(1) = 0$ was used to derive the second line. Now, by integrating eq. (4) from $t = t_0$ to $t = t_1$ and breaking it up term by term, we have the following decomposition:

$$P(t_1) - P(t_0) = \sum_j \left[X^j(t_0, t_1) + S^j(t_0, t_1)\right] + R(t_0, t_1) \tag{5}$$

where,

$$
\begin{aligned}
X^j(t_0,t_1) &\equiv \int_{t_0}^{t_1}\int_0^1 g'\left(\tilde{y}_{ht}\right)\tilde{y}_{ht}\dot{x}_{ht,j}\beta_{t,j}\mathbf{1}(\tilde{y}_{ht}\leq 1)\mathrm{d}h\mathrm{d}t & (6)\\
S^j(t_0,t_1) &\equiv \int_{t_0}^{t_1}\int_0^1 g'\left(\tilde{y}_{ht}\right)\tilde{y}_{ht}x_{ht,j}\dot{\beta}_{t,j}\mathbf{1}(\tilde{y}_{ht}\leq 1)\mathrm{d}h\mathrm{d}t & (7)\\
R(t_0,t_1) &\equiv \int_{t_0}^{t_1}\int_0^1 g'\left(\tilde{y}_{ht}\right)\tilde{y}_{ht}\dot{\epsilon}_{ht}\mathbf{1}(\tilde{y}_{ht}\leq 1)\mathrm{d}h\mathrm{d}t. & (8)
\end{aligned}
$$

Both X^j and S^j are the poverty change due to the jth covariate. However, the former is due to the change in the distribution of the jth covariate whereas the latter is due to the structural change in the relationship between the jth covariate and consumption. Therefore, we shall call X^j and S^j the covariate component and structural component for the jth covariate, respectively. The last term R is the poverty change due to the non-systematic component and called the residual component.

It is clear from eq. (6) that $X^j(t_1,t_0)=-X^j(t_0,t_1)$ holds once the path along which x, β, and ϵ change is fixed. It is also clear that $X^j(t_0,t_1)=X^j(t_0,t_2)+X^j(t_2,t_1)$ holds by the nature of integration and that similar relationships hold for S^j and R. Therefore, the poverty decomposition described in eq. (5) indeed satisfies time-reversion consistency and sub-period additivity.

While these properties are desirable, the decomposition given in eq. (5) cannot be implement in a typical empirical setting because it requires continuous observation of (y_{ht},x_{ht}^T). Therefore, we need to make some additional assumptions to implement eq. (5).

First, it should be clear that the decomposition described in eq. (5) requires the knowledge of changes in x and ϵ for each household h over time. Therefore, it is difficult to implemented it without a panel dataset. Even with a panel dataset, the observations of y_{ht} and x_{ht} are typically available only at $t=t_0$ and $t=t_1$. Therefore, to implement eq. (5), we make the following assumptions:

$$
\begin{aligned}
x_{ht} &= (1-\tau)x_{ht_0}+\tau x_{ht_1} & (9)\\
\beta_t &= (1-\tau)\beta_{ht_0}+\tau\beta_{ht_1} & (10)\\
\epsilon_{ht} &= (1-\tau)\epsilon_{ht_0}+\tau\epsilon_{ht_1} &\\
&= (1-\tau)(\ln\tilde{y}_{ht_0}-x_{ht_0}^T\beta_{t_0})+\tau(\ln\tilde{y}_{ht_1}-x_{ht_1}^T\beta_{t_1}), & (11)
\end{aligned}
$$

where $\tau\equiv(t-t_0)/(t_1-t_0)$.

Eq. (9) states that the covariates change smoothly and linearly between $t=t_0$ and $t=t_1$. While this is a reasonable assumption, it may not be readily applicable to discrete covariates which cannot change continuously. In this case, we can instead reinterpret x_{ht} as the expected value of the covariate for $t\in[t_0,t_1]$, which coincides with the observed value at $t = t_0$ and $t = t_1$.

In eq. (10), we assume that β also changes smoothly and linearly between $t = t_0$ and $t = t_1$. Finally, eq. (11) imposes a condition that ϵ_{ht} also changes smoothly and linearly.

Under these assumptions, eqs. (6)-(8) can be rewritten as follows:

$$X^j = \frac{1}{(t_1 - t_0)} \int_{t_0}^{t_1} \int_0^1 g'(\tilde{y}_{ht})\, \tilde{y}_{ht}(x_{ht_1,j} - x_{ht_0,j})\beta_{t,j} \mathbf{1}(\tilde{y}_{ht} \leq 1) \mathrm{d}h \mathrm{d}t \tag{12}$$

$$S^j = \frac{1}{(t_1 - t_0)} \int_{t_0}^{t_1} \int_0^1 g'(\tilde{y}_{ht})\, \tilde{y}_{ht} x_{ht,j}(\beta_{t_1,j} - \beta_{t_0,j}) \mathbf{1}(\tilde{y}_{ht} \leq 1) \mathrm{d}h \mathrm{d}t \tag{13}$$

$$R = \frac{1}{(t_1 - t_0)} \int_{t_0}^{t_1} \int_0^1 g'(\tilde{y}_{ht})\, \tilde{y}_{ht} \left[\ln \frac{\tilde{y}_{ht_1}}{\tilde{y}_{ht_0}} - \left(x_{ht_1}^T \beta_{t_1} - x_{ht_0}^T \beta_{t_0} \right) \right] \mathbf{1}(\tilde{y}_{ht} \leq 1) \mathrm{d}h \mathrm{d}t. \tag{14}$$

To implement these equations, we obtain estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 by regressions. Replacing β's with $\hat{\beta}$'s in eq. (10), we have an estimate $\hat{\beta}_t$ of β_t. Using this, the sample analogues of eqs. (12)-(14) can be obtained. For example, in the case of eq. (12), the sample analogues can be written as follows:

$$\hat{X}^j = \frac{1}{(t_1 - t_0)N} \int_{t_0}^{t_1} \left[\sum_{h \in \mathcal{S}} g'(\tilde{y}_{ht})\, \tilde{y}_{ht}(x_{ht_1,j} - x_{ht_0,j}) \hat{\beta}_{t,j} \right] \mathbf{1}(\tilde{y}_{ht} \leq 1) \mathrm{d}t, \tag{15}$$

where $\mathcal{S}$ is the index set for the households in the sample and $N \equiv \#\{\mathcal{S}\}$ is the sample size. We can similarly define the sample analogues of $\hat{S}^j$ and $\hat{R}^1$.

While we run ordinary least-squares (OLS) regressions for $t = t_0$ and $t = t_1$ separately to obtain $\hat{\beta}_{t_0}$ and $\hat{\beta}_{t_1}$, respectively, this does not necessarily require that ϵ_{ht} has to be uncorrelated for the same household h across different t. Because we use the same set of covariates for the two time periods, the OLS estimate is equivalent to the seemingly unrelated regression.

Once the coefficient estimates are obtained, the integral in eq. (15) is evaluated by numerical integration. We use the classical Runge-Kutta method with a sufficiently small time step to make the computational error negligible.

It should be noted that our decomposition does not depend on a particular estimation method. When each household is observed for a relatively large periods of time, it is possible to use fixed-effects. In this case, the error term is assumed to satisfy $\epsilon_{ht} = \eta_h + u_{ht}$ where the household-specific fixed effect η_h and the idiosyncratic term u_{ht} are independent, respectively, across h and across h and t and also independent of each other. Because we only have at most three observations per household in our datasets, we choose not to explore this method.

It is also possible to apply a common model for the two time periods. In this case, we are assuming that the coefficients are fixed over time (i.e., $\beta_t = \beta$ for $\forall t \in [t_0, t_1]$) and thus the structural component is identically equal to zero. The fixed coefficients can be estimated by pooling the data over the two (or more) periods. It is also possible to impose the constancy of coefficients for a subset of covariates. The fixed-coefficient model has an advantage that the coefficients can be more accurately estimated when the coefficients are indeed constant over time.

However, the use of fixed coefficients may lead to a very misleading conclusion, when the coefficients are in fact varying over time. In this case, the change in poverty due to the structural change is attributed to the changes in covariates. As a result, even when a covariate has no impact on consumption per capita, the observed change in poverty may be wrongly attributed to this covariate simply because it is changing over time. We shall explain this point again later with a graphical example. Because of this potential problem with the fixed-coefficient model, we prefer to use a model with flexible (i.e., time-varying) coefficients.

Another important aspect of our method is that the covariates x_{ht} are not required to be exogenous, because the decomposition exercise is an accounting exercise in which the observed poverty change is ascribed to the covariates of interest. Nevertheless, our decomposition exercise is still able to identify the factors that tend to change hand-in-hand with poverty. If we wish to make a causal inference, our decomposition can be conducted, for example, with instrumental-variables estimator of β, provided that suitable instruments are available. In this case, instead of using eqs. (12)-(14), it may be appropriate to use an alternative path of time evolution to take into account the endogeneity of x_{ht}.

While we have used a household as a unit of analysis to keep the presentation simple, it is straightforward to extend our method to the case where the unit of analysis is an individual. This can be achieved by using the population expansion factors as weights. The only difference is that the changes in weights over time due to, for example, changes in household size, also need be accounted for in the decomposition analysis and thus an additional term involving the time derivative of the weights has to be added to eq. (4).

Graphical representation for special cases

To facilitate the understanding of our decomposition, it is useful to consider a case in which the decomposition can be easily represented in a graph. To this end, we consider a decomposition of the Watts poverty measure on a unit time interval (i.e., $t_0 = 0$ and $t_1 = 1$) when the covariate of interest x is a scalar and the poverty line is normalised to be unity ($z = 1$). Notice here that $x_0 = x_{t_0}$, $x_1 = x_{t_1}$, and so on hold, because the unit time interval is considered. Further, for the sake of simplicity, we ignore ϵ (i.e., $\epsilon_{ht} = 0$ for all h and t) and assume that the intercept remains constant. Everyone is assumed to

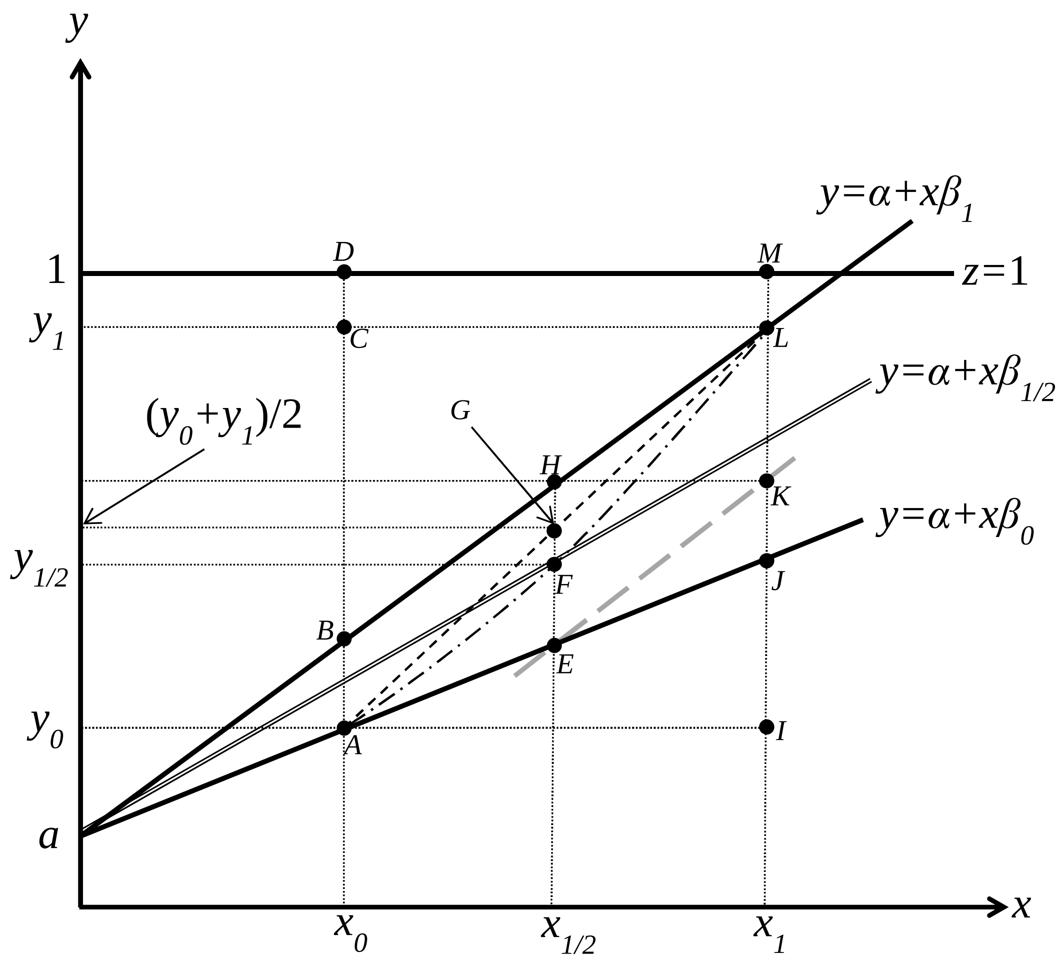

Figure 1: A graphical illustration of the consequence of using a fixed-coefficient model in poverty decomposition analysis.

be identical such that the subscript h can be dropped.

Figure 1 provides a graphical representation under these assumptions. The vertical axis measures the consumption per capita in a log scale, whereas the horizontal axis measures the covariate x. Figure 1 shows that the initial Watts poverty measure at $t = 0$ is given by AD,[1] because it is the shortfall in logarithmic consumption from the poverty line. Likewise, the terminal Watts poverty measure at $t = 1$ is represented by LM. Therefore, the goal of poverty decomposition is to decompose the change in the Watts poverty measure $AC(= IL)$ into relevant components. To highlight this point, it is instructive to consider a sequential change first. Thus, instead of assuming eqs. (9) and (10), suppose alternatively that x changes from x_0 to x_1 first and then b changes from b_0 to b_1 second. In this case, the Watts poverty measure reduces by IJ by the change in x and further reduces by JL by the change in b. Therefore, the covariate and structural components are $-IJ$ and -JL, respectively. It is straightforward to verify that they are $-BC$ and $-AB$ if b's change precedes x's change.

Now, let us consider our decomposition method under eqs. (9) and (10). In this case,

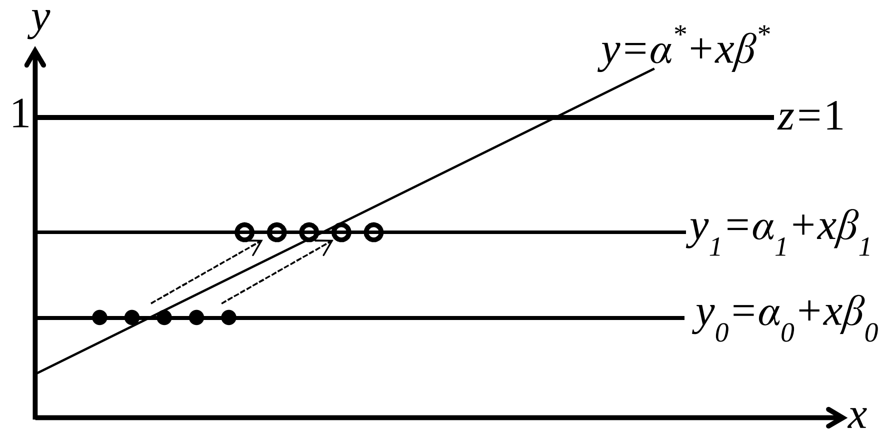

Figure 2: A graphical illustration of regression-based decomposition of the Watts poverty measure.

plugging these equations in eq. (3), we have:

$$\ln \tilde{y}_t = x_0\beta_0 + t\left[x_0(\beta_1 - \beta_0) + \beta_0(x_1 - x_0)\right] + t^2(x_1 - x_0)(\beta_1 - \beta_0),$$

This shows that the pair (x_t, y_t) varies over the arc AFL drawn by a dot and dash line rather than the line segment AGL. It is also straightforward to show that eqs. (12) and (13) reduces to the following:

$$X = -\frac{(x_1 - x_0)(\beta_1 + \beta_0)}{2} = -(x_1 - x_0)\beta_{1/2} \quad (16)$$

$$S = -\frac{(x_1 + x_0)(\beta_1 - \beta_0)}{2} = -(\beta_1 - \beta_0)x_{1/2} \quad (17)$$

What the covariate component X measures is the reduction in the Watts poverty measure when the covariate changes from x_0 to x_1 evaluated at the average slope $\beta_{1/2}$. In Figure 1, a grey dashed line going through point E with a slope of b_1 is drawn. Because the line segment AE has a slope b_0 and EK a slope b_1, a line that goes through points A and K (not drawn) has a slope $b_{1/2}$. Therefore, it can be seen that the covariate component is $-IK$ in this case. Similarly, the structural component S measures the reduction in the Watts poverty measure when the slope changes from β_0 to β_1 with x fixed at $x_{1/2}$. Therefore, the structural component is$-EH(= -KL)$.

It is straightforward to verify that X is the average of the covariate component for the two alternative sequential changes considered above (i.e., $IK = (JL + AB)/2$). Similarly, S is the average of the structural component for the two alternative sequential changes (i.e., $EH = (IJ + BC)/2$). Therefore, in the special case we considered here, our decomposition method can be interpreted as a type of Shapley decomposition.

Let us now use a figure similar to Figure 1 to highlight the potential problem with a fixed-coefficient model applied to different time periods. In the situation described in

Figure 3, each black dot and each ring represent the combination of the average consumption per capita as a ratio of poverty line and the value of its covariate x at $t = 0$ and $t = 1$, respectively. They make a parallel shift as the dashed arrow indicates. In this case, the covariate x does not affect y at all, hence $\beta_0 = \beta_1 = 0$. However, when we use a fixed-coefficient model, a misleading conclusion emerges. Because both x and y rise over the two time periods, even though x has no impact on y, x captures most of the effect of changing α in the fixed-coefficient model as the line $y = \alpha^* + x\beta^*$ indicates.

4 Data and Summary Statistics

For our empirical illustration, we use two separate datasets. The first dataset is the Kagera Health and Development Survey (KHDS) conducted in the Kagera region of Tanzania, which is located on the western shore of Lake Victoria adjacent to Uganda and Rwanda. This survey was conducted to measure the economic impact of adult mortality on surviving household members. The KHDS household sample was selected randomly, stratified on geography, community adult mortality rates, and indicators at the household level that were thought to be predictive of future adult deaths (See Ainsworth and Semali (2000) and studies cited therein). This dataset is suitable for our purpose because it is a panel dataset and contains consumption data crucial for poverty analysis (See World Bank (2004) and Beegle et al. (2006) for the description of the data). The KHDS dataset includes a community module that is necessary for analyzing the community-level characteristics. Because households are tracked over a long period of time, long-term impacts of various factors can be investigated with the KHDS dataset.

In this study, the KHDS data for years 1991 (Wave 1) and 2004 (Wave 5) are used. While there are other waves in 1992 (Wave 2), 1993 (Wave 3), 1994 (Wave 4), and 2010 (Wave 6), we chose not to use those waves for the following reasons. First, Waves 2-4 are not used because the design of consumption component in these waves is not comparable to that in Waves 1, 5, and 6 (World Bank, 2012), making them unsuitable for our analysis. Second, Wave 6 does not contain a community survey. Because the community survey contains some key variables of interest such as local infrastructure and market access, we chose not to use Wave 6 either. Because the KHDS does not contain sample weights that are applicable to all households in the sample, we only report unweighted results.

The second dataset we use is the Tanzania National Panel Survey (TZNPS). We use all three waves that are currently available, which are conducted in years 2008/09, 2010/11, and 2012/13. TZNPS was implemented by the Tanzania National Bureau of Statistics with technical support from the World Bank. The main objective of the TZNPS is to provide high-quality household-level data for monitoring poverty dynamics, tracking the progress of the MKUKUTA poverty reduction strategy[2], and evaluating the impact of other major socioeconomic factors (See National Bureau of Statistics (2010, 2012, 2014) for the description of the TZNPS dataset). The TZNPS sample is nationally representative and has both household and community components. The TZNPS dataset also comes with auxiliary geographic dataset that can be merged into the household-level data.[3] For all the results reported in this paper, we apply the sample weights for the TZNPS 2008/09 data and ignore the attrition of the sample.

While both KHDS and TZNPS are panel household surveys, the composition of a household may vary over time because of birth, death, marriage, separation, and migration of household members among other reasons. Birth and death cannot be dealt with in our analytical framework because we obviously do not observe the consumption before birth or after death. While we are in principle able to deal with other issues by tracking each household member, the linearity assumption is likely to be more problematic in this case. Therefore, we chose to leave issues like marriage, separation, and migration as a subject for future research and focus on those households whose head is the same during the observation period. We further restricted the sample to those households for which consumption and other key characteristics are fully observed in all the waves used in our analysis.

To compute the normalised consumption per capita $\tilde{y}$, we first need to compute real consumption aggregate per capita and then divide it by the poverty line. For the real consumption aggregate measure, the original KHDS dataset contains consumption aggregate measure based on prices for year 2010, adjusted for spatial price differences. We simply divide this by the household size to arrive at the real consumption aggregate per capita. The TZNPS dataset contains consumption aggregate measure adjusted for spatial prices. We further adjust it for inflation by the Consumer Price Index (CPI, base year=2010) taken from the World Development Indicators (WDI) database.[4] For the poverty line, we use the \$2 and \$1.25 international poverty lines for 2005 for moderate and extreme poverty, respectively. We convert them into Tanzanian Schillings using the Purchasing Power Parity conversion factor for private consumption taken from the WDI database and express it in 2010 prices using the CPI. While we adopted the same international poverty lines for both KHDS and TZNPS, the poverty figures may not be directly comparable because no attempt is made to make the consumption aggregates comparable.

Table 1 reports some key summary statistics for the KHDS data. This table shows that the household size and proportion of household head engaging in farming or fishing activities have dropped. Some improvements have been made in the education of household head and the proportion of electrified households has also increased in the Kagera region between 1991 and 2004. However, a number of indicators have worsened, including the distance to the nearest daily market and proportion of households residing in a community with a health center. Note that the panel households are used to compute

Table 1: Key summary statistics for the KHDS data.

Survey round	KHDS 1991		KHDS 2004	
Variable	Mean	(s.e.)	Mean	(s.e.)
Household size	6.214	(3.132)	5.519	(2.812)
Head has no job	0.014	(0.117)	0.063	(0.243)
Head's activity is farming/fishing	0.837	(0.370)	0.772	(0.420)
Head completed primary educ.	0.326	(0.469)	0.591	(0.492)
Electrified household	0.021	(0.143)	0.053	(0.225)
Distance to daily market (km)	8.547	(12.303)	13.044	(21.772)
Motorable road in community	0.949	(0.221)	0.958	(0.201)
Health center in community	0.119	(0.324)	0.086	(0.281)
Head count index for z =\$2	0.8860		0.8070	
Head count index for z =\$1.25	0.5651		0.5093	
# Obs	430		430	

the summary statistics. Therefore, the apparent increase in the proportion of household heads without a job may be driven by the retirement of the household head. The bottom part of Table 1 shows that the Kagera region has witnessed only a small improvement in the head count index (i.e., FGT_0) and that a majority of the population remains under the poverty line in 2004 regardless of whether \$2 or \$1.25 poverty line is used.

Table 2 provides summary statistics for the TZNPS data. We have tried to include the key characteristics reported in Table 1 wherever possible. However, the TZNPS data do not have information on whether there is a motorable road in the community. Therefore, we substitute it with the distance to a major road. Obviously, these two variables are different but both measure road accessibility. Also, the indicator variable that a health center is in the community of residence is constructed differently between the KHDS and TZNPS datasets. In the former dataset, this indicator variable is directly observed. However, in the latter, a separate question is asked for public and private health centers. Therefore, we defined this indicator variable as those households residing in a community with either a private or public health center or both. We also included in the set of covariates the distance to an office of the Savings and Credit Co-operatives (SACCOs), because the SACCOs play an important role in providing microfinance to the poor in Tanzania and thus it is also a variable of interest.

Because the time horizon involved in Table 2 is much shorter than Table 1, the changes in the reported variables during the observation period are generally small or moderate. However, there are a few variables that have changed relatively rapidly. First, the time that it takes to get water during the dry season, which includes the time to travel between the location of residence and the location water source as well as the waiting time to get water at the water source, has dropped substantially. The availability of electricity

Table 2: Key summary statistics for the TZNPS data.

Survey round	TZNPS 2008/09		TZNPS 2010/11		TZNPS 2012/13	
Variable	Mean	(s.d.)	Mean	(s.d.)	Mean	(s.d.)
Household size	5.217	(2.875)	5.557	(3.080)	5.537	(3.085)
Head has no job	0.021	(0.142)	0.031	(0.172)	0.039	(0.193)
Head's activity is farming/fishing	0.746	(0.435)	0.727	(0.446)	0.718	(0.450)
Head completed primary educ.	0.553	(0.497)	0.548	(0.498)	0.556	(0.497)
Electrified household	0.106	(0.308)	0.123	(0.328)	0.133	(0.339)
Distance to daily market (km)	6.090	(17.231)	9.538	(19.206)	12.050	(34.513)
Time to water in dry season (hrs)	1.425	(1.773)	1.209	(1.843)	1.144	(1.724)
Distance to major road (km)	16.914	(19.995)	16.809	(19.980)	16.801	(20.038)
Health center in community	0.473	(0.499)	0.506	(0.500)	0.551	(0.498)
Distance to SACCOs office (km)	6.070	(16.791)	3.741	(12.559)	6.678	(13.871)
Urban area	0.230	(0.421)	0.237	(0.425)	0.246	(0.431)
Head count index for z =\$2	0.6345		0.6697		0.6306	
Head count index for z =\$1.25	0.3378		0.3801		0.3601	
# Obs	2,355		2,355		2,355	

Note: All the figures are calculated with the household weights for TZNPS 2008/09.

at home and the availability of health center in the community of residence have also improved. On the other hand, the distance to daily market has substantially increased. This may reflect the relevance of daily market to the daily life of people in Tanzania has rapidly changed in recent years.

As with Table 1, the bottom part of Table 2 reports the head count index of poverty under alternative poverty lines. As this table shows, poverty has worsened during the period between TZNPS 2008/09 and TZNPS 2010/11. However, the situation has improved in the period between TZNPS 2010/11 and TZNPS 2012/13.

In the next section, the variables included in Tables 1 and 2 are used as covariates for our decomposition analysis. While the choice of covariates is admittedly arbitrary, these covariates plausibly affect poverty. Because their relative importance for poverty is not obvious, we apply the decomposition analysis developed in Section 3 to find which characteristics are particularly important for explaining observed poverty changes in our samples.

5 Results

Table 3 provides the regression results for the KHDS dataset. KHDS 1991 and KHDS 2004 columns provide the OLS estimates of β_{t_0} and β_{t_1}. The last column ("Pooled") provides the coefficient estimates for the fixed-coefficient model estimated with KHDS 1991 and KHDS 2004 samples pooled together. Whether \$2 or \$1.25 poverty line is used, all the coefficients are identical because the effect of the choice of poverty line is absorbed by the constant term.

Table 3 shows that most of the reported coefficients are relatively stable despite the long time horizon involved and the signs of all covariates remain unchanged for the two survey rounds. The regression results indicate that larger households and households headed by a jobless person or a person whose main activity is farming or fishing tend to have a lower standards of living, other things being equal. Neither completion of primary schooling nor whether there is a health center in the community residence appears to matter much for the standards of living.

There are two coefficients that exhibited a notable change. First, the coefficients on the indicator variable for electrified households have changed substantially. Its coefficient is insignificant for the KHDS 1991 survey but became both economically and statistically significant for the KHDS 2004 survey. Second, the distance to daily market is significant for the KHDS 1991 survey but not so for the 2004 survey. Therefore, the relative importance of electricity and market access for poverty reduction may have reversed over time.

Tables 4 gives the regression results for the TZNPS data. Because the time between the two contiguous waves is only about two years, the coefficients are generally similar across rounds. For each covariate, the difference in its coefficient between two contiguous waves is mostly below twice the standard error of the coefficient for either wave. While the signs for distance to major road and health center in community change over time, they are at best marginally significant statistically and close to zero.

The coefficients reported in the first three rows of Table 4 are all significant and negative as expected from Table 3. However, unlike the KHDS dataset, the indicator variable for the household head's completion of primary education is positive and significant for the TZNPS dataset. The indicator variable for electrified households is also positive and significant. The time to water in dry season is negative and significant for TZNPS 2008/09 and TZNPS 2012/13, suggesting that the distance to water is highly negatively related to the standards of living. Table 4 also suggests that urban residents tend to enjoy a higher standard of living.

Table 5 provides a summary of various decomposition results, where the estimated coefficients $\hat{\beta}_{t_0}$ and $\hat{\beta}_{t_1}$ are, respectively, taken from the estimates for KHDS 1991 and KHDS 2004 reported in Table 3. The initial poverty measure P_{t_0} and terminal poverty measure P_{t_1} correspond to the poverty measure calculated with the KHDS 1991 and KHDS 2004 samples, respectively. Table 5 shows that the poverty gap index (FGT_1) under the moderate poverty line of \$2 a day dropped from 38.71 percent to 35.34 percent between 1991 and 2004 in the KHDS sample. When the extreme poverty line of \$1.25 is used, the poverty gap has also dropped, though the drop is much smaller. A similar

Table 3: Ordinary least-squares estimates of β for KHDS data.

Survey round	KHDS 1991		KHDS 2004		Pooled	
Variable	Coef.	(s.e.)	Coef.	(s.e.)	Coef.	(s.e.)
Household size	-0.023^{***}	(0.008)	-0.041^{***}	(0.012)	-0.031^{***}	(0.007)
Head has no job	-0.670^{***}	(0.236)	-0.396^{**}	(0.155)	-0.414^{***}	(0.122)
Head's activity is farming/fishing	-0.395^{***}	(0.080)	-0.471^{***}	(0.092)	-0.426^{***}	(0.062)
Head completed primary educ.	0.045	(0.058)	0.038	(0.066)	0.056	(0.043)
Electrified household	0.093	(0.193)	1.161^{***}	(0.150)	0.874^{***}	(0.117)
Distance to daily market (km)	-0.009^{***}	(0.002)	-0.001	(0.002)	-0.002^{**}	(0.001)
Motorable road in community	0.226^{*}	(0.122)	0.189	(0.162)	0.166	(0.101)
Health center in community	-0.063	(0.085)	-0.114	(0.118)	-0.060	(0.072)
R^2	0.1285		0.2468		0.1789	
# Obs	430		430		860	

Note: *, **, *** respectively represent statistical significant at 1, 5, and 10 percent levels respectively. Standard errors are reported in parentheses. The dependent variable is the logarithmic real consumption per capita normalised by the poverty line. A constant term is included in the regression (not reported).

Table 4: Ordinary least-squares estimates of β for TZNPS data.

Survey round	TZNPS 2008/09		TZNPS 2010/11		TZNPS 2012/13		Pooled	
Variable	Coef.	(s.e.)	Coef.	(s.e.)	Coef.	(s.e.)	Coef.	(s.e.)
Household size	-0.122^{***}	(0.010)	-0.111^{***}	(0.009)	-0.123^{***}	(0.010)	-0.119^{***}	(0.005)
Head has no job	-0.484^{**}	(0.199)	-0.750^{***}	(0.167)	-0.849^{***}	(0.169)	-0.741^{***}	(0.101)
Head's activity is farming/fishing	-0.836^{***}	(0.085)	-0.741^{***}	(0.079)	-0.722^{***}	(0.087)	-0.763^{***}	(0.048)
Head has completed primary educ.	0.224^{***}	(0.058)	0.264^{***}	(0.056)	0.281^{***}	(0.064)	0.257^{***}	(0.034)
Electrified household	1.396^{***}	(0.104)	1.360^{***}	(0.095)	1.226^{***}	(0.105)	1.327^{***}	(0.058)
Distance to daily market (km)	0.005^{***}	(0.002)	0.001	(0.001)	-0.001	(0.001)	0.000	(0.001)
Time to water in dry season (hrs)	-0.033^{**}	(0.016)	-0.019	(0.015)	-0.042^{**}	(0.018)	-0.029^{***}	(0.009)
Distance to major road (km)	0.001	(0.001)	-0.001	(0.001)	0.002	(0.002)	0.001	(0.001)
Health center in community	0.102^{*}	(0.055)	0.038	(0.054)	-0.012	(0.060)	0.049	(0.032)
Distance to SACCOS office (km)	-0.005^{**}	(0.002)	-0.003	(0.002)	-0.003	(0.002)	-0.003^{**}	(0.001)
Urban area	0.218^{**}	(0.089)	0.111	(0.079)	0.293^{***}	(0.088)	0.216^{***}	(0.049)
R^2	0.3141		0.2989		0.2770		0.2932	
# Obs	2355		2355		2355		7065	

Note: $^{*}, ^{**}, ^{***}$ respectively represent statistical significant at 1, 5, and 10 percent levels respectively. Standard errors are reported in parentheses. The dependent variable is the logarithmic real consumption per capita normalised by the poverty line. A constant term is included in the regression (not reported).

Table 5: Summary of various decomposition results for KHDS dataset under a flexible-coefficient model.

Measure	z	P_{t_0}	P_{t_1}	Δ	X	S	R
FGT_1	2	0.3871	0.3534	-0.0337	-0.0044	-0.0230	-0.0063
	1.25	0.1660	0.1603	-0.0057	-0.0010	-0.0218	0.0171
FGT_2	2	0.2039	0.1900	-0.0138	-0.0012	-0.0192	0.0066
	1.25	0.0703	0.0694	-0.0009	0.0006	-0.0143	0.0128
Watts	2	0.5735	0.5291	-0.0444	-0.0040	-0.0471	0.0066
	1.25	0.2244	0.2174	-0.0071	0.0004	-0.0355	0.0280

Table 6: Summary of various decomposition results based for KHDS dataset under a fixed-coefficient model.

Measure	z	P_{t_0}	P_{t_1}	Δ	X	R
FGT_1	2	0.3871	0.3534	-0.0337	-0.0122	-0.0215
	1.25	0.1660	0.1603	-0.0057	-0.0088	0.0031
FGT_2	2	0.2039	0.1900	-0.0138	-0.0086	-0.0053
	1.25	0.0703	0.0694	-0.0009	-0.0044	0.0035
Watts	2	0.5735	0.5291	-0.0444	-0.0206	-0.0238
	1.25	0.2244	0.2174	-0.0071	-0.0120	0.0050

pattern is observed for the poverty severity index (FGT_2) and the Watts poverty measure as well. The observed change in poverty is reported in the fifth column (Δ) of Table 6. The remaining columns labeled X, S, and R report the covariate component, structural component, and residual component, respectively.

Whether the poverty line is set at \$2 or \$1.25, the contribution of the structural component is similar regardless of the choice of the poverty measure. Further, when the extreme poverty line is used, the size of the residual component tends to be larger in both absolute terms and relative terms compared with the total observed change in poverty. However, this observation depends on the estimation method used. For example, when we use a fixed-coefficient model (the coefficients reported in "Pooled" column in Table 3 are used), the residual under the moderate poverty line is much larger than under the extreme poverty line as shown in Table 6. Note that Table 6 does not contain the structural component as it is identically equal to zero under a fixed-coefficient model.

Table 7 provides the details of poverty decomposition result when the poverty line is set at \$2 a day for the poverty gap index (FGT_1), poverty severity index (FGT_2), and Watts poverty measure. For each poverty measure, we report the covariate component X and structural component S as well as the sum of these two components ($X + S$). The subtotal reported towards the bottom of the table provides the sum of each column above this row. Thus, Table 7 shows that the observed poverty change for the KHDS dataset is

Table 7: Details of the decomposition of moderate poverty for the KHDS dataset, 1991–2004.

Measure	FGT_1			FGT_2			Watts		
Variable	X	S	$X+S$	X	S	$X+S$	X	S	$X+S$
Household size	-0.0096	0.0571	0.0475	-0.0055	0.0423	0.0368	-0.0142	0.1008	0.0865
Head has no job	0.0098	-0.0040	0.0058	0.0064	-0.0030	0.0034	0.0167	-0.0073	0.0094
Head's activity is farming/fishing	-0.0098	0.0074	-0.0024	-0.0074	0.0056	-0.0018	-0.0179	0.0133	-0.0046
Head completed primary educ.	-0.0020	0.0050	0.0030	-0.0016	0.0036	0.0020	-0.0038	0.0087	0.0048
Electrified household	-0.0040	-0.0061	-0.0101	-0.0017	-0.0025	-0.0042	-0.0055	-0.0081	-0.0136
Distance to daily market (km)	0.0126	-0.0473	-0.0348	0.0098	-0.0393	-0.0295	0.0233	-0.0923	-0.0690
Motorable road in community	-0.0007	0.0515	0.0508	-0.0005	0.0376	0.0371	-0.0012	0.0901	0.0888
Health center in community	-0.0007	0.0005	-0.0002	-0.0006	0.0004	-0.0002	-0.0012	0.0008	-0.0004
Constant	0.0000	-0.0871	-0.0871	0.0000	-0.0639	-0.0639	0.0000	-0.1530	-0.1530
Subtotal	-0.0044	-0.0230	-0.0274	-0.0012	-0.0192	-0.0204	-0.0040	-0.0471	-0.0510
Residual (R)			-0.0063			0.0066			0.0066
Total			-0.0337			-0.0138			-0.0444

Note: Poverty line z =\$2. Coefficients are flexible.

mostly explained by the structural component. The residual component (R) is reported below the subtotal and this component is reasonably small relative to the total change in poverty.

Because our method involves numerical integration, we have also evaluated the computational errors by computing the sum of all the components ($\sum_j (X_j + S_j) + R$) and comparing it against the observed change in poverty directly calculated from the data. When the number of time step is sufficiently large, the sum of all the components is numerically very close to the observed change. For example, when we use 20,000 time steps to be conservative, the error does not exceed 1.0×10^{-6} and our decomposition results reported in this paper are accurate at least to the fifth decimal point.

To facilitate further understanding of Table 7, let us take the distance to a nearest daily market as an example of covariate of interest. As shown in Table 3, this covariate has a negative coefficient for both KHDS 1991 and KHDS 2004. Therefore, under the linear interpolation in eq. (10), an increase in this covariate would result in higher poverty. Indeed, as Table 1 shows, this covariate has increased on average between 1991 and 2004. As a result, all the reported poverty measures have increased because of the increasing distance to the nearest daily market.[5] For example, the poverty gap index has increased by 1.26 percentage points due to the increase in the distance to the nearest daily market.

The structural component for the distance to a daily market, on the other hand, is negative for all the reported poverty measures. This reflects the decreasing importance of the distance to a nearest daily market. Because the structural component is far larger in absolute value than the covariate component, the distance to daily market has on balance contributed to a decrease in poverty. For example, our decomposition analysis shows that 3.48 percentage points of poverty gap was reduced by this covariate.

Overall, household size and motorable road in community stand out as the factors that have contributed to the worsening of poverty in Kagera between 1991 and 2004. On the other hand, distance to daily market and electrified household contributed to poverty reduction. The biggest component of poverty reduction, however, is the systematic trend, which is captured by the change in the coefficient on the constant term. Joblessness of the household head, whether household head's main activity is farming/fishing, whether household head has completed primary school do not account much of poverty change. As reported in Table 10 in the Appendix, even when the extreme poverty line is used, the sign and relative importance of covariates are similar to those reported in Table 7.

Table 8 provides a summary of various decomposition exercises under the moderate poverty line for the TZNPS data. This table provides decomposition results for both the flexible- and fixed-coefficient models, all three poverty measures, and the period between TZNPS 2008/09 and 2010/11 as well as the period between TZNPS 2010/11 and 2012/13. Because our decomposition is subperiod additive, we also include the decomposition for

Table 8: Summary of the decomposition of moderate poverty for the TZNPS data, 2008/09–2012/13.

Measure	Coef	Period	P_{t_0}	P_{t_1}	Δ	X	S	$X+S$	R
FGT_1	Flex	2008/09-2010/11	0.2502	0.2728	0.0226	-0.0030	0.0284	0.0254	-0.0027
		2010/11-2012/13	0.2728	0.2592	-0.0137	-0.0001	-0.0159	-0.0160	0.0023
		2008/09-2012/13			0.0089	-0.0031	0.0125	0.0093	-0.0004
FGT_1	Fixed	2008/09-2010/11	0.2502	0.2728	0.0226	0.0005		0.0005	0.0221
		2010/11-2012/13	0.2728	0.2592	-0.0137	-0.0031		-0.0031	-0.0106
		2008/09-2012/13			0.0089	-0.0026		-0.0026	0.0115
FGT_2	Flex	2008/09-2010/11	0.1257	0.1405	0.0149	-0.0009	0.0183	0.0174	-0.0025
		2010/11-2012/13	0.1405	0.1359	-0.0047	0.0017	-0.0100	-0.0083	0.0036
		2008/09-2012/13			0.0102	0.0008	0.0083	0.0091	0.0011
FGT_2	Fixed	2008/09-2010/11	0.1257	0.1405	0.0149	0.0016		0.0016	0.0133
		2010/11-2012/13	0.1405	0.1359	-0.0047	0.0003		0.0003	-0.0050
		2008/09-2012/13			0.0102	0.0019		0.0019	0.0083
Watts	Flex	2008/09-2010/11	0.3614	0.3999	0.0385	-0.0034	0.0462	0.0428	-0.0044
		2008/09-2010/11	0.3999	0.3855	-0.0143	0.0026	-0.0256	-0.0230	0.0086
		2008/09-2012/13			0.0241	-0.0007	0.0206	0.0198	0.0043
Watts	Fixed	2010/11-2012/13	0.3614	0.3999	0.0385	0.0026		0.0026	0.0358
		2010/11-2012/13	0.3999	0.3855	-0.0143	-0.0015		-0.0015	-0.0128
		2008/09-2012/13			0.0241	0.0011		0.0011	0.0230

Note: Poverty line z =2.

the period between 2008/09 and 2012/13, which is obtained simply as a sum of the decomposition for the two sub-periods. For example, the covariate component for the poverty is -0.3 percentage points for the period between 2008/09 and 2010/11 and -0.01 percentage points for the period between 2010/11 and 2012/13. Therefore, the sum, or -0.31 percentage points, is the change in observed poverty accounted for by the changes in covariates.

Table 8 shows that poverty in Tanzania has increased between 2008/09 and 2010/11 but decreased between 2010/11 and 2012/13, regardless of the poverty measure used. It also shows that the poverty measures have worsened in the period between 2008/09 and 2012/13. As with the case of Kagera between 1991 and 2004, much of the change is explained by the structural change when a flexible-coefficient model is used. Given that the coefficients are relatively stable over time, this result may appear surprising. However, it does not have to be because the change in structural parameter affects everyone in the sample and thus a small change in β can translate into a relatively large change in poverty measures.

Table 8 also shows that the importance of the residual component is much larger when a fixed-coefficient model is used. Because the residual component provides little insight into the underlying cause of poverty change, the results for KHDS and TZNPS data both indicate that the fixed-coefficient model is not only potentially misleading but it may also be much less useful than the flexible-coefficient model. As Table 11 in the Appendix shows, this point is true even when an extreme poverty line is used.

Table 9 provides detailed decomposition results for the poverty gap index under a moderate poverty line. We report the covariate component, structural component, and their sum for the period between 2008/09 and 2010/11 in columns (A), (B), and (C), respectively. The corresponding results for the period between 2010/11 and 2010/11 are reported in columns (D)–(F). In column (G), we report the total contribution of each covariate to the poverty change in the period between 2008/09 and 2012/13, which is calculated as a sum of columns (C) and (F). All these calculations are based on a flexible-coefficient model. Column (H) reports a result for the fixed coefficient model.

Table 9 shows that the indicator variable for the household heads whose main activity is farming/fishing is one of the main drivers for poverty reduction. In addition, urbanization has also contributed to poverty reduction. Because of the structural change in the period between 2010/11 and 2012/13 the indicator for the primary-completed head has also negatively contributed to the poverty change in the entire period between 2008/09 and 20012/13.

However, these effects have been offset by a variety of factors. For example, factors such as the household size, time to water in dry season, and distance to a SACCOs office all contributed to poverty in the opposite direction of the observed change and each of

Table 9: Details of the decomposition of the poverty gap index (FGT_1) for the TZNPS data under the moderate poverty line, 2008/09–2012/13.

Period	2008/09-2010/11			2010/11-2012/13			2008/09-2012/13	
	X	S	X+S	X	S	X+S	[(C)+(F)]	Fixed Coef
Variable	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Household size	0.0000	0.0103	0.0103	0.0080	-0.0144	-0.0064	0.0039	0.0111
Head has no job	0.0006	-0.0001	0.0005	0.0012	0.0009	0.0021	0.0026	0.0034
Head's activity is farming/fishing	-0.0017	-0.0147	-0.0164	-0.0047	-0.0095	-0.0142	-0.0306	-0.0089
Head has completed primary educ.	-0.0006	0.0025	0.0018	0.0005	-0.0100	-0.0095	-0.0076	0.0000
Electrified household	-0.0011	0.0008	-0.0003	-0.0011	-0.0001	-0.0012	-0.0015	-0.0037
Distance to daily market (km)	-0.0001	0.0021	0.0020	-0.0026	0.0093	0.0067	0.0087	-0.0008
Time to water in dry season (hrs)	-0.0001	0.0075	0.0074	-0.0010	-0.0020	-0.0030	0.0044	-0.0017
Distance to major road (km)	0.0000	-0.0120	-0.0120	0.0000	0.0059	0.0059	-0.0061	0.0000
Health center in community	0.0001	0.0010	0.0011	-0.0004	0.0094	0.0090	0.0101	-0.0011
Distance to SACCOS office (km)	0.0032	0.0039	0.0071	-0.0020	-0.0015	-0.0035	0.0036	0.0005
Urban area	-0.0002	-0.0056	-0.0059	-0.0010	0.0000	-0.0010	-0.0069	-0.0016
Constant	0.0000	-0.0116	-0.0116	0.0000	0.0402	0.0402	0.0286	0.0000
Subtotal	-0.0001	-0.0159	-0.0160	-0.0030	0.0284	0.0254	0.0093	-0.0026
Residual (R)			0.0023			-0.0027	-0.0004	0.0115
Total			-0.0137			0.0226	0.0089	0.0089

Note: Poverty line z =$2. Coefficients are flexible except for the calculation of Column (H).

these factors overall contributed to the worsening of poverty in the entire period between 2008/09 and 2012/13, even though the size of their contributions is small. Because the coefficient on the health center in community is falling while the health centers are getting more accessible, the variable for the health center in the community of residence has actually contributed to an increase in poverty in our decomposition.

Finally, as the comparison between columns (G) and (H) make clear, the relative importance of various covariates are generally different between flexible- and fixed-coefficient models. Therefore, making appropriate assumptions is crucial for our decomposition analysis. The choice of poverty measures, on the other hand, does not appear to be as important because the observations made above are generally applicable to the poverty severity index and Watts poverty measure as reported in Tables 12 and 13 in the Appendix.

6 Discussion

In this paper, we have developed a new method of dynamic poverty decomposition. The novelty of our method is that it is based on a regression model, which allows us to find how much contribution is made by each covariate. Unlike the standard approach to poverty decomposition, which typically utilises the population-level characteristics to explain the observed change in poverty, our method uses household-level characteristics to account for the poverty change. Therefore, we can associate the observed poverty change directly with household characteristics and describe the decomposition results in an intuitive manner. Furthermore, our decomposition also satisfies both the time-reversion consistency and subperiod additivity because it is derived from the integration of partial derivatives.

However, there are at least three limitations in our method. First, because our method is model-based, we will be always left with the residual component R, which is the part of the observed poverty change that cannot be explained by X or S. In particular, when the explanatory power of the model is weak, most of the changes will be explained by the changes in the coefficient on the constant term (i.e., intercept) and the residual term, where the former essentially describes the change in mean and the latter describes the change in inequality. On the other hand, if the residual component is small relative to other components, the observed change in poverty can be ascribed to covariates. Therefore, having a good consumption model is important for our method to yield informative decomposition results.

Second, our decomposition method requires the availability of a panel dataset and assume that the set of households in the population is stable over time. As a result, the results of our decomposition analysis presented in Section 5 do not necessarily extend to those households in which the household head has changed over the observation period. For example, households that have fallen into poverty because of the death of the current household head will not be appropriately accounted for by our decomposition method. Similarly, our analysis excludes those households that have been lifted out of poverty because the head has migrated abroad to obtain a better job. Formation of new households due to marriage and disappearance or split of existing households are also ignored in the current method.

Finally, the current method does not take into account the sampling errors. While it is straightforward to artificially incorporate the sampling error by bootstrapping the samples at time $t = t_0$ and $t = t_1$, the interpretation of the standard errors computed in this way is ambiguous and may not be relevant. Further, the sampling errors are often omitted in decomposition studies. For these reasons and for the sake of simplicity of presentation, we chose not to incorporate standard errors.

Our method was applied to two separate datasets from Tanzania, both of which include various indicators of interest, including electrification, market access, road access, and the presence of health centers in the community. In our applications, the structural component S turns out to be overall more important than the effect of covariate component X. The relative importance of covariates vary over time and across different locations. For example, in the Kagera district between 1991 and 2004, we find that the market and road access variables are among the most important factors contributing to the observed poverty change, though the signs of their contributions are opposite. On the other hand, the presence of health center in the community did not contribute much to the observed poverty change.

In contrast to these observations, the presence of health center in the community is found to be more important than the road and market access variables in the TZNPS dataset for the period between 2008/09 and 2012/13. Further, unlike the decomposition based on the KHDS dataset, the indicator variables for the household head whose main activity is farming/fishing and for the household head who has completed primary education are comparatively more important for the TZNPS dataset. These results show that both relative and absolute importance of various factors in explaining the observed poverty change may vary over time

Because we are primarily interested in poverty, the logarithmic consumption per capita was used as the left-hand-side variable in our regression models. However, the decomposition approach proposed in this paper is general and applicable to the cases where an alternative left-hand-side variable is used. For example, if we use the logarithmic wage as a left-hand-side variable and gender or ethnicity as covariates, we have a decomposition similar to the Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973), which allows us to analyze the changes in labor market discrimination over time (See Fortin et al. (2011) for a review of the Oaxaca-Blinder decomposition and other related decomposition methods).

Our approach may also prove useful in a variety of settings. In particular, when we wish to simultaneously evaluate the impact of various policies on poverty, our method is particularly valuable, because it allows researchers to identify the poverty change that can be attributed to each of the multiple policies that have been implemented already. We can do so by including various policy variables in the set of regressors and apply our decomposition method. The only major requirement in this approach is that the mix of policies to be evaluated must vary over the households.

This point makes a clear contrast with the randomised control trial (RCT) approach, which can also be used to evaluate the impact of a set of policies on poverty. While RCT studies generally rest on a less restrictive set of assumptions than our approach, it is costly to evaluate the impacts of various mixtures of policies because the number of control groups increases geometrically as the number of policies to be jointly evaluated increases. The RCT approach also requires researchers to design a study before policies are implemented. As a result, it is not possible to evaluate the policies that have already been implemented by the RCT approach.[6] Therefore, our approach allows us to evaluate policies that cannot be randomised for operational or ethical reasons and serves as a practical and economical alternative to RCT for evaluating the impacts of policies on poverty.

Appendix: Additional tables

Table 10 is the same as Table 7 except that the extreme poverty line of \$1.25 is used instead of the moderate poverty line of \$2. Similarly, Table 11 is the same as Table 8 except that the extreme poverty line is used. Tables 12 and 13 are the same as Table 9 except that the poverty measures used are poverty severity index and Watts poverty measure, respectively, instead of the poverty gap index.

Table 10: Details of the decomposition of extreme poverty for the KHDS dataset, 1991–2004.

Measure	FGT_1			FGT_2			Watts		
Variable	X	S	$X+S$	X	S	$X+S$	X	S	$X+S$
Household size	-0.0041	0.0452	0.0411	-0.0018	0.0213	0.0196	-0.0052	0.0630	0.0578
Head has no job	0.0064	-0.0034	0.0031	0.0035	-0.0016	0.0019	0.0097	-0.0049	0.0048
Head's activity is farming/fishing	-0.0082	0.0061	-0.0020	-0.0044	0.0030	-0.0014	-0.0120	0.0086	-0.0034
Head completed primary educ.	-0.0018	0.0038	0.0020	-0.0009	0.0018	0.0008	-0.0026	0.0052	0.0026
Electrified household	-0.0013	-0.0016	-0.0029	-0.0006	-0.0004	-0.0010	-0.0017	-0.0019	-0.0036
Distance to daily market (km)	0.0094	-0.0444	-0.0349	0.0054	-0.0247	-0.0193	0.0143	-0.0666	-0.0522
Motorable road in community	-0.0006	0.0402	0.0396	-0.0004	0.0189	0.0185	-0.0009	0.0558	0.0549
Health center in community	-0.0009	0.0004	-0.0005	-0.0003	0.0001	-0.0002	-0.0011	0.0005	-0.0006
Constant	0.0000	-0.0683	-0.0683	0.0000	-0.0326	-0.0326	0.0000	-0.0954	-0.0954
Subtotal	-0.0010	-0.0218	-0.0229	0.0006	-0.0143	-0.0137	0.0004	-0.0355	-0.0351
Residual (R)			0.0171			0.0128			0.0280
Total			-0.0057			-0.0009			-0.0071

Note: Poverty line z =\$1.25. Coefficients are flexible.

Table 11: Summary of the decomposition of extreme poverty for the TZNPS data, 2008/09–2012/13.

Measure	Coef	Period	P_{t_0}	P_{t_1}	Δ	X	S	$X+S$	R
FGT_1	Flex	2008/09-2010/11	0.0987	0.1124	0.0138	0.0005	0.0176	0.0181	-0.0043
		2010/11-2012/13	0.1124	0.1100	-0.0025	0.0032	-0.0094	-0.0062	0.0037
		2008/09-2012/13			0.0113	0.0037	0.0082	0.0120	-0.0006
FGT_1	Fixed	2008/09-2010/11	0.0987	0.1124	0.0138	0.0054		0.0054	0.0084
		2010/11-2012/13	0.1124	0.1100	-0.0025	0.0035		0.0035	-0.0060
		2008/09-2012/13			0.0113	0.0089		0.0089	0.0025
FGT_2	Flex	2008/09-2010/11	0.0403	0.0477	0.0074	0.0002	0.0079	0.0081	-0.0007
		2010/11-2012/13	0.0477	0.0489	0.0012	0.0021	-0.0043	-0.0022	0.0034
		2008/09-2012/13			0.0086	0.0023	0.0036	0.0059	0.0027
FGT_2	Fixed	2008/09-2010/11	0.0477	0.0489	0.0012	0.0029		0.0029	-0.0017
		2010/11-2012/13	0.0403	0.0477	0.0074	0.0024		0.0024	0.0050
		2008/09-2012/13			0.0086	0.0052		0.0052	0.0034
Watts	Flex	2008/09-2010/11	0.1306	0.1513	0.0207	0.0008	0.0241	0.0249	-0.0041
		2008/09-2010/11	0.1513	0.1517	0.0004	0.0051	-0.0129	-0.0078	0.0082
		2008/09-2012/13			0.0211	0.0059	0.0112	0.0171	0.0041
Watts	Fixed	2010/11-2012/13	0.1306	0.1513	0.0207	0.0074		0.0074	0.0133
		2010/11-2012/13	0.1513	0.1517	0.0004	0.0060		0.0060	-0.0056
		2008/09-2012/13			0.0211	0.0135		0.0135	0.0077

Note: Poverty line z =\$1.25.

Table 12: Details of the decomposition of the poverty severity index (FGT_2) for the TZNPS data under the moderate poverty line, 2008/09–2012/13.

Period	2008/09-2010/11			2010/11-2012/13			2008/09-2012/13	
	X	S	X+S	X	S	X+S	[(C)+(F)]	Fixed Coef
Variable	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Household size	0.0007	0.0069	0.0077	0.0056	-0.0096	-0.0040	0.0036	0.0088
Head has no job	0.0007	0.0000	0.0006	0.0007	0.0005	0.0012	0.0018	0.0025
Head's activity is farming/fishing	-0.0010	-0.0099	-0.0109	-0.0026	-0.0064	-0.0090	-0.0199	-0.0052
Head has completed primary educ.	-0.0005	0.0015	0.0011	0.0004	-0.0061	-0.0057	-0.0047	0.0001
Electrified household	-0.0005	0.0003	-0.0002	-0.0005	0.0000	-0.0005	-0.0007	-0.0017
Distance to daily market (km)	-0.0001	0.0014	0.0013	-0.0017	0.0061	0.0044	0.0057	-0.0006
Time to water in dry season (hrs)	-0.0001	0.0051	0.0050	-0.0007	-0.0013	-0.0020	0.0029	-0.0012
Distance to major road (km)	0.0000	-0.0083	-0.0083	0.0000	0.0040	0.0040	-0.0043	0.0000
Health center in community	0.0001	0.0007	0.0007	-0.0002	0.0062	0.0059	0.0067	-0.0007
Distance to SACCOS office (km)	0.0025	0.0027	0.0052	-0.0013	-0.0010	-0.0022	0.0030	0.0007
Urban area	-0.0001	-0.0028	-0.0029	-0.0006	0.0000	-0.0006	-0.0035	-0.0009
Constant	0.0000	-0.0076	-0.0076	0.0000	0.0260	0.0260	0.0184	0.0000
Subtotal	0.0017	-0.0100	-0.0083	-0.0009	0.0183	0.0174	0.0091	0.0019
Residual (R)			0.0036			-0.0025	0.0011	0.0083
Total			-0.0047			0.0149	0.0102	0.0102

Note: Poverty line z =\$2. Coefficients are flexible except for the calculation of Column (H).

Table 13: Details of the decomposition of the Watts poverty measure for the TZNPS data under the moderate poverty line, 2008/09–2012/13.

Period	2008/09-2010/11			2010/11-2012/13			2008/09-2012/13	
	X	S	X+S	X	S	X+S	[(C)+(F)]	Fixed Coef
Variable	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Household size	0.0013	0.0173	0.0186	0.0136	-0.0240	-0.0104	0.0082	0.0206
Head has no job	0.0013	-0.0001	0.0011	0.0019	0.0014	0.0033	0.0044	0.0059
Head's activity is farming/fishing	-0.0025	-0.0248	-0.0273	-0.0071	-0.0158	-0.0228	-0.0502	-0.0135
Head has completed primary educ.	-0.0011	0.0039	0.0028	0.0010	-0.0158	-0.0148	-0.0121	0.0001
Electrified household	-0.0016	0.0011	-0.0005	-0.0015	-0.0001	-0.0016	-0.0021	-0.0051
Distance to daily market (km)	-0.0002	0.0036	0.0034	-0.0042	0.0153	0.0111	0.0144	-0.0013
Time to water in dry season (hrs)	-0.0003	0.0128	0.0125	-0.0018	-0.0033	-0.0051	0.0074	-0.0031
Distance to major road (km)	0.0000	-0.0205	-0.0205	-0.0001	0.0099	0.0098	-0.0107	0.0001
Health center in community	0.0002	0.0017	0.0019	-0.0006	0.0155	0.0150	0.0168	-0.0017
Distance to SACCOS office (km)	0.0059	0.0067	0.0126	-0.0032	-0.0024	-0.0056	0.0069	0.0014
Urban area	-0.0003	-0.0080	-0.0083	-0.0015	0.0000	-0.0015	-0.0097	-0.0022
Constant	0.0000	-0.0192	-0.0192	0.0000	0.0655	0.0655	0.0463	0.0000
Subtotal	0.0026	-0.0256	-0.0230	-0.0034	0.0462	0.0428	0.0198	0.0011
Residual (R)			0.0086			-0.0044	0.0043	0.0230
Total			-0.0143			0.0385	0.0241	0.0241

Note: Poverty line z =\$2. Coefficients are flexible except for the calculation of Column (H).

Notes

[1]For the sake of simplicity, we simply use AD to represent the length of line segment AD. We hereafter use similar notations.

[2]MKUKUTA is a Kiswahili acronym for the National Strategy for the National Strategy for Growth and Reduction of Poverty. The first phase started in 2005 and finished in 2010. The second phase started in 2010 and finishes in 2015.

[3]Both the KHDS and TZNPS original datasets and relevant documentations are available from the Living Standards Measurement Surveys (LSMS) website See, `http://go.worldbank.org/IFS9WG7EO0`.

[4]For each wave, the data collection starts in October and lasts twelve month. Therefore, we use the average CPI for the first and second years of survey weighted by the ratio of 1:3.

[5]Logically speaking, poverty measures could drop even when the covariate increases on average and its coefficient is (always) negative during the period of decomposition. This is because poverty measures are non-linear transformation of consumption per capita and insensitive to the upper tail of distribution.

[6] When there is an "exogenous" change, we may be able to adopt the quasi-experimental design but the applicability of such a design is generally limited.

References

Ainsworth, M. and Semali, I. (2000). The impact of adult deaths on children's health in northwestern Tanzania. World Bank Policy Research Working Paper 2266, World Bank.

Banerjee, A., Duflo, E., and Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in China. NBER Working Paper 17897, National Bureau of Economic Research.

Beegle, K., De Weerdt, J., and Dercon, S. (2006). *Kagera Health and Development Survey 2004: Basic Information Document*. World Bank, Washington, DC.

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8:436–455.

Datt, G. and Ravallion, M. (1992). Growth and redistribution components of changes in poverty measures: A decomposition with applications to Brazil and India in the 1980s. *Journal of Development Economics*, 38:275–295.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from South Africa. *American Economic Review*, 101:3078–3108.

Duflo, E. and Pande, R. (2007). Dams. *Quarterly Journal of Economics*, 122(2):601–646.

Fernald, J. (1999). Roads to prosperity?: Assessing the link between public capital and productivity. *American Economic Review*, 89(3):619–638.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, volume 4A, chapter 1, pages 1–102. Elsevier.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52(3):761–766.

Fujii, T. (2014). Dynamic poverty decomposition analysis: An application to the Philippines. ADBI Working Paper Series 466, Asian Development Bank Institute.

Gramlich, E. (1994). Infrastructure investment: A review essay. *Journal of Economic Literature*, 32:1176–1196.

Grogan, L. and Sadanand, A. (2013). Rural electrification and employment in poor countries: Evidence from Nicaragua. *World Development*, 43(0):252–265.

Grootaert, C. (1995). Structural change and poverty in Africa: A decomposition analysis for Côte d'Ivoire. *Journal of Development Economics*, 47:375–401.

Jain, L. and Tendulkar, S. (1990). The role of growth and distribution in the observed change in head-count ratio-measure of poverty: A decomposition exercise for India. *Indian Economic Review*, 25(2):165–205.

Kakwani, N. and Subbarao, K. (1990). Rural poverty and its alleviation in India. *Economic and Political Weekly*, 25(13):A2–A16.

Kolenikov, S. and Shorrocks, A. (2005). A decomposition analysis of regional poverty in Russia. *Review of Development Economics*, 9(1):25–46.

Lipscomb, M., Mobarak, M., and Barham, T. (2013). Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics*, 5(2):200–231.

National Bureau of Statistics (2010). *Basic Information Document: National Panel Survey 2008-2009*. Downloadable from `http://go.worldbank.org/EJMAC1YDY0`.

National Bureau of Statistics (2012). *Basic Information Document: National Panel Survey 2010-2011*. Downloadable from `http://go.worldbank.org/EJMAC1YDY0`.

National Bureau of Statistics (2014). *Basic Information Document: National Panel Survey 2012-2013*. Downloadable from `http://go.worldbank.org/EJMAC1YDY0`.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Econoic Review*, 14:693–709.

Peters, J., Vance, C., and Harsdorff, M. (2011). Grid extension in rural Benin: Micro-manufacturers and the electrification trap. *World Development*, 39(5):773–783.

Ravallion, M. and Huppi, M. (1991). Measuring changes in poverty: A methodological case study of Indonesia during an adjustment period. *World Bank Economic Review*, 5(1):57–82.

Röller, L. and Waverman, L. (2001). Telecommunications infrastructure and economic development: A simultaneous approach. *American Economic Review*, 91:909–923.

Rud, J. (2012). Electricity provision and industrial development: Evidence from India. *Journal of Development Economics*, 97:352–367.

Sahn, D. and Stifel, D. (2000). Poverty comparisons over time and across countries in Africa. *World Development*, 28(12):2123–2155.

Shorrocks, A. (2013). Decomposition procedures for distributional analysis: a unified framework based on Shapley value. *Journal of Economic Inequality*, 11:99–126.

Straub, S. (2008). Infrastructure and development: A critical appraisal of the macro level literature. World Bank Policy Research Working Paper 4590, The World Bank.

Watts, H. (1968). An economic definition of poverty. In Moynihan, D., editor, *On Understanding Poverty*, pages 316–329. Basic Books, New York.

World Bank (2004). *User's Guide to the Kagera Health and Development Survey Datasets*. Development Research Group, World Bank, Washington, DC.

World Bank (2012). *Kagera Health and Development Survey: Consumption Expenditure Data*. World Bank, Washington, DC.

Zheng, B. (1993). An axiomatic characterization of the Watts poverty index. *Economics Letters*, 42:81–86.