Smoothing Sporadic Poverty and Inequality Estimates: Pakistan, 1985-2016

Jamal, Haroon

Social Policy and Development Centre (SPDC)

October 2018

Smoothing Sporadic Poverty and Inequality Estimates:
Pakistan, 1985-2016

Haroon Jamal

Social Policy and Development Centre (SPDC)
Karachi-Pakistan

October 22, 2018

**Smoothing Sporadic Poverty and Inequality Estimates:
Pakistan, 1985-2016**

ABSTRACT

Poverty and income inequality is estimated from household surveys which provide detailed information on household income and consumption along with socio-economic characteristics. However, these surveys are conducted sporadically with irregular intervals. Thus the resultant estimates are not in the form of a continuous time-series, which is a prerequisite for a rigorous analysis of the relationship between macro variables and the estimates of poverty and inequality in the national context. To cater to the need of researchers and students, this research provides a continuous time-series of poverty incidence and the various measures of income inequality for the period 1985-2016 after applying the interpolation techniques on sporadic counts. These series are then used to explore the relationship between poverty, inequality and growth.

JEL Classification: I32
Keywords: Consistent Poverty Estimates, Inequality, Interpolation, Poverty Elasticity, Pakistan

1. Introduction

Detailed information on income and expenditure of households are required to estimate monetary or traditional poverty incidence and the extent of income inequality. In the context of Pakistan, Household Integrated Economic Survey (HIES) is the only source which provides such household data that is representative at national and regional (urban/rural) level. HIES is conducted by the Pakistan Bureau of Statistics (PBS) and includes standard and detailed income and consumption modules besides information on household characteristics.[1] The first HIES was conducted in 1963 and has been repeated periodically since then, though with irregular intervals.[2] Thus the monetary poverty and inequality estimates are available in sporadic counts for this period.

However, for conducting rigorous econometric and time-series analysis to explore the impact of microeconomic policies on poverty reduction and income distribution, methodologically consistent time-series estimates are required. In the context of Pakistan, Jamal (2006) provided interpolated times series of poverty incidence (headcount) and *Gini* coefficients for the period 1973 to 2003. Nonetheless, the observed poverty estimates used in Jamal (2006) were taken from earlier research which adopted crude and to some extent flawed methodology for measuring and updating poverty line.[3] Therefore, this research not only incorporates the new available HIES data (post-2003) but also estimates poverty line and poverty aggregates with a well-defined and consistent methodology.

Twelve household level HIES data sets during the period of 1985 to 2016 are employed for measuring poverty and inequality numbers.[4] These sporadic estimates are then interpolated to obtain continuous time-series. These series may be used as an input in the analysis of relationship between macro variables and poverty and inequality estimates by researchers and students.

This research report is organized in seven sections. Section 2 presents a brief methodology for estimating poverty line and poverty aggregates. The procedure for updating poverty line with the new consumption data is explained in Section 3. The measures of income inequality considered for this research are described in Section 4, while the subsequent section summarizes interpolation techniques. The interpolated annual estimates are then used to empirically establish the nexus between poverty, inequality and growth in the context of Pakistan. Section 6 furnishes the results of this exercise. The last section is reserved for some concluding remarks.

---

[1] Sampling design and sampling methodology are briefly described in the Appendix-1.

[2] The HIES acronym used herein refers to its updated name 'Household Integrated Economic Survey'; previously the acronym HIES was used for 'Household Income and Expenditure Survey' during 1963-1998.

[3] According to Jamal (2006), Malik (1988) generated five poverty observations during the period 1963-64 to 1984-85 based on HIES surveys. They have applied a consistent methodology to compute poverty lines for these particular years. Amjad and Kemal (1999) added three more observations for the years 1987-88, 1990-91 and 1992-93 by inflating poverty line for 1984-85 using the consumer price index. Three more observations were added by Jamal (2006) for the years 1996-97, 1998-99 and 2001-02 again by inflating poverty line for the year 1992-93.

[4] Unfortunately author has no access to detail household level raw data of four (1963-64, 1966-67, 1969-70, and 1979) HIES surveys.

2. Brief Methodology of Poverty Estimation

Among the various approaches of defining absolute monetary (income/consumption) or traditional poverty, 'calorific approach' is the most popular in developing countries due to its practicality. In almost all studies of poverty in developing countries including Pakistan, the poverty level is defined in terms of food inadequacy which is typically measured by the lack of nutritional (calorie) requirements.

This research follows the methodology used by Jamal (2002) for estimation of poverty through calorific approach. The details of various methodological options and recommended steps are provided in Jamal (2002), while a brief description of the major steps to compute the poverty line and poverty indices is furnished below.[5]

- Food quantities consumed in household are translated into calories (food energy).[6] Food Consumption Tables for Pakistan (GoP, 2001) provides proximate conversion factor for each food item.

- To compute the poverty line, minimum required calories (calorie norms) and the estimated coefficients of the Calorie-Consumption Function (CCF) are required. The CCF provides the estimates of expenditure (rupees) needed to obtain additional (marginal) one calorie.[7]

- The estimation approach of Jamal (2002), recommends 2,550 and 2,230 calories per day per adult as minimum requirement for rural and urban areas, respectively.[8] It is argued that consumption behavior, purchasing patterns, dietary habits, taste and ecology are significantly different for urban and rural inhabitants.

- The CCFs are estimated separately for urban and rural areas by regressing household per adult daily calorie consumption on total household expenditure.[9] Moreover, these

---

[5] A schematic view of various approaches to estimate poverty line is furnished in the Appendix-2.

[6] The consumption includes not only actual purchases but also self-produced and consumed items, consumption of items received as gifts, plus items provided in place of monetary compensation.

[7] For the purpose of CCF estimation, purchase of durable assets and expenses such as marriage and taxes are not included in the consumption aggregate as these are not truly reflect the current status of living standard.

[8] The Government of Pakistan does not estimate separate urban and rural poverty lines. The rural lifestyle in general requires a greater consumption of calories than the urban lifestyle. It is not irrational to assume that for any given level of income, rural households are likely to consume more calories, on average, than their urban counterparts. Thus, poverty estimates derived from official methodology using a unique poverty line for both urban and rural households underestimate rural poverty and overestimate urban poverty.

[9] Food Consumption Tables for Pakistan (2001) provide the recommended daily allowance for the Pakistani population for various age and sex composition. These requirements are used to compute adult equivalent unit for each household.

functions are estimated from the lowest quartile of distribution after ranking households for a better reflection of consumption behavior/pattern of the poor. The estimated coefficients of calorie-consumption functions are used to derive the poverty line separately for urban and rural areas. The poverty line thus reflects an estimation of total household expenditure (food plus non-food) needed to obtain the minimum required calories.

- Household poverty status (poor or non-poor) is determined by linking poverty line and household reported consumption expenditure.

- Various measures have been suggested in the literature to aggregate the individual household poverty status into a single index that may be used as a proxy for the status of a group of households (national, provincial, regional etc.) The issue in this regard primarily relates to assigning weights to different intensities of poverty. The most popular measure, namely the Head Count Index (HCI), assigns equal weights to all the poor regardless of the extent of poverty[10]. HCI or incidence of poverty reflects the proportion of households whose consumptions fall below the poverty line.

3. Inter-Temporal Comparison of Poverty

For a meaningful comparison of poverty over time, it is essential to adhere to a consistent methodology and the calorie norms. Two options are available for monitoring poverty over time; poverty line for the latest survey year may either be updated by utilizing previously estimated poverty line after adjusting with some appropriate index of inflation or it may be re-estimated with the help of new available survey data.

There are many criticisms on using Consumer Price Index (CPI) for updating the poverty line in the case of Pakistan due to its very low geographical coverage. CPI only covers major urban centres for tracking inflation and ignores price movement in rural areas and small urban locations. To circumvent this deficiency, survey based price index – Tornqvist Price Index (TPI) – is suggested as an alternative. However, it is not a problem-free option, since TPI can only incorporate homogenous goods like specific food items. Further, the household survey does not report the consumption of non-food quantities and provides only expenditures on these items. These complications make TPI an inappropriate measure of inflation. The extent of adjustment in TPI can be ascertained from the fact that TPI includes only 75 items, whereas CPI includes more than 400 items.

[10] There are several other measures, which have been suggested in the poverty literature. These measures are sensitive to distribution among the poor. A class of functional forms, which has been suggested by Foster, Greer, and Thorbeke (FGT), uses various powers of the proportional gap between the observed and the required expenditure as the weights to indicate the level of intensity of poverty (Appendix-3). For this exercise however, only headcount or poverty incidence is considered for developing interpolated poverty series.

Re-estimation of poverty line is also criticized on the ground that for monitoring and tracking poverty numbers, the bundle of goods and services should remain same and one should adjust the magnitude of the poverty line with the movement of price. However, this criticism does not seem valid if 'calorific approach' is used in deriving poverty line instead of 'basic need approach'[11] . With fixed calorie norms, it is estimated on the basis that how much expenditures are required for the particular year to obtain minimum required calories.

Thus, in the absence of any appropriate price index for inflating the previous poverty line, it is perhaps reasonable and is also preferred for this research to re-estimate the poverty line from the latest survey to circumvent the problems associated with price indices.

4. Measure of Income Inequality

The measurement of inequality is an arduous task and no single statistical measure is able to capture its myriad dimensions. However, the *Gini* Index is easily interpreted and widely used in the empirical literature of inequality. Following Kakwani (1980), the *Gini* is computed as follows:

$$Gini \equiv \left(\frac{1}{2\bar{I}}\right)\left(\frac{1}{n(n-1)}\right)\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\left|I_i - I_j\right|\right]$$

Where I_i is the value of resource indicator (here household per capita income), $\bar{I}$ is the mean value of the indicator and n is the number of households. The *Gini* Index provides a measure of resource inequality within a population. It is the most popular measure of inequality and summarizes the extent to which actual distribution of resource differs from a hypothetical distribution in which each household receives an identical share. *Gini* is a dimensionless index scaled to vary from a minimum of zero to a maximum of one; zero representing no inequality and one representing the maximum possible degree of inequality.

A limitation of the *Gini* coefficient or index as a measure of inequality is that it is most sensitive to the middle part of income distribution than to that of extremes because it depends on the rank order weights of income recipients and on the number of recipients within a given range. Thus, to capture small changes in extreme parts of income distribution (tails), several ancillary measures have been developed that focus on measuring certain types of inequality.

The Chilean economist Gabriel Palma (2011) observed that the income share of those in 5-9 deciles of income distribution is usually stable, across countries and across time. The remaining 50 percent is shared amongst the very top earners in 10th decile and those in deciles 1 to 4, but its distribution varies widely between countries and across time. This proposition led to the development of the famous Palma Ratio by two economists, Alex Cobham and Andy Sumner (2013), who state that "it gives a more accurate picture of income inequality than the *Gini* coefficient because the *Gini* is not sensitive to data in the tails, where the inequality actually

[11] See Appendix-2 for the methodological consideration and choices.

lies”. The Palma Ratio, which compares the share of income of the richest 10 percent in a society to that of the poorest 40 percent focuses on the locus of inequality, and is also sensitive to extreme inequality unlike the *Gini*.

Income Quintile Share Ratio (IQR) is another measure of income inequality which is widely used to capture extreme inequality. IQR focuses on comparing the incomes of those at the top of the income distribution to those at the bottom. It is calculated as the ratio of the average income received by the 20 percent of persons with the highest income (top quintile) to the average income received by the 20 percent of persons with the lowest income (bottom quintile).

5. Interpolation Techniques

The software SPSS (Statistical Package for the Social Sciences) is used to interpolate various series of poverty and inequality for this research. SPSS provides two procedures for interpolating; ‘curve fitting’ and ‘replacing missing values’.

Curve fitting is the process of constructing a curve that has the best fit to a series of data points. The statistical routine (CURVEFIT) in SPSS offers 11 different regression models for the estimation of curves (Exhibit 5.1). The syntax for the selected regression model produces regression statistics and predicted values along with a chart, which displays observed and predicted values.

Exhibit – 5.1
Regression Models for Curve Fitting

Type/Function	Equation
Linear	$Y=b_0+b_1t$
Logarithmic	$Y=b_0+b_1\ln(t)$
Inverse	$Y=b_0+b_1/t$
Quadratic	$Y=b_0+b_1t+b_2t^2$
Cubic	$Y=b_0+b_1t+b_2t^2+b_3t^3$
Compound	$Y=b_0b_1^t$
Power	$Y=b_0t^{b1}$
S	$Y=e^{b0+b1/t}$
Growth	$Y=e^{b0+b1t}$
Exponential	$Y=b_0e^{b1t}$
Logistic	$Y=(1/u+b_0b_1^t)-1$

Where:
- b_0 a constant
- b_n regression coefficient
- t Time value or Year
- ln the natural logarithm
- u upper-bound value for Logistic Model

Source: IBM, SPSS Statistics 20.

Due to small number of observations, it is preferred to empirically assess all these statistical models for the best fit in terms of regression statistics (t-value, F-value, sign of the coefficient, adjusted R^2 etc.) and visuals of observed and predicted values.

In contrast, statistical procedure for replacing missing values (RMV) creates new variables by copying existing variables and replacing any missing values between two observations with the estimates computed by one of several methods. SPSS offers five different methods for computing missing value: linear interpolation, mean of surrounding values, median of surrounding values, variable mean, and linear trend at that point. For this research however, linear trend is preferred for comparing model for the best fit.

Before selecting the final interpolated series of poverty and inequality, stationarity is also checked by applying the Augmented Dickey–Fuller (ADF) test.[12] Empirical work based on time series data assumes the underlying series is stationary. Stationarity of a time series is crucial for the application of various econometric techniques, especially related with forecasting. Further, non-stationary time-series may produce spurious or nonsense regression. Exhibit-5.2 describes the interpolation technique as well as value of ADF for selected series.

Exhibit – 5.2
Selected Interpolation Techniques and Test for Unit Root

				Augmented Dickey-Fuller Test (ADF)	
Series		Interpolation Technique		Test for Unit Root in:	Test-Statistics
Poverty Incidence	Overall	RMV	Trend	First Difference	-7.640
	Urban	RMV	Trend	First Difference	-8.763
	Rural	RMV	Trend	First Difference	-9.363
Gini Coefficient	Overall	Curve Fitting	Logarithmic	Level	-3.855
	Urban	Curve Fitting	Power	Level	-6.282
	Rural	Curve Fitting	Logarithmic	Level	-4.043
Palma Ratio	Overall	Curve Fitting	Logarithmic	Level	-4.167
	Urban	Curve Fitting	Logarithmic	Level	-2.959
	Rural	Curve Fitting	Logarithmic	Level	-4.769
Quintile Ratio	Overall	Curve Fitting	Power	Level	-12.018
	Urban	Curve Fitting	Logarithmic	Level	-11.338
	Rural	Curve Fitting	Logarithmic	First Difference	-5.457

Note: Barring Palma Ratio for Urban areas, all series reject the hypothesis of Unit Root at one percent level of significance. Dickey Fuller Critical Values for 1, 5 and 10 percent are -3.716, -2.986 and -2.624 respectively for 30 observations.

6. Relationship between Poverty Inequality and Growth

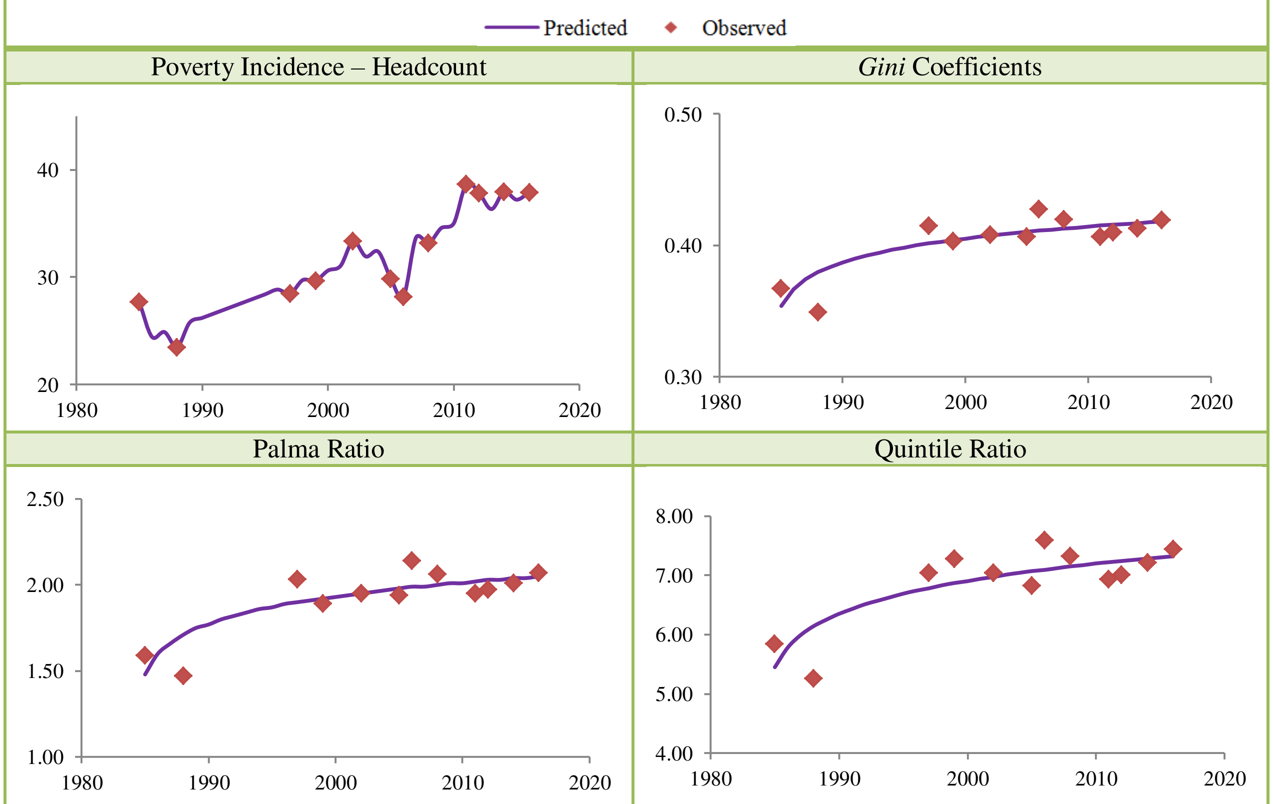
The observed and selected interpolated annual estimates of poverty and inequality are provided in the Appendix-4, whereas Exhibit 6.1 in this section facilitates the assessment of accuracy of interpolation exercise by organizing observed and predicted values of poverty and inequality. The selected poverty series is developed by 'replacing missing value' method of interpolation and as such it just fills the gap between two points with the help of estimated values of regression

[12] Broadly speaking, a stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed. In the time-series literature, such a stochastic process is known as a weakly stationary or stochastic process.

coefficient. In contrast, selected series of inequality measures are based on the ‘curve estimation’ technique of interpolation. The results however show that none of the observed value is far from the estimated line.

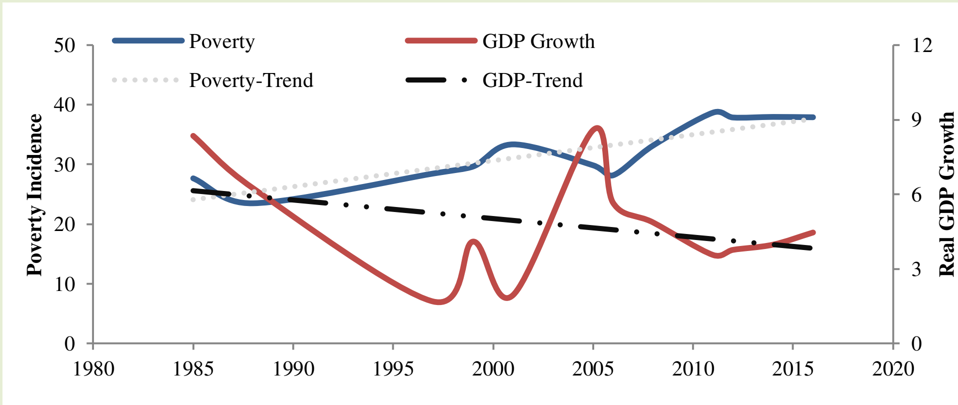
There is consensus among researchers and analysts that economic growth may not always be a sufficient condition for poverty reduction but it certainly is a necessary one. Exhibit 6.2 which graphically depicts the relationship between poverty incidence and economic growth reiterate the phenomena and in general suggests an inverse relationship between these two series. Nonetheless, income distribution and thus measures of inequality change gradually with a considerable lag of time. Therefore, the graphical presentation of income inequality and economic growth (Exhibit 6.3) does not provide strong indication regarding the nature of relationship. Similarly, the association depicted in Exhibit 6.4 between poverty incidence and *Gini* coefficient only gives a rough idea regarding the movement of these two series.

Exhibit – 6.1
Predicted versus Observed Estimates



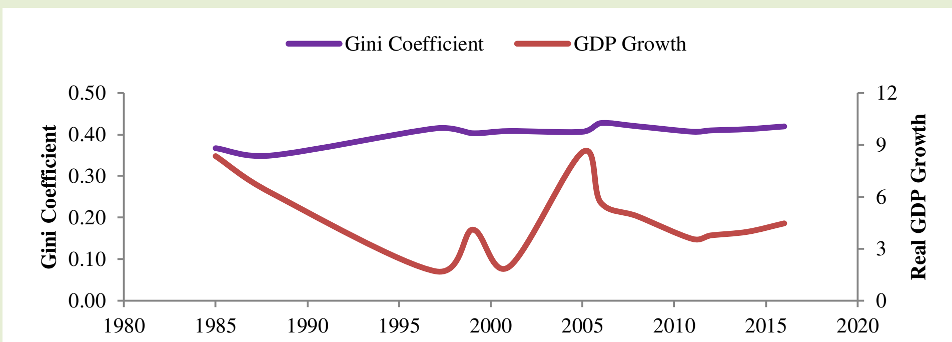
Source: Estimated from HIES (household level) data; various years.

Exhibit – 6.2
Poverty Incidence and Real GDP Growth



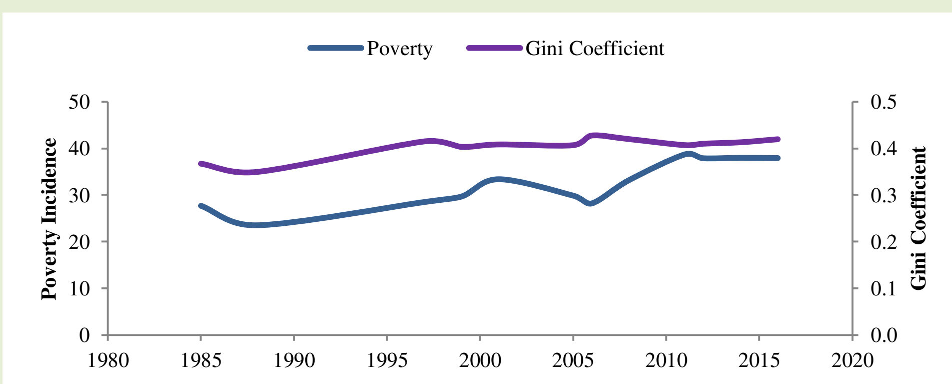
Source: Estimated from HIES (household level) data; various years.

Exhibit – 6.3
Income Inequality and Real GDP Growth
[*Gini* Coefficient – Per Capita Income]



Source: Estimated from HIES (household level) data; various years.

Exhibit – 6.4
Poverty Incidence and *Gini* Coefficient



Source: Estimated from HIES (household level) data; various years.

Consequently, three statistical procedures are used in this research to empirically establish the relationship among poverty, inequality and growth: cointegration, causality and multivariate regression. The interpolated annual estimates of poverty and inequality are used for these econometric techniques.

Cointegration of two (or more) time series suggests that there is a long-run, or equilibrium, relationship between them. Among the various methods for testing cointegration, the most powerful test developed by Johansen is applied here to determine the existence of long-run relationship. Johansen (1991) estimates the cointegration vectors and tests for the order of cointegration vectors and linear relationship in a multivariate model. In a vector auto-regression, cointegration between variables gives an indication that a shock to any one of the equation will trigger response from the rest of the equations in the system. Essentially, the Johansen tests are likelihood-ratio tests which include the "maximum eigenvalue test" and the "trace test". For both test statistics, the initial Johansen test is a test of the null hypothesis of no cointegration against the alternative of cointegration.

Exhibits 6.5 and 6.6 summarize the results of Johansen Cointegration Test for poverty versus GDP and poverty versus inequality respectively. The results however are very much sensitive to the assumptions used for the model specification in terms of intercept, time trend and deterministic trend. Therefore, it is preferred to apply all combinations of assumptions provided in the EView software for testing the evidence of cointegration.

Exhibit – 6.5

Johansen Cointegration Test for
Poverty Incidence and GDP Per Capita

Test Name	Model Specification				
	Assuming No Deterministic Trend in Data		Allow for Linear Deterministic Trend in Data		Allow for Quadratic Deterministic Trend in Data
	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
	Number of Cointegrating Relations				
Trace	2	1	1	1	2
Maximum Eigen Value	2	1	1	1	2
Note: EViews software is used to estimate co-integration Models. Cointegration relations are selected at 0.05 probability value.					

Exhibit – 6.6

Johansen Cointegration Test for
Poverty Incidence and *Gini* Coefficient

Test Name	Model Specification				
	Assuming No Deterministic Trend in Data		Allow for Linear Deterministic Trend in Data		Allow for Quadratic Deterministic Trend in Data
	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
	Number of Cointegrating Relations				
Trace	1	1	2	1	2
Maximum Eigen Value	1	1	2	1	2
Note: EViews software is used to estimate co-integration Models. Cointegration relations are selected at 0.05 probability value.					

The results presented in the both exhibits confirm that there is at least one cointegration vector in the given set of variables and thus reject the null hypothesis of no cointegration at 5 percent level of significance (probability value). Consequently the findings of this research suggest the existence of long-run or equilibrium relationship between interpolated poverty incidence, interpolated *Gini* coefficient and GDP per capita growth.

Standard Granger-type causality test is also applied to the interpolated series to ascertain the causal link between poverty, growth and inequality. According to Granger (1969), the term "Granger Causality" means "precedence". For instance, do movements in per capita income precede movements in poverty, or its opposite, or the movement contemporaneous? This is the approach of Granger causality which is a popular method for studying casual links between random variables. Granger causality assumes linear interactions by virtue of the autoregressive model structure. The null hypothesis for the test is that lagged x-values do not explain the variation in y. In other words, it assumes that x(t) doesn't Granger-cause y(t).

The results of Granger causality exercise are furnished in the Exhibit 6.7, which clearly indicate that there is a unidirectional causality between GDP growth and poverty. In contrast, a two-way directional causality between poverty and inequality is evident. An important finding of this research is that no causal link is found between the inequality (*Gini)* and the GDP growth. The phenomenon crudely reflects the absence of pro-poor policies in the growth process during the period of the study.

Exhibit – 6.7
Pairwise Granger Causality Test
[1985-2016 with Lag One]

Null Hypotheses	F-Statistics	Probability	Decision
GDP GROWTH does not Granger Cause POVERTY	6.27	0.0183	Causality
POVERTY does not Granger Cause GDP GROWTH	0.44	0.5141	No Causality
GINI does not Granger Cause Growth	0.77	0.3867	No Causality
GROWTH does not Granger Cause *GINI*	0.99	0.3276	No Causality
GINI does not Granger Cause POVERTY	11.08	0.0025	Causality
POVERTY does not Granger Cause *GINI*	45.71	2.E-07	Causality

Note: Statistical software EViews is used to estimate co-integration models.

Regression framework is used to estimate elasticity of poverty with respect to GDP growth and the interpolated series of *Gini* coefficients. Log-Linear specification is preferred simply for the ease in the interpretation of regression coefficients.[13] The findings from this estimation exercise are provided in Exhibit - 6.8.

The model is estimated using OLS with White's correction for heteroscedasticity. The summary statistics show a good fit with adjusted R^2 value of 0.94, while the value of Durbin-Watson test confirms the absence of autocorrelation. Moreover, all determinants of poverty are statistically significant with a priori expected signs. A negative relationship between poverty incidence and preceding GDP growth rate is evident in the exhibit with the estimated elasticity of 1.13 percent.

[13] The logarithmic transformation is a monotone transformation which preserves the ordering between x and $f(x)$. In the log-linear model, the literal interpretation of the estimated coefficient β is that a one-unit increase in X will produce an expected increase in log Y of β units. In terms of Y itself, this means that the expected value of Y is multiplied by e^{β}. To be more precise, the relationship between the percent change in y and change in x may be estimated as $[(e^{\beta i} - 1)*100]$.

In contrast, the elasticity of poverty with respect to inequality (*Gini*) is significantly higher (4.677) than the elasticity with respect to growth.[14] The phenomenon endorses the findings of Jamal (2006) in terms of trend of poverty elasticity. He also estimated significantly higher poverty with respect to income inequality.

Exhibit – 6.8
Responsiveness of Poverty to Growth and Inequality (Log-Linear Model)
[Dependent Variable: Log(Headcount)]

Explanatory Variables	Coefficient	Std. Error	t-Statistic	Probability	Elasticity of Poverty [Percentage]
Intercept (Constant)	1.542127	0.45796	3.36736	0.0023	--
GDP Growth (with Lag 1)	-0.011285	0.00529	-2.13028	0.0424	1.134892
GINI Coefficient	0.045705	0.01139	4.01029	0.0004	4.676557
Inflation (GDP Deflator)	0.000651	0.00001	6.89577	0.0000	0.065121
R-squared	0.942	Mean dependent variable		3.422	
Adjusted R-squared	0.936	S.D. dependent variable		0.143	
S.E. of regression	0.036	Akaike info criterion		-3.686	
Sum squared residuals	0.035	Schwarz criterion		-3.502	
Log likelihood	61.148	F-statistic		148.794	
Durbin-Watson stat	1.985	Probability (F-statistic)		0.000	

Source: Estimated by author using EView software.

7. Concluding Remarks

This research provides continuous time-series of poverty incidence and various measures of income inequality. Methodologically consistent estimates were generated using the consumption and income data of 12 household surveys conducted during the period 1985 to 2106. Continuous series were then developed by applying various interpolation techniques to these observed sporadic estimates. These series may be used as an input in the analysis of relationship between macro variables and poverty and inequality estimates.

The study also investigated the nature of relationship among poverty, inequality and growth in the context of Pakistan. Various statistical procedures are applied to interpolated series. The cointegration exercise confirms the existence of long run equilibrium relationship among GDP growth and the estimated series of poverty incidence and *Gini* coefficient. The test of Granger causality indicates a unidirectional causality between economic growth and poverty incidence, while no causal link is confirmed between growth and inequality. Finally, the poverty elasticity with respect to inequality is statistically significant and also the magnitude is relatively high as compared with poverty elasticity of growth. This result clearly reveals the relative importance of income inequality in poverty reduction.

[14] The estimates of elasticities are not comparable with the findings of Jamal (2006) due to differences in the poverty estimation methodology, specification of multivariate regression model and study period.

References:

Cobham, A. and Sumner, A. (2013a), 'Putting the Gini Back in the Bottle? "The Palma" as a Policy-Relevant Measure of Inequality', Mimeograph, London: King's College London

Cobham, A. and Sumner, A. (2013b), "*Is It All About the Tails? The Palma Measure of Income Inequality",* CGD Working Paper, Washington DC, CGD

Government of Pakistan (GOP, 2001), "Food Consumption Table for Pakistan", Department of Agricultural Chemistry, NWFP Agriculture University, Peshawar

Granger, C W J., "Investigating Causal Relationships by Econometric Models and Cross-spectral Methods", *Econometrica*, 1969

Jamal Haroon (2006), "Does Inequality Matter for Poverty Reduction? Evidence from Pakistan's Poverty Trends", *Pakistan Development Review*, Autumn

Jamal, Haroon (2002), "On the Estimation of an Absolute Poverty Line: An Empirical Appraisal", *The Lahore Journal of Economics,* July-December

Johansen S., (1991), "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control*
1
Kakwani, N. (1980), Income Inequality and Poverty: Methods of Estimation and Policy Applications, Oxford University Press, Oxford

Palma, G. (2011), "Homogeneous Middles vs. Heterogeneous Tails, and the End of the "Inverted-U": The Share of the Rich is What It's All About", *Development and Change*, 42 (1), pp. 87–153

Appendix – 1

HIES Survey Design and Sampling Methods:

The Household Integrated Economic Survey (HIES) covers all urban and rural areas of the four provinces and the capital territory (Islamabad) of Pakistan. It however excludes some parts of northern areas, protected areas of Khyber Pakhtunkhwa (KPK) and military restricted areas. Pakistan Bureau of Statistics (PBS) uses separate sampling frames for urban and rural areas. For urban areas, PBS has developed a sample frame using quick count listing methods for households in major cities and town. Each area is subdivided into enumeration blocks consisting 200 to 250 households. For rural areas the list of village/mouzas/dehs published in population and housing census of 1998 is used as a sampling frame.

In urban areas each large-size city is treated as an independent stratum and further divided into low, middle, and high income sub-strata in the light of information from enumeration blocks. The remaining urban areas in all provinces are grouped together and treated as an independent stratum. In rural areas, the population of each district in Punjab, Sindh and KPK province have been grouped together to make a stratum while for Balochistan province each of defunct administrative division is taken as a stratum.

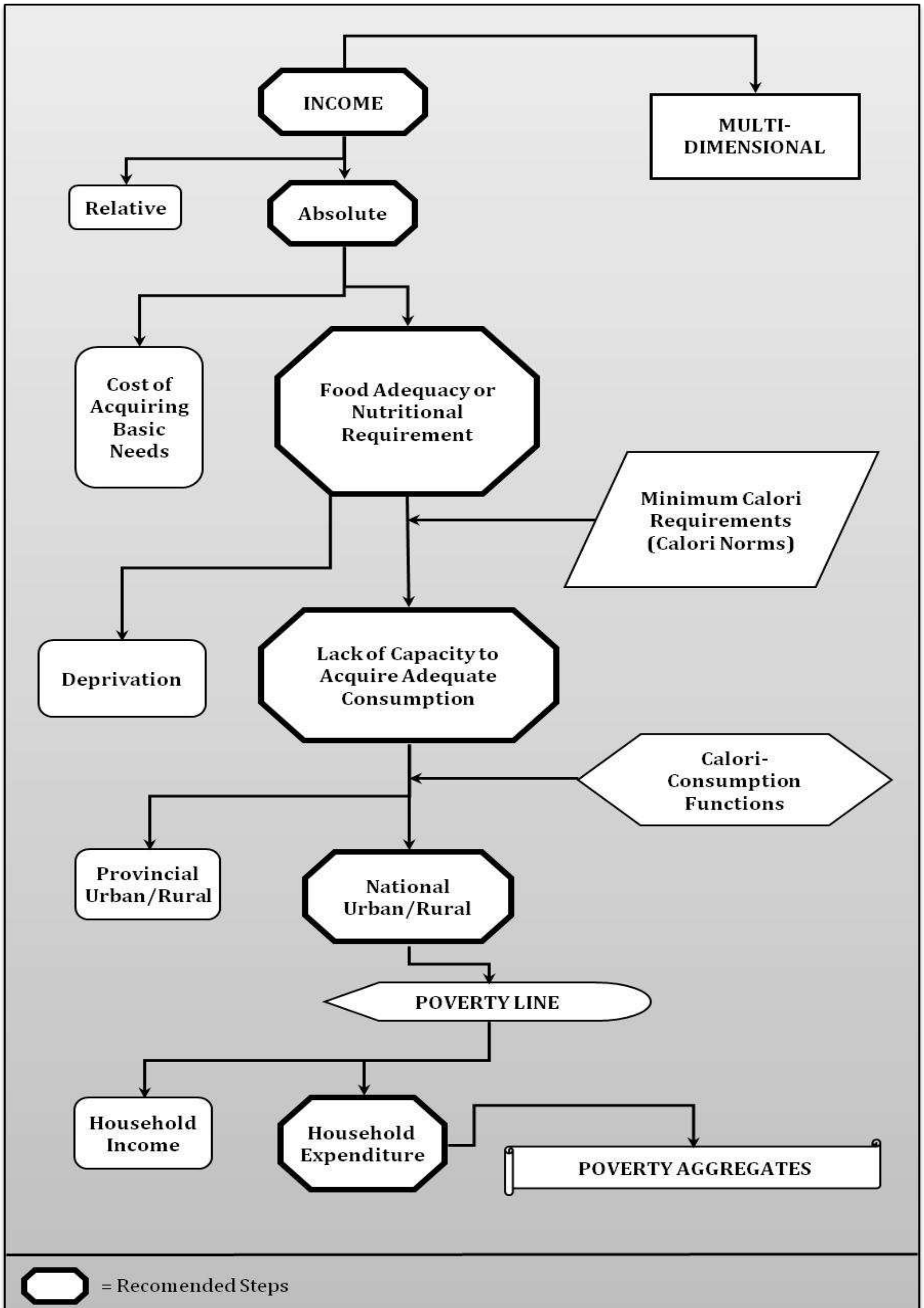
In all surveys, a two-stage stratified random sample design is adopted to select the households. In the first stage, Primary Sampling Units (PSUs) are selected in the urban and rural areas. Enumeration blocks in the urban areas and mouzas/dehs/village in the rural areas are PSUs. The sample PSUs are selected by probability to size (PPS) based on the number of households in the PSU. The households within PSU were taken as secondary sampling units (SSUs) and chosen using systematic sampling scheme with a random start. Sixteen and twelve households are selected from rural and urban areas respectively from each primary sampling unit.

An extensive cleaning process have been applied for each HIES dataset. Household data was scrutinized in terms of food shares, item-wise per capita food expenditures and missing information on family size, food quantities and expenditure. The exhibit below furnishes the number of observations (households) used in the estimation of poverty and income inequality after the cleaning process.

HIES – Year	Number of Observations Used in Poverty and Inequality Estimation
1985	16484
1988	18107
1997	14229
1999	14599
2001	14831
2005	14391
2006	15412
2008	15476
2011	16323
2012	15730
2014	17802
2016	24197
Source: Author’s own estimates.	

<u>**Appendix – 2**</u>

Approaches to Estimate Poverty Line:



Source: Reproduced from Jamal (2002).

Appendix – 3

POVERTY INDICES

The measures which are used to estimate poverty indices are sensitive to distribution among the poor. A class of functional forms, which has been suggested by Foster et al (1984), uses various powers of the proportional gap between the observed and the required expenditure as the weights to indicate the level of intensity of poverty. The higher the power the greater the weight assigned to a given level of poverty. It therefore, combines both the incidence and intensity. The following formula is used for measuring various poverty aggregates.

$$P^{\alpha} = \left({}^{1}/_{N}\right) * \sum \left[\frac{Z - EXP}{Z}\right]^{\alpha}$$

Where;

P^{α}	=	Aggregation measure
N	=	Total number of households
EXP	=	Observed household total expenditure
Z	=	Poverty line
$\sum$	=	Summation for all households below the poverty line

Putting α=0, the formula shows the HCI, i.e., proportion of households whose consumption fall below the poverty line. This simple measure ignores the depth of poverty. This popular measure, assigns equal weights to all the poor regardless of the extent of poverty.

Putting α=1, the Proportionate Gap Index or Poverty Gap Index (PGI) is calculated. It measures the average distance from the poverty line. Although, PGI shows the depth of poverty, it is insensitive to the distribution among the poor.

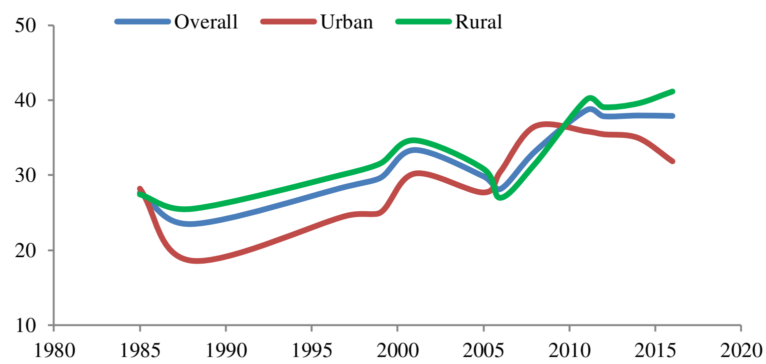
Putting α=2, FGT2 index is calculated. The index takes into account inequality amongst the poor and shows the severity of poverty by assigning greater weights to those households who are far from the poverty line.

Appendix – 4

Observed and Predicted Estimates:

Exhibit – A4.1
Observed Estimates – Poverty Incidence (Headcount)
[Percentage of Poor Population]

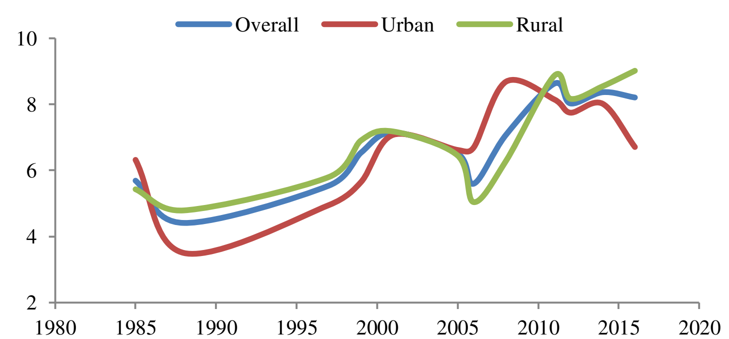
Year	Poverty Incidence		
	Overall	Urban	Rural
1985	27.67	28.21	27.44
1988	23.49	18.61	25.51
1997	28.47	24.58	30.19
1999	29.65	25.02	31.57
2001	33.36	30.22	34.65
2005	29.85	27.70	30.85
2006	28.18	30.52	26.99
2008	33.16	36.51	31.52
2011	38.68	35.85	40.09
2012	37.86	35.46	39.07
2014	37.96	34.97	39.57
2016	37.90	31.85	41.16



Source: Estimated from HIES (Household Level) Data; Various Waves.

Exhibit – A4.2
Observed Estimates – Poverty Gap

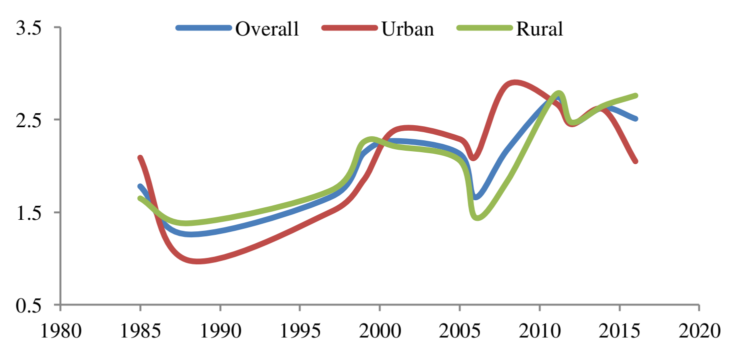
Year	Poverty Gap		
	Overall	Urban	Rural
1985	5.69	6.32	5.43
1988	4.41	3.50	4.79
1997	5.54	4.95	5.80
1999	6.54	5.65	6.91
2001	7.15	7.08	7.18
2005	6.51	6.62	6.45
2006	5.60	6.70	5.04
2008	7.08	8.69	6.29
2011	8.63	8.15	8.87
2012	8.03	7.75	8.17
2014	8.37	8.02	8.55
2016	8.21	6.71	9.02



Source: Estimated from HIES (Household Level) Data; Various Waves.

Exhibit – A4.3
Observed Estimates – Poverty Severity

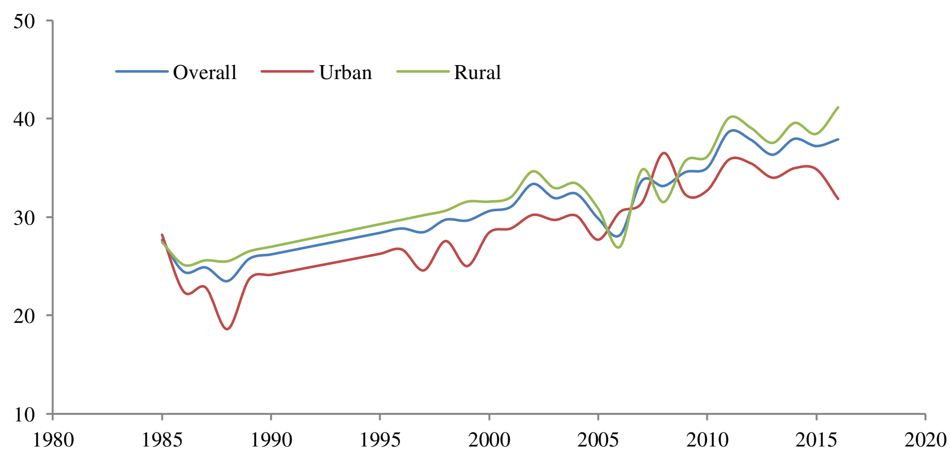
Year	Poverty Severity		
	Overall	Urban	Rural
1985	1.78	2.09	1.65
1988	1.26	0.98	1.38
1997	1.67	1.51	1.74
1999	2.14	1.85	2.26
2001	2.27	2.39	2.21
2005	2.13	2.29	2.06
2006	1.66	2.1	1.44
2008	2.18	2.88	1.84
2011	2.74	2.68	2.77
2012	2.46	2.45	2.47
2014	2.63	2.61	2.65
2016	2.51	2.05	2.76



Source: Estimated from HIES (Household Level) Data; Various Waves.

Exhibit – A4.4
Predicted Series – Poverty Incidences

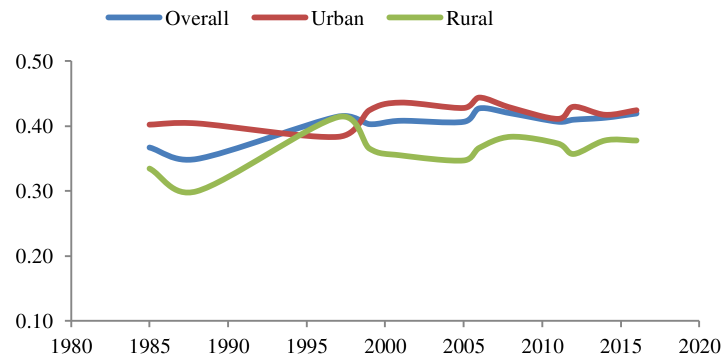
Year	Overall	Urban	Rural
1985	27.67	28.21	27.44
1986	24.44	22.41	25.14
1987	24.88	22.84	25.60
1988	23.49	18.61	25.51
1989	25.76	23.70	26.52
1990	26.20	24.13	26.98
1991	26.64	24.56	27.44
1992	27.08	24.99	27.90
1993	27.52	25.41	28.36
1994	27.96	25.84	28.82
1995	28.40	26.27	29.28
1996	28.84	26.70	29.73
1997	28.47	24.58	30.19
1998	29.73	27.56	30.65
1999	29.65	25.02	31.57
2000	30.61	28.42	31.57
2001	31.05	28.85	32.03
2002	33.36	30.22	34.65
2003	31.93	29.71	32.95
2004	32.37	30.14	33.41
2005	29.85	27.70	30.85
2006	28.18	30.52	26.99
2007	33.69	31.43	34.79
2008	33.16	36.51	31.52
2009	34.57	32.28	35.71
2010	35.01	32.71	36.16
2011	38.68	35.85	40.09
2012	37.86	35.46	39.07
2013	36.34	34.00	37.54
2014	37.96	34.97	39.57
2015	37.22	34.86	38.46
2016	37.90	31.85	41.16



Source: Estimated from HIES (Household Level) Data; Various Waves.

Exhibit – A4.5
Per Capita Income Inequality – *Gini* Coefficients [Observed]

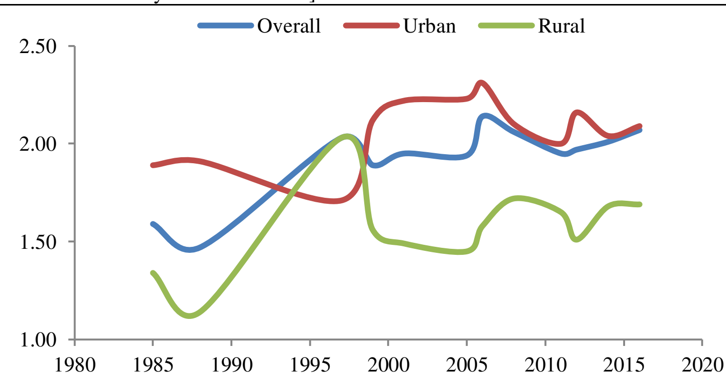
Year	***Gini* Coefficients**		
	Overall	Urban	Rural
1985	0.3670	0.4023	0.3345
1988	0.3492	0.4042	0.2995
1997	0.4147	0.3835	0.4139
1999	0.4030	0.4245	0.3653
2001	0.4082	0.4362	0.3548
2005	0.4066	0.4279	0.3470
2006	0.4273	0.4440	0.3666
2008	0.4197	0.4284	0.3835
2011	0.4067	0.4111	0.3731
2012	0.4098	0.4298	0.3570
2014	0.4129	0.4174	0.3779
2016	0.4194	0.4244	0.3778



Source: Estimated from HIES (Household Level) Data; Various Waves.

Exhibit – A4.6
Per Capita Income Inequality – Palma Ratio [Observed]
[Share – Richest 10% by the Poorest 40%]

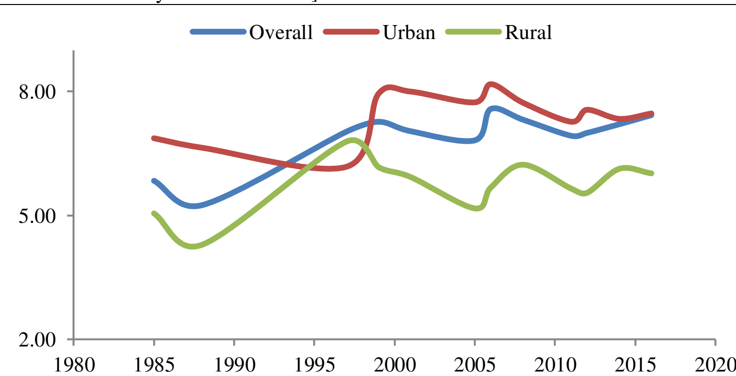
Year	**Palma Ratio**		
	Overall	Urban	Rural
1985	1.59	1.89	1.34
1988	1.47	1.91	1.14
1997	2.03	1.71	2.03
1999	1.89	2.12	1.56
2001	1.95	2.22	1.49
2005	1.94	2.23	1.45
2006	2.14	2.31	1.58
2008	2.06	2.10	1.72
2011	1.95	2.00	1.65
2012	1.97	2.16	1.51
2014	2.01	2.04	1.68
2016	2.07	2.09	1.69



Source: Estimated from HIES (Household Level) Data; Various Waves.

Exhibit – A4.7
Per Capita Income Inequality – Quintile Ratio [Observed]
[Share – Richest 20% by the Poorest 20%]

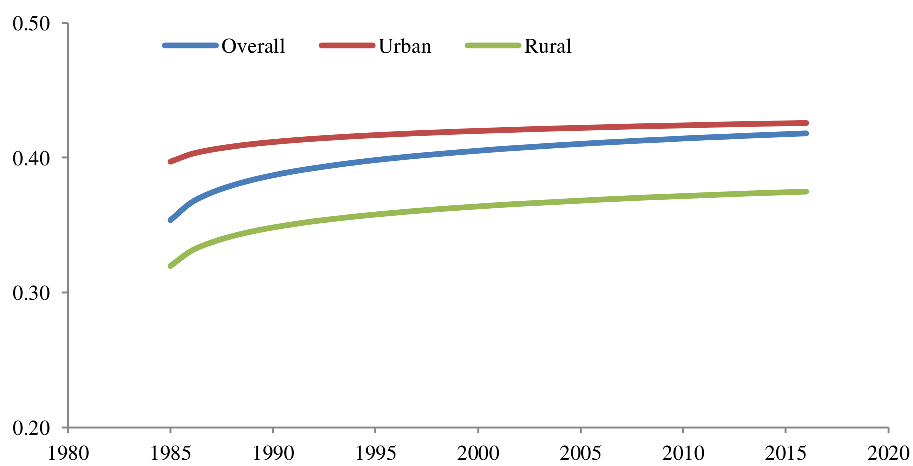
Year	**Quintile Ratio**		
	Overall	Urban	Rural
1985	5.84	6.87	5.05
1988	5.25	6.64	4.29
1997	7.03	6.18	6.79
1999	7.27	7.95	6.17
2001	7.04	8.00	5.93
2005	6.82	7.74	5.17
2006	7.58	8.18	5.68
2008	7.32	7.73	6.23
2011	6.93	7.27	5.65
2012	7.00	7.56	5.55
2014	7.21	7.34	6.13
2016	7.43	7.47	6.02



Source: Estimated from HIES (Household Level) Data; Various Waves.

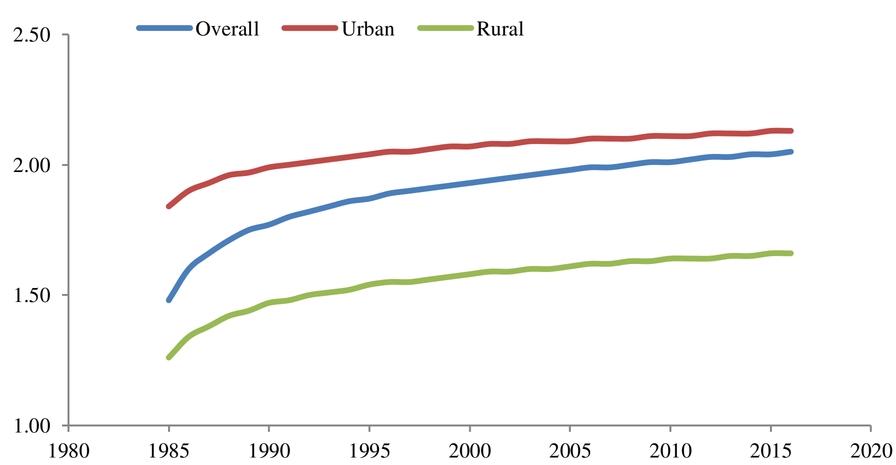
Exhibit – A4.8
Predicted Series – *Gini* Coefficients

Year	Overall	Urban	Rural
1985	0.3536	0.3970	0.3196
1986	0.3665	0.4026	0.3307
1987	0.3740	0.4059	0.3371
1988	0.3793	0.4082	0.3417
1989	0.3835	0.4101	0.3453
1990	0.3869	0.4116	0.3482
1991	0.3898	0.4129	0.3506
1992	0.3922	0.4140	0.3528
1993	0.3944	0.4150	0.3546
1994	0.3964	0.4159	0.3563
1995	0.3982	0.4167	0.3578
1996	0.3998	0.4174	0.3592
1997	0.4013	0.4181	0.3605
1998	0.4026	0.4187	0.3617
1999	0.4039	0.4193	0.3628
2000	0.4051	0.4198	0.3638
2001	0.4063	0.4203	0.3648
2002	0.4073	0.4208	0.3657
2003	0.4083	0.4213	0.3665
2004	0.4093	0.4217	0.3673
2005	0.4102	0.4221	0.3681
2006	0.4111	0.4225	0.3689
2007	0.4119	0.4229	0.3696
2008	0.4127	0.4233	0.3703
2009	0.4134	0.4236	0.3709
2010	0.4142	0.4239	0.3715
2011	0.4149	0.4243	0.3721
2012	0.4155	0.4246	0.3727
2013	0.4162	0.4249	0.3733
2014	0.4168	0.4252	0.3738
2015	0.4174	0.4254	0.3743
2016	0.4180	0.4257	0.3748



Source: Estimated from HIES (Household Level) Data; Various Waves.

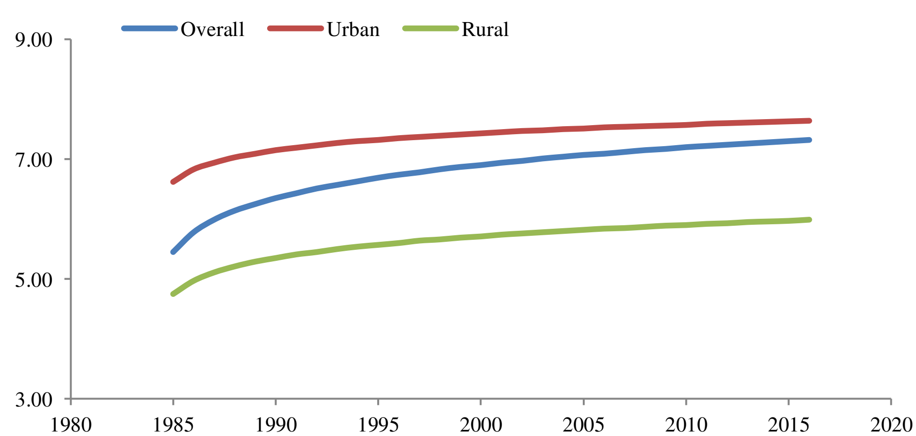
Exhibit – A4.9 Predicted Series – Palma Ratio			
Year	Overall	Urban	Rural
1985	1.48	1.84	1.26
1986	1.60	1.90	1.34
1987	1.66	1.93	1.38
1988	1.71	1.96	1.42
1989	1.75	1.97	1.44
1990	1.77	1.99	1.47
1991	1.80	2.00	1.48
1992	1.82	2.01	1.50
1993	1.84	2.02	1.51
1994	1.86	2.03	1.52
1995	1.87	2.04	1.54
1996	1.89	2.05	1.55
1997	1.90	2.05	1.55
1998	1.91	2.06	1.56
1999	1.92	2.07	1.57
2000	1.93	2.07	1.58
2001	1.94	2.08	1.59
2002	1.95	2.08	1.59
2003	1.96	2.09	1.60
2004	1.97	2.09	1.60
2005	1.98	2.09	1.61
2006	1.99	2.10	1.62
2007	1.99	2.10	1.62
2008	2.00	2.10	1.63
2009	2.01	2.11	1.63
2010	2.01	2.11	1.64
2011	2.02	2.11	1.64
2012	2.03	2.12	1.64
2013	2.03	2.12	1.65
2014	2.04	2.12	1.65
2015	2.04	2.13	1.66
2016	2.05	2.13	1.66



Source: Estimated from HIES (Household Level) Data; Various Waves.

Exhibit – A4.10
Predicted Series – Quintile Ratio

Year	Overall	Urban	Rural
1985	5.45	6.62	4.75
1986	5.78	6.83	4.97
1987	5.99	6.94	5.11
1988	6.14	7.03	5.21
1989	6.25	7.09	5.29
1990	6.35	7.15	5.35
1991	6.43	7.19	5.41
1992	6.51	7.23	5.45
1993	6.57	7.27	5.50
1994	6.63	7.30	5.54
1995	6.69	7.32	5.57
1996	6.74	7.35	5.60
1997	6.78	7.37	5.64
1998	6.83	7.39	5.66
1999	6.87	7.41	5.69
2000	6.90	7.43	5.71
2001	6.94	7.45	5.74
2002	6.97	7.47	5.76
2003	7.01	7.48	5.78
2004	7.04	7.50	5.80
2005	7.07	7.51	5.82
2006	7.09	7.53	5.84
2007	7.12	7.54	5.85
2008	7.15	7.55	5.87
2009	7.17	7.56	5.89
2010	7.20	7.57	5.90
2011	7.22	7.59	5.92
2012	7.24	7.60	5.93
2013	7.26	7.61	5.95
2014	7.28	7.62	5.96
2015	7.30	7.63	5.97
2016	7.32	7.64	5.99



Source: Estimated from HIES (Household Level) Data; Various Waves.