RESEARCH ARTICLE

Education Attainment Forecasting and Economic Inequality in the United States

Sudeshna Ghosh

Economics Department, Scottish Church College, Kolkata, India

Abstract: This paper explores a major issue in the context of economics of inequality, namely how far the stark inequalities are retarding educational attainment in the United States in the recent period. An application of the time series ARIMA model is made to forecast both high school and college education in the United States for the period 2018-2022. The study obtains fluctuating trends (forecast) in certain categories of school educational attainment. A Multi Variate Error Correction Model is utilized to assess how far socio economic factors are responsible in the process of educational attainment. The study obtains a bidirectional long run causal association across educational completion, income growth, inflation, and income inequality. Income inequality and the distribution of income in the United States has a positive bearing on the education sector. Further distributive polices and the behaviour of government spending have a significant effect on the quality of learning. Last, attainment of education and its ramification has an important bearing on the level of income in the United States in the long run.

Keywords: United States; inequality; education; ARIMA model; time series; vector error correction

1. Introduction

The World Economic Forum in Davos (Switzerland), 2014 has underlined, the challenges of expanding income inequalities in the major advanced economies of the world, in the recent decade. United States is a major advanced economy among others which is facing rising economic inequality during the current times. There is a growing consensus in the literature that the distribution of income, assets, wealth, and income earned through wages are facing significant skewness. According to Bernstein (2013)[1], between 1979 and 2007 per capita consumption rose slightly higher than 2.2 percent annually but income enjoyed by the top one percent rose up to 13.5 percentage points. The literature has debated upon the merits and demerits of rising inequality. As far as the benefits from inequality is concerned it provides the motivation to work harder to achieve higher rates of income. However, inequality causes macro-economic instability, Stiglitz (2012)[2], it further generates distortionary growth, Alesina & Roderik (1994)[3]. The investment decision making of the households in human capital formation gets reduced. The recent upsurge in inequality in income distribution can be attributed to globalization and expansion of product and factor markets; greater participation of workers with low economies of scale; skill augmenting changes in technology; policy thrust of cut in tax rates for the top income quintile. Alvaredo *et al* (2013)[4], Hoeller *et al*. (2012)[5] emphasis that these factors have a detrimental effect on income inequality but Chen and Ravallion (2010)[6] emphasis that these factors are contributory to reduction in relative poverty at the world level.

Education was traditionally a level player in the United States, it enabled the less better off children to move up the socio economic ladder. However, recently the society is experiencing widening gap across the rich and the poor children as far as educational achievements is concerned, Oded (2011)[7]. In this study we explore how educational achievements

Copyright © 2018 Sudeshna Ghosh.
doi: 10.18063/csa.v1i2.787
This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

is influenced by income inequality in the United States, by applying time series data for the period 1972 to 2016. The purpose of the current exercise is twofold, first by applying a univariate time series model we forecast the situation of education in the United States both at high school levels and at college levels. Next we attempt to seek explanation for the differences in the educational achievement through a multivariate VAR model. Last, we compare the forecasting efficacy of the univariate versus the multivariate model. The specific objectives of this study are:

1) First to forecast the status of educational attainment both at high school level completion and college level, across various age groups, sex, and race over the period 2018-2022.

2) To model the heterogeneity in educational achievements in the United States, how far income and income inequality is explaining the human capital formation- A Vector Autoregressive Framework is developed in the study.

3) Finally, the study explores whether univariate forecasting or multivariate forecasting provides better results in the context of educational achievements.

The paper is organized as follows the Section II discusses about the data sources and the methodology applied, Section III delves on the broad results, the discussion is in the Section V. Finally the paper is concluded in the Section VI.

2. Materials and Methods

2.1 Materials

The major variables utilized in the analysis are percentage of population (male and female) who have completed different levels of school and college education in the United States at different age groups and across varying ethnic groups; logarithmic per capita income and its square; logarithmic public expenditure in the United States; logarithmic trade openness denoted by [exports+imports/GDP]; inflation denoted by average annual CPIlogarithmic of urbanization denoted by percentage of population living in urban areas and inequality in income measured by the income GINI. The Table 1 (a) provides the details of the variables utilized and the source of data. The time series of observations run from 1972 to 2016. Further Table 1 (b) provides the summary of the descriptive statistics. The descriptive statistics show that the standard deviations differ among variables and the statistics range from 0.01 to 4.63. The skewness for GINI INCOME, LEXP, and LTR are negatively ended while for other variables it is positively ended, Table 1 (b). The time series plots for each variable are illustrated in the Figure (1).

Variables	Description	Source
ED	ED denotes educational completion at different levels of school and college education.	Period: 1972-2016. Source: Current Population Survey, Annual Social and Economic Supplement to the Current Population Survey, U.S. Census Bureau, Education, and Social Stratification Branch[8].
LGDP	LGDP per capita is gross domestic product divided (in Logarithmic terms) by midyear population. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. Data are in constant local currency.	Period: 1972-2016. Source: World Bank's World Development Indicators (WDI)[9].
$LGDP^2$	$LGDP^2$ is square of gross domestic product in logarithmic terms.	Period: 1972-2016. Source: World Bank's World Development Indicators (WDI).
GINI INCOME	GINI INCOME denotes the Gini coefficient of	Period: 1972-2016. Source: U.S.

	disposable income inequality, it is the popular measure used to measure income inequality.	Census Bureau, Current Population Survey, 1968 to 2017 with Annual Social and Economic Supplements.
LU	LU is logarithmic of urban population, refers to people living in urban areas as defined by national statistical offices.	Period: 1972-2016. Source: World Bank's World Development Indicators (WDI).
CPI	CPI denotes inflation, it is measured by the consumer price index, reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used.	Period: 1972-2016. Source: World Bank's World Development Indicators (WDI). World Bank's World Development Indicators (WDI).
LEXP	Public spending in the United States in logarithmic terms, includes social security funds.	Period: 1972-2016. Source: Public spending is available from IMF, Government Finance Statistics Yearbook[10].
LTR	LTR stands for logarithmic of trade openness denoted by Exports+Imports/GDP. Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. Imports of goods and services represent the value of all goods and other market services received from the rest of the world.	Period: 1972-2016. Source: World Bank's World Development Indicators (WDI). World Bank's World Development Indicators (WDI).

Table 1 (a): Data Description

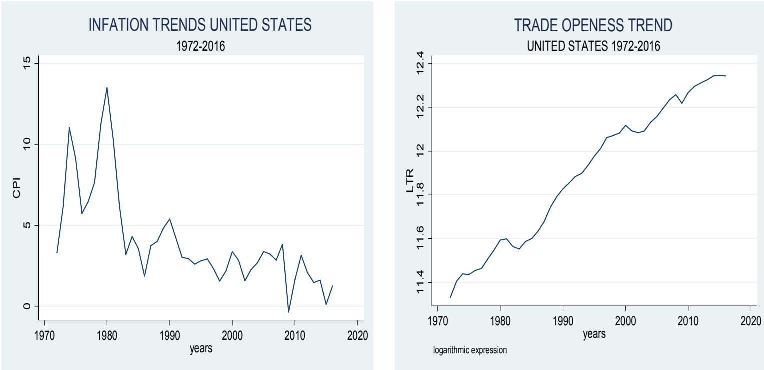

Figure 1. Time series plots for all variables under review

Measures	ED	LGDP	$LGDP^2$	GINI INCOME	LU	CPI	LEXP	LTR
Mean	26.54	1.41	2.01	0.41	1.88	4.07	12.13	11.89
Maximum	36.1	1.55	2.42	0.46	1.91	13.50	12.62	12.34
Minimum	19	1.27	1.63	0.35	1.86	0.35	11.35	11.33
Standard Deviation	4.63	0.07	0.21	0.03	0.01	3.00	0.36	0.31
Skewness	0.51	0.29	0.36	-0.43	0.12	1.44	-0.46	-0.15
Kurtosis	2.12	1.93	1.97	1.74	1.41	4.68	2.26	1.62

Source: World Development indicators, World Bank; IMF, Government Finance Statistics Yearbook and Current Population Survey, U.S. Census Bureau. Compilation Self

Table 1 (b): Summary of Data

According to Kuznets inverted–U hypothesis a positive correlation exists between income inequality and per capita income at low levels of income, but when income reaches higher levels this association becomes negative. However, the United States during the recent period is experiencing high income along with high income inequality. This is generating increased social instability and there is an unfavourable climate to household investment in human capital formation. There is an urgent role for the state in the area of social spending and the thrust of emphasis should be for redistributive policies. Again, globalization and trade openness have enhanced the premium on skilled wage returns, yet the educational gap remains. Such, puzzling situation makes this study investigate why people do not participate in the education process? At the outset a forecasting exercise is undertaken through univariate ARIMA forecasting model to see the future trends in completion at different levels of education. Next a multivariate framework is constructed to investigate the causal factors associated with low educational levels.

The Autoregressive Integrated Moving Average (ARIMA) model is a notable stochastic time series model. The central assumption applied to build this model is the time series observations is linear and follows a normal distribution. Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) are various sub classification of the ARIMA model. The model is built on the principle of parsimony. Box Jenkins[11] methodology is used for ideal model building process in ARIMA, which has made ARIMA popular. An exercise of such kind will enhance efficiency in policy decision making, relating to allocation of future resources to improving educational outcomes. Further, the study utilizes multivariate dynamic model building exercise in a Vector Error Correction framework to study the impact of macroeconomic and social factors in determining educational outcomes in the United States. Vector autoregression (VAR) model was formulated by Sims (1980)[12], this technique is basically used to describe the joint dynamic behavior of a group of time series variables, and further this method does not require any strong restrictions to identify the underlying structural parameters of the equations.

2.2 Methods

2.2.1 ARIMA model

The autoregressive integrated moving average (ARIMA) time series model is utilized here to forecast educational attainment at different levels of school and college education in the United States. The advantage of adopting the ARIMA model is, it takes care of serial correlations, considers fluctuations in trends, predictions in errors, non-stationarity of the data and helps to increase the accurateness in forecasting. The ARIMA model was established by Box and Jenkins (1970) and has been widely applied in forecasting different socio-economic variables. The central supposition made to use the ARIMA model is that the given time series is linear and possesses a certain statistical distribution, namely the normal distribution.

ARIMA (p,d,q) is a linear model originating from the autoregressive model AR (p) , the moving average model MA (q) and thus the combination of the two AR (p) and MA (q) is the ARIMA (p,d,q). The model is developed as follows-

=

$E(\varepsilon_i)$=0,$Var(\varepsilon_i)$=σ_ε^2, $E(\varepsilon_t\,\varepsilon_s)$ =0,$s\ \neq t$ (1)

$Ex_s\,\varepsilon_t$=0, $\forall_s\ < t$

Where p, q are orders of the AR model and MA model respectively, d is the number of series difference. Here p, d, q are all integers. ε_i indicates the estimated residual at each time period. For optimal conditions the model should be independent and distributed as normal random variables with mean=0. σ_ε^2 is the variance of the residuals.

$\Phi(B)$=$1-\emptyset_l B-\cdots-\emptyset_p\,B^p$

⊖(B)=1-θ_(l) B - ⋯ - θ_(q) B^q

Here $\Phi(B)\ and$ ⊖(B) are polynomials in B of degrees p and q. B is backward shift operator.

So in the ARIMA (p,d,q) p,d and q indicates the orders of auto regression, differencing and moving average respectively. The parameter ‘p’ gives evidence on the nature of structural dependencies between neighbouring observations, so it is an indicator of autocorrelation; ‘d’ shows the number of times the series of observations has to be differenced in order to make the series stationary and finally ‘q’ indicated the number of moving average terms. The ARIMA model of forecasting is often chosen over other methods because it provides projections over the future by smoothing of data based on extrapolation. In this study forecasting of people completing school education at various grades across different age groups and forecasting of population completing college education is obtained across different population groups.

Econometric Model Specification : Multi Variate Frame work.

Here we specify how educational attainment is affected by income dispersion in the United States and try to explore the causal relation among the explanatory variables. To test the relationship between educational attainment and income growth and its dispersion we use the model specification described in the equation 1(a).

$ED_t=\beta_0\ +\ \beta_1 LGDP_t+\beta_2 LGDP_t^2+\ \beta_3 GINI\ INCOME_t+\beta_4 X_t+\epsilon_t$ 1(a)

ED is the measure of educational attainment, it is the percentage of population above 25 years of age who have competed four years or more college education; LGDP is logarithmic of income (per capita) and $LGDP^2$ is the square of income (per capita). The effect of economic growth on educational attainment is captured through the income variable. The squared term is utilized to verify the occurrence of the Kuznets Hypothesis which states that when income levels are low then the distribution implications are unequal. However as income rises the distribution tends to be equal. So β_1’s expected sign is positive and β_2 ‘s expected sign is negative. The expected sign of β_3 is negative. When there is rising income inequality educational attainment declines. X denotes the other control variables that affect educational attainment independent of income. ϵ_t is the residual term. The other explanatory variables used in the analysis are urbanization (logarithmic expression) (LU), inflation denoted by the consumer price index (CPI), public expenditure (logarithmic expression) (LEXP) and trade openness (logarithmic expression) (LTR). Levels of urbanization has a positive impact on educational development because the urban areas open opportunities of skilled employment. The impact of inflation on the household’s decision to invest in human capital formation has been consistently utilized in the literature. Rising inflation will tend to lower educational attainments. Public spending programmes with targeted intervention on the educational sector will progressively raise the educational attainments. Trade openness raises the skewness of income of high skilled workers in skill abundant countries (Heckscher-Ohlin Theorem). This will generate the demand for education in these countries.

2.2.2 Cointegration Analysis and the Vector Error Correction Model

Unit Root Test

To obtain the long run causal relationship involving the time series variables it is necessary to specify the stationarity of the series of observations. Stationarity of the series of the observations can be obtained by applying the unit root test. This paper utilizes the Augmented Dickey Fuller (ADF) (1979) test and the Phillips Perron (PP) test (1988) unit root test to examine the stationarity and order of integrability of the series of the observations. The ADF test require the need of running the regression of a first differenced series (of the concerned variable) on its first lag, the lagged difference terms and specified components like the intercept and the time trend. Suppose we run the stationarity test of the time series y_t the ADF test requires the estimation of the equation (2).

$$\Delta y_t = a_0 + a_1 t + a_2 y_{t-1} + \sum_{i=2}^{k} \beta_i \, \Delta y_{t-1+1} + \varepsilon_t \ldots.. \quad (2)$$

Here ε_t denotes the usual uncorrelated stationary error terms possessing a zero mean and constant variance. k denotes the optimum lag length, that should be determined in such a way that will make autocorrelation free. The unit root test is carried on the coefficient of y_{t-1}, if the coefficient is not significantly different from zero then the null hypothesis is accepted. This implies that the equation has a unit root. The rejection of the null hypothesis implies stationarity of the series. The Dickey Fuller test specifies the test statistic and the corresponding critical values.

The Phillips Perron (PP) (1988) test is built on the null hypothesis of ADF test. If $H_0: \alpha=0$ is the null hypothesis then from the equation (3) we obtain the PP test statistic.

$$\Delta y_t = \alpha y_{t-1} + u_t \ldots. \quad (3)$$

Here y_t is the time series of observations and u_t denotes the sequence of innovations. Contrary to the ADF test the PP test alters the test statistic of the α parameter thus the serial correlation does not affect the asymptotic distribution of the concerned test statistic. If all the set of the observations are of the order I (1) then the cointegration test can be applied. This paper applies the Johansen and Juselius (1990) method of cointegration.

Johansen Cointegration Method

The long run equilibrium relation is examined through the Johansen technique of cointegration. The prerequisite of the Johansen method of cointegration is that the sample data fits a finite order of a vector autoregression (VAR) model. Cointegration is use to explore whether the variables share a common stochastic trend and their first difference has to be stationary. This process helps to determine the long run relationships among the variables. In the Johansen-Juselius cointegration method (1990) the cointegration rank of the time series variables (indicated by r) is tested by two test (eigen value and trace statistics) statistics. Let the number of cointegrating vectors be denoted by r_0, the maximum eigen value test is calculated using the null hypothesis $r_0=r$ and the alternative hypothesis is $r_0 > r$. The trace statistics is obtained under the null hypothesis $r_0 \leq r$, the alternative hypothesis is $r0 > r$.

Johansen's technique takes into consideration the vector autoregresssion (VAR) of the order p as elaborated in the equation (4),

$$X_t = \prod_1 X_{t-1} + \prod_2 X_{t-2} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots + \prod_p X_{t-p} + u_t \ \ldots\ldots.. \quad (4)$$

Where X_t is is a nx1 vector of variables that are integrated of order one, I(1) u_t is the error term, $\prod$ to $\prod_p$ is the mXm coefficient matrices. By subtracting X_{t-1} from both sides of equation (4) , the reparametrized version is as follows, in equation (5),

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \ldots\ldots\ldots + \Gamma_{p-1} \Delta X_{t-p+1} - \prod X_{t-p} + u_t \ \ldots.. \quad (5)$$

Where $\Gamma_1 = \prod_1 - I$. $\Gamma_2 = \prod_2 - \Gamma_1 \Gamma_3 = \Gamma_3 - \Gamma_2$; $\prod = I - \prod_1 - \prod_2 - \ldots\ldots\ldots - \prod_p$. The matrix $\prod$ determines how the system is cointegrated.

Based on the equation (5) the first equation of the system can be written as, equation (6)

$$\Delta X_{1t} = \acute{\gamma}_{11} \Delta X_{t-1} + \acute{\gamma}_{12} \Delta X_{t-2} + \ldots\ldots\ldots + \acute{\gamma}_{1p-1} \Delta X_{t-p+1} - \acute{\prod}_1 X_{t-p} + u_{1t} \ldots\ldots \quad (6)$$

where $\acute{\gamma}_{ij}$ is the first row of Γ_j, j=1,2,…..p-1 and $\acute{\prod}_1$ is the first row of $\prod$

The matrix $\prod$ is of order mXm, if this has rank m ,then m is the number of linearly independent rows or columns, this forms the basis of m-dimensional vector space. Any linear combination of the row is stationary.

$\prod$ can be written as $\prod=\beta_{-}\acute{\alpha}_{-}$ for suitable mXr matrices, here

$$\acute{\alpha}_{-}=\begin{bmatrix} \acute{\alpha}_1 \\ \acute{\alpha}_2 \\ . \\ . \\ . \\ . \\ \acute{\alpha}_r \end{bmatrix} \quad (7)$$

$$\beta_{-}=[\beta_1 \quad \beta_2 ....\beta_r]....... \quad (8)$$

Then $\prod X_{t-p} = \beta_{-}\acute{\alpha}_{-}X_{t-p}$ and further all the linear combinations of $\acute{\alpha}_{-}X_{t-p}$ are stationary.Johansen method estimates the VAR for various values of r number of cointegrating vectors, based on the maximum likelihood procedure.The estimate can be written as

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \ldots\ldots\ldots\ldots + \Gamma_{p-1} \Delta X_{t-p+1} - \prod X_{t-p} - \beta_{-}\acute{\alpha}_{-}X_{t-p} + u_t \ \ldots\ldots \quad (9)$$

The number of cointegrating vectors are detected through two likelihood test- the trace test and the maximum eigen value.

Trace test

This test tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of n cointegrating vectors, the test statistics is thus,

$$J_{trace} = -T\sum_{i=r+1}^{n} In(1-\grave{\lambda}_i) \ \ldots\ldots\ldots \quad (10)$$

Maximum eigenvalue

The maximum eigen value tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of(r+1)cointegrating vectors. The test statistic is as follows,

$$J_{max} = -T\ (1-\grave{\lambda}_{r+1})\ldots\ldots \quad (11)$$

here T is the sample size and $\grave{\lambda}_i$ stands for the i th largest canonical equation.

If a cointegrating relation exists then the dynamics of the model can be estimated through the Vector Error Correction (VEC) model. The VEC model has the cointegrating relation inbuilt in the specification that makes the long run relation of the endogenous variables to converge into the cointegrating relation. It further allows for the short run dynamic adjustment. The error correction term shows the deviation from the long run equilibrium. The size of the error correction term shows the speed of adjustment of any disequilibrium towards the long run equilibrium state. The deviation from the long run equilibrium is corrected through a sequence of short term adjustments.

The Vector Error Correction Model

Estimable VEC model involving educational completion, logarithmic per capita income, square of logarithmic per capita oncome, Gini income inequality, logarithmic of urbanization, consumer price index, logarithmic of public spending and logarithmic of trade openness consists of the set eight equations running from equation (12) to equation (19).

$$\Delta ED_t = \vartheta_1 + \sum_{i=1}^{k}\beta_i\ \Delta ED_{t-i} + \sum_{i=0}^{k}\varphi_i\ \Delta LGDP_{t-i} + \sum_{i=0}^{k}\emptyset_i\ \Delta LGDP_{t-i}^2 + \sum_{i=0}^{k}\theta_i\ \ \Delta GINI\ INCOME_{t-i} + \sum_{i=0}^{k}\rho_i\ \Delta LU_{t-i} + \sum_{i=0}^{k}\gamma_i\ \ \Delta CPI_{t-i} + \sum_{i=0}^{k}\alpha_i\ \ \Delta LEXP_{t-i} + \sum_{i=0}^{k}\mu_i\ \ \Delta LTR_{t-i} + \delta_1\ EC_{t-1} + \varepsilon_{1t} \quad (12)$$

$$\Delta LGDP_t = \vartheta_2 + \sum_{i=0}^{k}\beta_i\ \Delta ED_{t-i} + \sum_{i=1}^{k}\varphi_i\ \Delta LGDP_{t-i} + \sum_{i=0}^{k}\emptyset_i\ \Delta LGDP_{t-i}^2 + \sum_{i=0}^{k}\theta_i\ \ \Delta GINI\ INCOME_{t-i} + \sum_{i=0}^{k}\rho_i\ \Delta LU_{t-i} + \sum_{i=0}^{k}\gamma_i\ \ \Delta CPI_{t-i} + \sum_{i=0}^{k}\alpha_i\ \ \Delta LEXP_{t-i} + \sum_{i=0}^{k}\mu_i\ \ \Delta LTR_{t-i} + \delta_2\ EC_{t-1} + \varepsilon_{2t} \quad (13)$$

$$\Delta LGDP_t^2 = \vartheta_3 + \sum_{i=0}^{k}\beta_i\ \Delta ED_{t-i} + \sum_{i=0}^{k}\varphi_i\ \Delta LGDP_{t-i} + \sum_{i=1}^{k}\emptyset_i\ \Delta LGDP_{t-i}^2 + \sum_{i=0}^{k}\theta_i\ \ \Delta GINI\ INCOME_{t-i} + \sum_{i=0}^{k}\rho_i\ \Delta LU_{t-i} + \sum_{i=0}^{k}\gamma_i\ \ \Delta CPI_{t-i} + \sum_{i=0}^{k}\alpha_i\ \ \Delta LEXP_{t-i} + \sum_{i=0}^{k}\mu_i\ \ \Delta LTR_{t-i} + \delta_3\ EC_{t-1} + \varepsilon_{3t} \quad (14)$$

$\Delta GINI_t = \vartheta_4 + \sum_{i=0}^{k} \beta_i \ \Delta ED_{t-i} + \sum_{i=0}^{k} \varphi_i \, \Delta LGDP_{t-i} + \sum_{i=0}^{k} \emptyset_i \, \Delta LGDP_{t-i}^2 + \sum_{i=1}^{k} \theta_i \quad \Delta GINI\ INCOME_{t-i} + \sum_{i=0}^{k} \rho_i \, \Delta LU_{t-i} + \sum_{i=0}^{k} \gamma_i \ \ \Delta CPI_{t-i} + \sum_{i=0}^{k} \alpha_i \ \ \Delta LEXP_{t-i} + \sum_{i=0}^{k} \mu_i \ \ \Delta LTR_{t-i} + \delta_4 \, EC_{t-1} + \varepsilon_{4t}$ (15)

$\Delta LU_t = \vartheta_5 + \sum_{i=0}^{k} \beta_i \ \Delta ED_{t-i} + \sum_{i=0}^{k} \varphi_i \, \Delta LGDP_{t-i} + \sum_{i=0}^{k} \emptyset_i \, \Delta LGDP_{t-i}^2 + \sum_{i=0}^{k} \theta_i \quad \Delta GINI\ INCOME_{t-i} + \sum_{i=1}^{k} \rho_i \, \Delta LU_{t-i} + \sum_{i=0}^{k} \gamma_i \ \ \Delta CPI_{t-i} + \sum_{i=0}^{k} \alpha_i \ \ \Delta LEXP_{t-i} + \sum_{i=0}^{k} \mu_i \ \ \Delta LTR_{t-i} + \delta_5 EC_{t-1} + \varepsilon_{5t}$ (16)

$\Delta CPI_t \vartheta_6 + \sum_{i=0}^{k} \beta_i \ \Delta ED_{t-i} + \sum_{i=0}^{k} \varphi_i \, \Delta LGDP_{t-i} + \sum_{i=0}^{k} \emptyset_i \, \Delta LGDP_{t-i}^2 + \sum_{i=0}^{k} \theta_i \quad \Delta GINI\ INCOME_{t-i} + \sum_{i=0}^{k} \rho_i \, \Delta LU_{t-i} + \sum_{i=1}^{k} \gamma_i \ \ \Delta CPI_{t-i} + \sum_{i=0}^{k} \alpha_i \ \ \Delta LEXP_{t-i} + \sum_{i=0}^{k} \mu_i \ \ \Delta LTR_{t-i} + \delta_6 EC_{t-1} + \varepsilon_{6t}$ (17)

$\Delta LEXP_t = \ \vartheta_7 + \sum_{i=0}^{k} \beta_i \ \Delta ED_{t-i} + \sum_{i=0}^{k} \varphi_i \ \ \Delta LGDP_{t-i} + \sum_{i=0}^{k} \emptyset_i \, \Delta LGDP_{t-i}^2 \ + \ \sum_{i=0}^{k} \theta_i \ \ \Delta GINI\ INCOME_{t-i} + \sum_{i=0}^{k} \rho_i \, \Delta LU_{t-i} + \ \sum_{i=0}^{k} \gamma_i \ \ \Delta CPI_{t-i} + \sum_{i=1}^{k} \alpha_i \ \ \Delta LEXP_{t-i} + \sum_{i=0}^{k} \mu_i \ \ \Delta LTR_{t-i} + \delta_7 EC_{t-1} + \varepsilon_{7t}$ (18)

$\Delta LTR_t = \ \vartheta_8 + \sum_{i=0}^{k} \beta_i \ \Delta ED_{t-i} + \sum_{i=0}^{k} \varphi_i \ \ \Delta LGDP_{t-i} + \sum_{i=0}^{k} \emptyset_i \, \Delta LGDP_{t-i}^2 \ + \ \sum_{i=0}^{k} \theta_i \ \ \Delta GINI\ INCOME_{t-i} + \sum_{i=0}^{k} \rho_i \, \Delta LU_{t-i} + \sum_{i=0}^{k} \gamma_i \ \ \Delta CPI_{t-i} + \sum_{i=0}^{k} \alpha_i \ \ \Delta LEXP_{t-i} + \sum_{i=1}^{k} \mu_i \ \ \Delta LTR_{t-i} + \delta_8 EC_{t-1} + \varepsilon_{8t}$ (19)

The denotes the one period lagged error correction term which is derived from the cointegrationg vector. The residuals ε_{it} are serially uncorrelated and normally distributed. The t statistics of the error correction term (is used to examine the long run causal association while the short run Granger causality is examined by calculating the Likelihood Ratio Statistics (LR) on the first difference of lagged explanatory variables.

3. Results

3.1 ARIMA Forecasting

In the present study a projection of school and college completion is made of the residents of the United States over the period 2018-2022, disaggregated by age groups, sex and ethnicity, using ARIMA model. The Table 1 shows the predicted values of people in the years of school completed, the prediction is based across four age groups. The results show an upward trend though there are fluctuations across some age groups. The Table 2 shows the prediction across different levels of education between the male and female population in the age groups, 25 years and older. There is a declining tendency for lower levels of school completion both at elementary and high school levels, this implies that only some students who are joining schools are going to pursue to its highest levels. The Table 3 reports forecasted completion of all the different school and college levels by the male and the female population of the United States in the age group of 25 -34 years. A decline in completion of the elementary and high school education is predicted for the female population in the age group of 25-34 years. The Table 4 shows the predicted values of completion of education at school and college levels of the male and the female population in the age group of 35-54 years. As far as the male population is concerned the school level completion predicts a declining trend. The same tendency is noticeable for the female population, Table 4. The Table 5 reports the educational completion prediction levels for the male and the female population in the age group 55 years and above. For the female population there is a decline in prediction for school level completion (except for high school level more than four years) in the particular age group, again for the male population a decline tendency is noticed in the completion levels for the elementary level(5 to 8 years) and college level(1 to 3 years). The tables 6 (a) and 6 (b) show the forecast of percentage of students both male and female who complete school and college level of education respectively, disaggregated by race and age classification. Across all races the trend is increasing though the rise among the female population is higher. The figures 1(a) to 1(h) show graphically for all population taken together the forecasting trends for different age groups. The situation is one of rising trends for the age group 25 years and older. However the ager group 25-29 years witnesses fluctuating tendencies.

Sex	Age Group	Fitted ARIMA model	Predicted Values				
			2018	2019	2020	2021	2022
Male	25 years and older	(2,1,0)	105518.7	106597.6	107729.1	108835.4	109953.1
Female	25 years and older	(1,1,0)	113745.3	114966.9	116216.1	117475.8	118739.4
Male	25-34 years	(1,1,0)	22348.53	22555.93	22754.98	22950.56	23144.7
Female	25-34 years	(0,1,1)	23044.85	23762.53	24391.4	24953.68	25467.49
Male	35-54 years	(2,1,0)	40266.45	40234.01	40279.72	40360.14	40485.22
Female	35-54 years	(1,1,0)	42604	43702.94	44563.24	45335.03	46056.45
Male	55 years older	(2,1,0)	40266.45	40234.01	40279.72	40360.14	40485.22
Female	55 years older	(1,1,0)	41891	42305.63	42720.25	43134.87	43549.49

Source: 1947, and 1952 to 2002, March Current Population Survey, 2003 to 2017 Annual Social and Economic Supplement to the Current Population Survey. Compilation: Author

Table 1: Predicted Values of People in Years of School Completed, by Age and Sex, using ARIMA model, 2018-2022.

Sex	Educational Levels	Fitted ARIMA model	Predicted Values				
			2018	2019	2020	2021	2022
Male	Elementary 0 to 4 years	(1,1,0)	1041.29	961.3	903.4	829.9	767.42
	Elementary 5 to 8 years	(0,1,1)	17568.42	17798.73	18029.04	18259.34	18489.65
	High School 1 to 3 years	(2,1,1)	4959.6	4524.53	4126.18	3751.16	3398.1
	High School 4years	(1,1,1)	31749.7	32192.45	32635.19	33077.94	33520.69
	College 1 to 3 years	(1,1,0)	26949.82	27355.83	27761.84	28167.85	28573.86
	College 4 years or more	(2,1,1)	35680.68	36225.83	36768.84	37311.66	37854.47
Female	Elementary 0 to 4 years	(1,1,0)	1108.9	1024.49	995.06	926.03	885.53
	Elementary 5 to 8 years	(0,1,1)	16926.17	17146.08	17365.99	17585.89	17805.8
	High School 1 to 3 years	(0,1,2)	6667.9	6644.77	6665.06	6685.35	6705.64
	High School 4years	(1,1,0)	31489.13	31865.95	32281.42	32711.3	33146.56
	College 1 to 3 years	(1,1,0)	31571.18	32028.65	32499.04	32971.77	33444.93
	College 4 years or more	(2,1,0)	40253.74	41399.54	42459.24	43445	44371.59

Source: 1947, and 1952 to 2002, March Current Population Survey, 2003 to 2017 Annual Social and Economic Supplement to the Current Population Survey. Compilation: Author

Table 2: Predicted Values of People in Years of School Completed, by Age 25 years and older and Sex, using ARIMA model, 2018-2022.

Sex	Educational Levels	Fitted ARIMA model	Predicted Values				
			2018	2019	2020	2021	2022
Male	Elementary 0 to 4 years	(1,1,0)	116.36	109.27	98.64	89.09	79.21
	Elementary 5 to 8 years	(2,1,0)	353.85	266.86	217.93	124.35	100.65
	High School 1 to 3 years	(1,1,0)	1396.43	1381.81	1367.51	1353.11	1338.75
	High School 4years	(1,1,0)	6533.21	6612.51	6692.39	6772.46	6852.57
	College 1 to 3 years	(1,1,1)	6389.15	6500.02	6604.13	6704.2	6801.86
	College 4 years or more	(1,1,0)	7551.12	7551.12	7551.12	7551.12	7551.12
Female	Elementary 0 to 4 years	(0,1,1)	76.15	68.04	59.99	51.93	43.86
	Elementary 5 to 8 years	(2,1,1)	329.14	286.9	259.07	216.11	178.14
	High School 1 to 3 years	(2,1,2)	1161.09	1163.53	1168.47	1151.27	1143.74
	High School 4years	(1,1,0)	1684.74	1099.74	733.51	498.85	321.45
	College 1 to 3 years	(2,1,0)	6602	6651.44	6734.77	6814.79	6903.08
	College 4 years or more	(2,1,1)	9204.15	9383.63	9554.67	9719.31	9878.85

Source: 1947, and 1952 to 2002, March Current Population Survey, 2003 to 2017 Annual Social and Economic Supplement to the Current Population Survey. Compilation: Author

Table 3: Predicted Values of People in Years of School Completed, by Age and Sex, (25-34 years age group) using ARIMA model, 2018-2022.

Sex	Educational Levels	Fitted ARIMA model	Predicted Values				
			2018	2019	2020	2021	2022
Male	Elementary 0 to 4 years	(1,1,0)	380.46	341.41	320.32	286.07	261.46
	Elementary 5 to 8 years	(0,1,2)	1197.83	1098.65	979.06	859.47	739.88
	High School 1 to 3 years	(1,1,0)	2086.22	1954.92	1894.11	1862.47	1842.88
	High School 4years	(1,1,0)	12162.15	12323.43	12492.84	12664.38	12836.47
	College 1 to 3 years	(1,1,0)	10049.74	10152.74	10285	10428.05	10575.09
	College 4 years or more	(1,1,2)	14026.9	14224.03	14423.18	14624.06	14826.42
Female	Elementary 0 to 4 years	(1,1,0)	292.81	264.42	238.52	211.76	185.3

	Elementary 5 to 8 years	(2,1,1)	1123.59	1074.17	1046.38	1009.4	973.09
	High School 1 to 3 years	(1,1,3)	2335.49	2318.41	2306.08	2299.37	2297.99
	High School 4years	(4,1,0)	2327.38	2298.58	2266.04	2248.36	2230.69
	College 1 to 3 years	(0,1,2)	11714.09	11789.55	11963.81	12138.06	12312.32
	College 4 years or more	(1,1,2)	16715.02	17054.09	17386.97	17714.17	18036.17

Source: 1947, and 1952 to 2002, March Current Population Survey, 2003 to 2017 Annual Social and Economic Supplement to the Current Population Survey. Compilation: Author

Table 4: Predicted Values of people in Years of School Completed , by Age and Sex, (35-54 years age group) using ARIMA model, 2018-2022.

Sex	**Educational Levels**	**Fitted ARIMA model**	**Predicted Values**				
			2018	2019	2020	2021	2022
Male	Elementary 0 to 4 years	(3,1,2)	1011.38	1238.37	1318.04	1459.94	1544.62
	Elementary 5 to 8 years	(1,1,0)	1551.92	1492.72	1434.86	1377.3	1319.81
	High School 1 to 3 years	(2,1,1)	2847.39	2874.62	2905.35	2936.83	2969.7
	High School 4years	(2,1,1)	13018.94	13275.3	13512.81	13748.26	13978.49
	College 1 to 3 years	(2,1,0)	186.21	134.62	116.02	101.23	99.1
	College 4 years or more	(1,1,0)	14251.46	14602.81	14903.22	15173.48	15425.9
Female	Elementary 0 to 4 years	(1,1,0)	714.67	689.85	672.19	652.03	632.74
	Elementary 5 to 8 years	(1,1,0)	1816.25	1779.89	1741.6	1702.28	1662.41
	High School 1 to 3 years	(2,1,1)	3232.04	3223.78	3184.63	3159.28	3130.97
	High School 4years	(2,1,2)	16319.03	16456.76	16612.79	16828.24	17095.34
	College 1 to 3 years	(2,1,1)	13514.13	13936.7	14359.35	14777.15	15191.09
	College 4 years or more	(2,1,0)	40253.74	41399.54	42459.24	43445	44371.59

Source: 1947, and 1952 to 2002, March Current Population Survey, 2003 to 2017 Annual Social and Economic Supplement to the Current Population Survey. Compilation: Author

Table 5: Predicted Values of People in Years of School Completed, by Age and Sex, (55 years and older age group) using ARIMA model, 2018-2022.

Completed four years of high school or more (25 years and older) Race (Sex)	**Fitted ARIMA model**	**Predicted Values**				
All Races(Male)	(1,1,2)	2018	2019	2020	2021	2022
		89.42	90.14	90.9	91.7	92.53
All Races(Female)	(1,1,0)	90.69	91.58	92.6	93.7	94.85
White (Male)	(2,1,1)	90.43	91.33	92.37	93.42	94.57
White (Female)	(2,1,2)	91.34	92.37	93.68	95.29	97.17
Black (Male)	(2,1,1)	87.25	87.88	88.56	89.24	89.94
Black (Female)	(2,1,1)	88.65	89.35	90.08	90.83	91.59
Completed four years of high school or more (25-29 years) All Races (Male)	(1,1,2)	2018	2019	2020	2021	2022
		92.5	93.3	94.2	95.2	96.2
All Races (Female)	(0,1,1)	94.36	95.33	96.29	97.26	98.22
White (Male)	(2,1,0)	91.92	92.71	93.6	94.56	95.61
White (Female)	(1,1,0)	94.84	96.1	97.47	98.94	100.5
Black (Male)	(2,1,3)	91.2	91.57	92.12	92.83	93.68
Black (Female)	(1,1,2)	92.73	93.31	93.95	94.63	95.37

Table 6 (a): Predicted Values of People (percent) who have completed high school, by Age, Race and Sex, using ARIMA model, 2018-2022.

Completed four years of high school or more (25 years and older) Race (Sex)	**Fitted ARIMA model**	**Predicted Values**				
All Races(Male)	(2,1,1)	2018	2019	2020	2021	2022
		34.22	34.69	35.18	35.65	36.13
All Races(Female)	(1,1,1)	35.1	35.6	36.2	36.7	37.2
White (Male)	(1,1,1)	34.47	34.95	35.43	35.92	36.41
White (Female)	(4,1,4)	35.52	36.12	36.6	36.91	37.5
Black (Male)	(1,1,1)	21.98	22.03	22.17	22.39	22.65
Black (Female)	(1,1,0)	25.61	26.09	26.45	26.86	27.25
Completed four years of college or more (25-29 years) Race (Sex)	Fitted ARIMA model	Predicted Values				
All Races(Male)	(3,1,2)	2018	2019	2020	2021	2022
		32.67	33.04	32.82	33.17	34

All Races(Female)	(1,1,2)	39.77	40.58	41.05	41.67	42.22
White (Male)	(2,1,2)	31.79	30.18	29.33	29.9	31.28
White (Female)	(2,1,1)	42.04	42.16	42.95	43.56	44.11
Black (Male)	(2,1,3)	20.3	20.87	19.48	20.55	21.22
Black (Female)	(0,1,1)	25.01	25.38	25.75	26.13	26.5

Table 6(b): Predicted Values of People (percent) who have completed college, by Age, Race and Sex, using ARIMA model, 2018-2022.

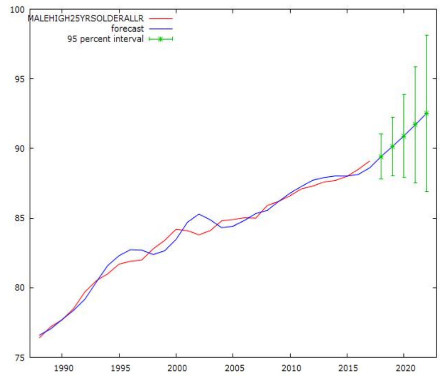

Figure 1(a): Forecast of percent of people who have completed four or more years of high school, (male) (all races)(25 years older age group),(years 2018-2022)

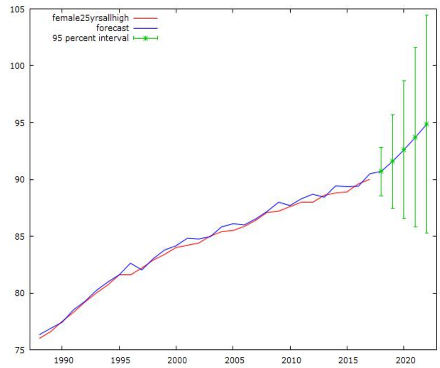

Figure 1(b): Forecast of percent of people who have completed four or more years of high school, (female) (all races)(25 years older age group),(years 2018-2022)

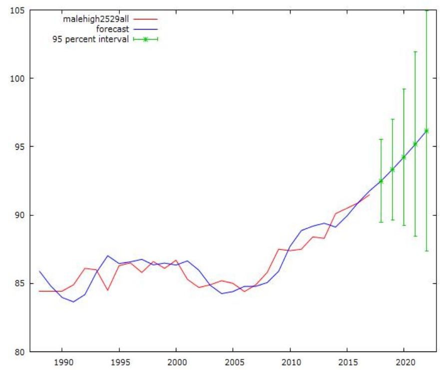

Figure 1(c): Forecast of percent of people who have completed four or more years of high school, (male) (all races) (25-29 age group),(years 2018-2022)

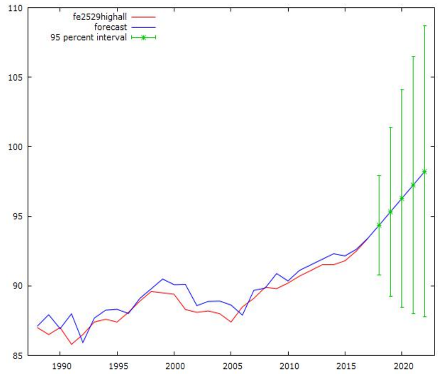

Figure 1(d): Forecast of percent of people who have completed four or more years of high school, (female) (all races) (25-29 age group),(years 2018-2022)

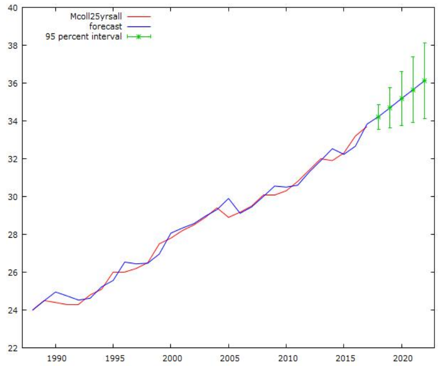

Figure 1(e): Forecast of percent of people who have completed four or more years of college (male) (all races) (25 years older age group), (years 2018-2022)

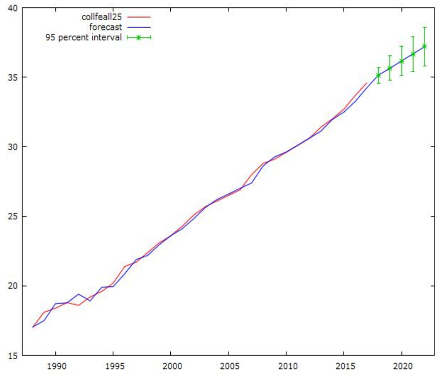

Figure 1(f): Forecast of percent of people who have completed four or more years of college (female) (all races)(25 years older age group),(years 2018-2022)

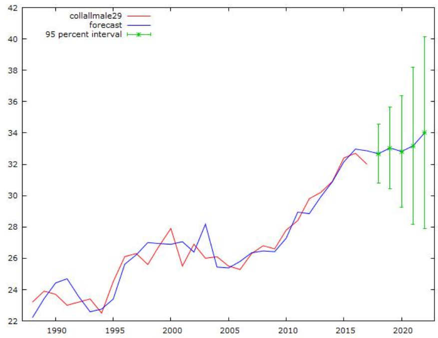

Figure 1(g): Forecast of percent of people who have completed four or more years of college (male) (all races)(25 -29 years age group), (years 2018-2022)

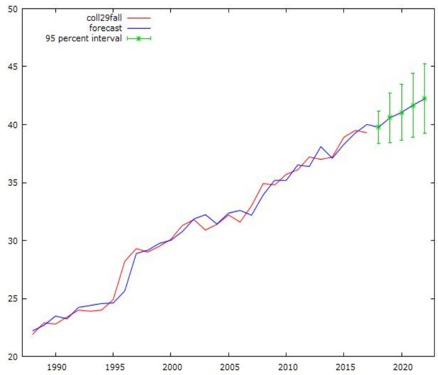

Figure 1(h): Forecast of percent of people who have completed four or more years of college female (all races)(25 -29 years age group),(years 2018-2022)

The situation of completion of college education (forecast) over the period 2018-2022 across the male and female population over the age group 25-29 years show intense fluctuations with a tendency to decline, so it is a cause for serious concern. The prime age group's educational needs in the future call for concerted public action.After explaining school and college educational forecasting trends based on the ARIMA for the United States, the subsequent section attempts to find the causality association of factors responsible for the varied performance in the education system in the United States. A Vector Error Correction Model is chosen since the underlying set of the observations are cointegrated. Further a comparison of univariate forecasting versus multivariate forecasting is made based on the forecast performance measures. The discussion on multivariate model follows the econometric methodology of Unit Root Testing, Johansen Cointegration test and finally building the Vector Error Correction Model.

3.2 Multi Variate Model

Unit Root Tests

The Table 7 reports the results of the unit root test for the concerned variables. Here the Augmented Dickey Fuller test and Phillpis Perron test statistics are reported. The results based on the Table 7 confirm that the underlying series of observations are integrated of order I(1). So we can proceed with the verification of the cointegrating rank of the variables.

Variables	**At Level**		**First Difference**	
	Dickey Fuller Test	**Phillips Perron Test**	**Dickey Fuller Test**	**Phillips Perron Test**
LGDP	-2.00	-11.38	-5.48	-29.73
$L(GDP)^2$	-1.83	-9.94	-5.54	-30.73
GINI INCOME	-1.46	-5.08	-6.51	-35.65
LU	-3.05	-5.70	-5.41	-34.01
CPI	-3.07	-3.34	-5.51	-29.08
LEXP	-2.27	-3.92	-4.81	-28.53
LTR	-1.38	-7.83	-5.11	-32.15
ED	-1.48	-7.0	-5.7	-33.02

Source: Refer to the Table 1(a) for a detailed description of the data sources of the variables chosen

Note: Critical Values at 1%, 5% and 10 % respectively -4.205 -3.524 and -3.194 for Dickey Fuller Test for unit root. Critical Values [Z(rho)], 1%,5% and 10% respectively -24.932; -19.344 and -16.512 for Phillips- Perron Test for Unit root . Compilation Author

Table 7: Unit Root Tests

Cointegration.

Null	Alternative	Statistic	5% Critical Value
Eigen value statistics			
r=0	r=1	176.10	45.28
r≤1	r=2	123.79	39.37
r≤2	r=3	98.97	33.46
r≤3	r=4	20.2442	27.07
r≤4	r=5	0.7843	3.76
Trace statistics			
r=0	r ≥ 1	424.81	124.24
r≤1	r ≥ 2	302..67	47.21
r≤2	r ≥ 3	202.51	29.68
r≤3	r ≥ 4	10.27	15.41
r≤4	r = 5	0.7843	3.76

Source: Refer to the Table 1(a) for a detailed description of the data sources of the variables chosen, Compilation Author

Note: Akaire Criteria was used to select the number lags for the cointegration test, r denotes the number of cointegrated vectors.

Table 8: Johansen–Juselius likelihood cointegration tests

As it has been found that educational attainments rates and the explanatory variables under consideration are integrated I(1), then the cointegration method of Johnsen and Juselius (1990&1992) can be applied. This will help to determine the available cointegrating vectors. Before the application of Johansen technique a sufficient lag length is required for the VAR model estimation. The best specification shows a lag length of 4, so the order of the model is VAR (3). Table 8 presents the cointegration analysis of the long run relationship, according to maximum eigen value criteria, the null hypothesis of no cointegrationis (r=0) is rejected at 5% level of significance in favour of alternative hypothesis Again when trace statistics test is concerned, r=0 is rejected against r ≥ 1 at 5% level of significance. So both trace statistics and maximum eigen value tests confirm the existence of more than one co integrating vector among the variables. As the cointegrating relationship is established the residuals can be used in a Vector Error correction Model (VECM).

Granger causality Tests

After establishing the cointegrating vector among the variables in the model, the residuals are then used as an error correction term in the VEC model, which is obtained from the long the relationship. The VECM shows how the short run dynamics of the time series ultimately converge into a stable long-run equilibrium state. The estimates of the VEC model are reported in the Table 9.

Variables		Likelihood ratio (LR) statistics [p-values]						
	$\Delta \mathrm{ED_t}$	$\Delta \mathrm{LGDP_t}$	$\Delta \mathrm{LGDP_t^2}$	$\Delta \mathrm{GINI\ INCOME_t}$	$\Delta \mathrm{LU_t}$	$\Delta \mathrm{CPI_t}$	$\Delta \mathrm{LEXP_t}$	$\Delta \mathrm{LTR_t}$
$\sum \Delta \mathrm{ED_{t-i}}$	-----	2.79 [0.24]	4.14 [0.12]	0.31 [0.57]	13.81 [0.079]	0.047 [0.82]	30.12 [0.74]	19.95* [0.002]
$\sum \Delta \mathrm{LGDP_{t-i}}$	24.52* [0.001]	---------	27.67* [0.00]	0.12* [0.008]	10.57* [0.0017]	0.18* [0.004]	0.16* [0.002]	0.28 [0.075]
$\sum \Delta \mathrm{LGDP_{t-i}^2}$	18.47* [0.00]	10.93 [0.09]	----------	0.33 [0.27]	31.31* [0.001]	30.12* [0.002]	0.047 [0.067]	0.0079 [0.001]
$\sum \Delta GINI\ INCOME_{t-i}$	27.62* [0.001]	24.53* [0.0014]	12.01 [0.14]	----------	28.69* [0.001]	23.21 [0.31]	36.36* [0.004]	18.29 [0.006]

$\sum \Delta LU_{t-i}$	12.62* [0.0028]	4.14* [0.001]	23.31 [0.43]	0.28* [0.0054]	-------	1.66 [0.23]	0.81 [0.70]	10.96* [0.009]
$\sum \Delta CPI_{t-i}$	31.19* [0.0079]	23.21 [0.21]	0.31* [0.001]	0.21* [0.002]	40.68* [0.001]	-------	32.12 [0.54]	43.21 [0.58]
$\sum \Delta LEXP_{t-i}$	22.58 [-0.45]	3.03 [0.32]	5.15* [0.01]	7.66* [0.0071]	15.80 [0.53]	0.02 [0.73]	------	0.12* [0.013]
$\sum \Delta LTR_{t-i}$	66.12 [-0.51]	0.57* [0.003]	10.25* [0.001]	42.23* [0.002]	0.28 [2.25]	12.23 [0.51]	0.80 [0.30]	-------
EC_{t-1} [t statistics]	-0.614* [-0.0087]	-0.22* [-0.072]	-0.27* [-0.025]	-0.75* [-0.024]	-0.52 [-2.88]	-0.39* [0.042]	0.96 [-4.31]	0.02 [-2.21]

Note:(*) denotes the level of significance at 5 percent. The optimal lag order is obtained from Akaike's information criterion (AIC). Δ denotes the difference of the times series of observations. Compilation Author

Table 9: Granger Causality results –VECM

The Granger Causality results are reported in the Table 9. As far as the long run causality effect is concerned it is based on the significance of the one period lagged error correction term [ECT(-1)]. The error correction coefficients are negative and less than unity for all the variables except LU, LEXP and LTR. The results imply that long run equilibrium is attainable and there is no over correction. Further, the error correction terms are statistically significant. So a bidirectional causality exists in the long run across educational completion, income growth, income inequality, and inflation. Turning to the discussion on short run causality we find that economic growth, income inequality, urbanization and inflation are statistically significant at 1 percent level when educational development is the dependent variable, so economic growth, income inequality, urbanization and inflation Granger cause educational development in the short run. Again educational development does Granger cause income growth or inequality in the short run. Income growth, inflation, and trade Granger cause urbanization. In sum there is unidirectional causality in the short run from income growth and inequality towards educational development. After establishing the causality association among the variables, the subsequent task is to forecast educational achievements in the VEC model, to compare with the univariate ARIMA forecasting.

Years	**Forecasted Observation**
2018	35.22
2019	36.21
2020	35.33
2021	36.21
2022	36.13

Compilation Author

Table 9 (b): Post Sample period Forecast, educational attainments, The United States VEC model

Years	**Forecasted Observation**
2018	36.12
2019	37.13
2020	36.12
2021	37.12
2022	37.23

Compilation Author

Table 9 (c): Post Sample period Forecast, educational attainments, The United States, ARIMA model

Measures	Statistics
Mean Forecast Error	0.03
Mean Absolute Error	0.1
Mean Absolute Percentage Error	0.45
Root Mean Squared Error	0.10
Theil's U-Statistics	0.51
Mean Percentage Error	0.08

Compilation: Author

Table 10 (a): Forecast Performance Measure VEC model

Measures	Statistics
Mean Forecast Error	0.002
Mean Absolute Error	0.031
Mean Absolute Percentage Error	0.32
Root Mean Squared Error	0.01
Theil's U-Statistics	0.649
Mean Percentage Error	0.01

Compilation: Author

Table 10 (b): Forecast Performance Measure ARIMA model

The forecast based estimates from the VEC model is reported in the Table 9 (b), the tendency is towards a stable level of educational completion in the United States. So educational accomplishments are not rising in the near future in the United States. The forecast accuracy estimates based on the VEC model is reported in the Table 10 (a). The forecast accuracy estimates based on the ARIMA model is show in the Table 10 (b), they show better results.

This can be owing to some inherent assumptions in the multi variate VEC model. In the VEC model usually it is supposed that the model will be projected with the similar number of lags for all the variables in the system of equations. So, this causes the occurrence of the low lag lengths. Moreover, the VEC model makes enriched parameterization, where the degrees of freedom are limited in number. Thus preference bias in the choosing of the lag length and the parameterization gives relatively lower forecast precision estimates. As the possibility of risk related with forecasting is crucial, one should consider forecasting results on a collection of tools to reduce the risks, rather than depending on one model. Achieving imprecise approximation of the parameters in the VEC model is not uncommon because the number of parameters often outnumber the number of observations. VEC model is used to explore, how much change in the dependent variable is detected with one unit change in any particular independent variables. We apply VEC model to investigate factors associated with educational completion in the United States. The study shows the existence of long-run stationary relations across the dependent variable namely educational completion at college level and the explanatory factors like income growth, income inequality and inflation.

4. Discussion

A large body of the recent literature has discussed about the relation between income inequality, economic growth and its impact on human development. The seminal paper of Kuznets (1955)[13] discuss that at the initial levels of economic growth income inequality rises , reaches an optimum level and then declines subsequently when economic growth advances further. A number of empirical studies have explored the Kuznets hypothesis for example, Ahluwalia (1974)[14]; Robinson (1976)[15]; Stewart (1978)[16],Winegarden (1979)[17]; Nielson and Alderson (1995)[18]; Checchi(2000)[19] and Wells (2006)[20]. At the theoretical level Galor and Zeira (1993)[21] discuss that inequality and economic growth exhibit an inverse relation. According, to Aghion and Bolton (1997)[22] owing to imperfections in the capital market the poorer households are unable to invest for human capital formation, thus the gains from productivity remains unaltered. Interestingly there are variant views on the relation between growth and inequality, Galor and Tsiddon (1997)[23] discuss that technological change enhance the payment of factors in the sectors with high economies of scale; Siebert (1998)[24] and Furman and Stiglitz (1998)[25] observe that inequality fosters flexibility in the labour market which intensifies the effort thereby generating positive economic growth. Park (1996a)[26] and (1998)[27] focusses the analyses of the impact of income inequality on the society. Further Park (1986a)[26] discuss that income inequality may lead to political instability. Forbes (2000)[28] and Partridge (1997[29], 2005[30]) based on empirical discussion conclude that inequality has positive bearing on the levels of economic growth. Followed by Becker's (1964)[31] work on the importance of human capital in fostering economic growth, a large number of studies explore the importance of education in shaping income inequality both at the theoretical and at the empirical levels. Park (1996b)[32] explore the behaviour of different levels of enrollment and their influence on economic growth. Ahluwalia (1976)[33]; Barro (2000)[34] and Alderson & Nielsen (2002)[35] focus their discussion on the flow variables of the educational sector namely institutional enrollment, again Winegarden (1979)[36], Ram (1984)[37], Gregorio and Lee (2002)[38] use the stock component of education namely mean years of schooling to examine its impact on labour force participation and productivity growth. Tinbergen (1972)[39] and Park (1996b)[32] discuss that higher schooling years and less skewness in schooling has positive impact on growth and reduces income inequality, however Ram's (1984) discussion refute the conclusion established by Tinbergen (1972)[39] and Park (1996b)[32]. Barro (2000)[34] concludes that the impact of education on income inequality varies with variation in the levels of education. Alderson and Nielsen's (2002)[35] conclude that in the developed nations income inequality negatively affects average years of school completion, (such finding is in conformity with the present study). The ARIMA forecasting establishes that there is a declining future rates of completion when disaggregated across age groups Capital market imperfections lack of social mobility, intergenerational human capital formation, and bequests of wealth to children by rich parents has strong negative implications on the dispersion of income.

Chani *et al*. (2014)[40] observe for Pakistan (based on Johanson method of co-integration and Granger Causality tests) that there is bidirectional causality between educational inequality and income inequality over the period 1973-2009. According to Jenkins (1995)[41] the progress of an economy is positively related to the investment of skill formation of labour. The study shows that unequal investment in human capital formation exacerbates income inequality in a country. For the different cities in the United States, Moretti (1999)[42] concludes that differential human capital formation draw workers to cities which have high wage structure. This process leads to skewness in wages among workers. Based on cross country comparisons (Germany and the United States), Freeman and Schettkat (2001)[43] observe that differences in learning abilities between these two countries explain the variation in earnings inequality in these two countries. The study concludes that the United States has more inequality in learning accomplishments and this is reflected in greater income inequality than Germany. The findings of Freeman and Schettkat (2001)[43] with respect to the United States reinforces the results based in the current exercise.

Lopez *et al* (1998)[44] explain that human capital formation measured through average educational completion is not contributing to expansion in the growth of the economy, in a statistically significant way. According to the study the initial distribution of human capital is important. Bhargava *et al* (2001)[45] opine through an empirical study across both developing and developed nations that poverty generates a vicious cycle of ill health which is detrimental to the productivity levels of the workers in the developing nations. So developed nations have high productivity rates, this leads to earnings differential across the developing and the developed nations. According to Appiah and McMahon (2002)[46] for African countries long term investment in human and physical capital is a necessary prerequisite to the enable the nations to raise their productivity levels at par with other mature economies of the world. Yang (2002)[47] discusses in the context of China that differential human capital formation between the rural and urban areas is positively related to wage inequality between the rural and urban areas. This study obtains an association with urbanization and educational attainment in the United States, Table 9. Chani *et al* (2011)[40] discuss that poverty intensifies income inequality in Pakistan, so investment in human development can ameliorate poverty. Jamal and Khan

(2003)[48] observe that for the different districts of Pakistan the variation in the levels of income is owing to differential investment in skill formation. In line with the discussion of the present study, the writer concludes that human capital inequality is positively related to income inequality. Morgues and Carter (2005)[49] discuss about the importance of social capital formation in fostering economic prosperity across the poorer nations. For a panel set of OECD and non OECD countries Földvári and Leeuwen (2010)[50] conclude through an empirical investigation that the effect of education inequality on income inequality is insignificant for the set of non OECD countries. Further the study shows that more equitable investment in human capital formation is not positively related to economic progress. Checchi (2001)[51] concludes that inequality in income distribution of a nation has a positive association with the capital output ratio and on the behaviour of government spending towards the education sector. Checchi and de Werfhorst (2014)[52] explored the relation between income inequality, educational distribution, and educational policies. The paper concludes that policies related to education have a significant effect on the quality and quantity of learning. Further attainment of education and its distribution has a significant bearing on the earnings distribution function. Thus the study opines that educational policies go a long way in affecting income distribution. According to Galor (2012)[53] households facing credit constraints are unable to make educational investments, this may ultimately affect earnings inequality. Based on the results of cognitive test scores in the United States, Blau and Kahn(2005)[54] discuss that higher the inequality in the achievement levels in the test scores greater is the wage differential in the US economy.

5. Conclusion

In searching for explanation on the differences of skill formation and educational attainments, the study observes interesting variations across different age groups and at different levels of educational completion in the United States. There is an urgent need to address these heterogeneity in levels of educational performance. This paper makes univariate ARIMA based forecasting of educational attainment in the United States disaggregated by age, sex and racial groups. The paper further looks into the major determinants of educational attainments and explores the long run and short run causal association among the variable by developing a Vector Error Correction framework. A bidirectional long run association exists across educational completion, income growth, income inequality, and inflation. Thus policy on educational expansion needs to address the important issue of income inequality.

The United States faces a situation of rising inequality however the scope for redistributive polices are controlled by the fiscal regime. Nevertheless, equitable distribution in the labour market may increase the scope for a tendency towards equitable society. Investment in education in an equitable way could be a proper policy towards redistribution of income and wealth. It should be noted that the de facto opportunities of educational resources are also important at given level of distribution. As far as the United States is concerned many of the youth start education but do not complete. The reasons for discontinuance may be numerous; lack of motivation, insufficient proficiency, and family background. Proper attention to the quality dimension along with the quantitative aspect of education would ensure rising productivity growth from education and it may outpace the productivity gains from education associated with high levels of specialization.

Conflict of Interest

The author has reported “no conflict of interest”.

Funding and Acknowledgements

No funding is associated with this work.

References

1. Bernstein, Jared. The impact of inequality on growth. 2013; Center for American Progress, Washington.
2. Stiglitz, J. E. The price of inequality: How today's divided society endangers our future. 2012; WW Norton Company.
3. Alesina, A., Roderick, D. Distributed politics and economic growth. Quarterly Journal of Economics .1994; 109 (2): pp. 465-490.
4. Alvaredo, F., Atkinson, A. B., Piketty, T., Saez, E. The top 1 percent in international and historical perspective.

Journal of Economic perspectives. 2013; 27(3): 3-20.
5. Hoeller, P., I. Joumard, M. Pisu, *et al*. Less Income Inequality and More Growth – Are They Compatible? Part 1. Mapping Inequality Across the OECD. 2012; OECD Economics Department Working Paper, No. 924.
6. Chen, S., Ravallion, M. The developing world is poorer than we thought, but no less successful in the fight against poverty. The Quarterly Journal of Economics. 2010; 125(4): 1577-1625.
7. Oded, G. Inequality, human capital formation, and the process of development. In Eric A. Hanushek, Stephen Machin and Ludger Woessmann (eds). Handbook of the Economics of Education. Vol. 4, 441-49. North Holland: Elsevier; 2011.
8. CPS Historical Time Series Tables. United States Census. 2017. Retrieved from URL: https://www.census.gov/data/tables/time-series/demo/educational-attainment/cps-historical-time-series.html.
9. World Bank's World Development Indicators (WDI) 2017. Retrieved from URL:. http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators.
10. IMF, Government Finance Statistics Yearbook. Retrieved from URL: https://data.imf.org/?sk=a0867067-d23c-4ebc-ad23-d3b015045405.
11. Box, G. E. P, GM Jenkins. Time Series Analysis, Forecasting, and Control. Francisco Holden-Day; 1970.
12. Sims, C. A. Macroeconomics and reality. Econometrica: Journal of the Econometric Society. 1980; 1-48.
13. Kuznets, S. Economic growth and income inequality. The American Economic Review. 1955; 45(1): 1-28.
14. Ahluwalia, M.S. Income Inequality: Some dimensions of the problem, In H. Chenery *et al*., (eds.). Redistribution with Growth .Oxford: Oxford University Press, 1974; 3-37.
15. Robinson, S. A note on the U hypothesis relating income inequality and economic development. The American Economic Review. 1976; 66(3): 437-440.
16. Stewart, F. Inequality, technology and payments systems. World Development. 1978; 6(3): 275-293.
17. Winegarden, C. R. Schooling and income distribution: Evidence from international data. Economica. 1979; 83-87.
18. Nielsen, F., Alderson, A. (1995). Income inequality, development, and dualism: results from an unbalanced cross-national panel'. American Sociological Review.1995; 60 (5): 674-701.
19. Checchi, D. Does Educational Achievement Help to Explain Income Inequality? Working Paper No. 208 (Helsinki: United Nations University, World Institute for Development Economics Research; 2000.
20. Wells, R. Education's effect on income inequality: an economic globalisation perspective. Globalisation, Societies and Education. 2006; 4(3): 371-391.
21. Galor, O., Zeira, J. Income distribution and macroeconomics. The review of economic studies. 1993; 60(1): 35-52.
22. Aghion, P., Bolton, P. A theory of trickle-down growth and development. The Review of Economic Studies. 1997; 64(2): 151-172.
23. Galor, O., Tsiddon, D. Technological progress, mobility, and economic growth. The American Economic Review; 1997: 363-382.
24. Siebert, H. Commentary: Economic consequences of income inequality. In Symposium of the Federal Reserve Bank of Kansas City on Income Inequality: Issues and Policy Options p. 1998: 265-281.
25. Jason Furman Joseph E. Stiglitz. Economic Consequences of income inequality. Proceedings - Economic Policy Symposium - Jackson Hole. Federal Reserve Bank of Kansas City. p.221-263; 1998.
26. Park, K. H. Income inequality and economic progress: An empirical test of the Institutionalist approach. American Journal of Economics and Sociology. 1996 a; 55(1): 87-96.
27. Park, K. H. Distribution and growth: cross-country evidence. Applied Economics. 1998; 30(7): 943-949.
28. Forbes, K. J. A reassessment of the relationship between inequality and growth. American economic review. 2000; 90(4): 869-887.
29. Partridge, M. D. Is inequality harmful for growth? Comment. The American Economic Review. 1997; 87(5): 1019-1032.
30. Partridge, M. D. Does income distribution affect US state economic growth? Journal of Regional Science. 2005; 45(2): 363-394.
31. Becker, G. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. New York: Columbia Univ. Press; 1964.
32. Park, K. H. Educational expansion and educational inequality on income distribution. Economics of education review. 1996b; 15(1): 51-58.
33. Ahluwalia, M. S. Inequality, poverty, and development. Journal of development economics. 1976; 3(4): 307-342.
34. Barro, R. J. Inequality and Growth in a Panel of Countries. Journal of economic growth. 2000; 5(1): 5-32.
35. Alderson, A. S., Nielsen, F. Globalization and the great U-turn: Income inequality trends in 16 OECD countries. American Journal of Sociology. 2002; 107(5): 1244-1299.
36. Winegarden, C. R. Schooling and income distribution: Evidence from international data. Economica. 1979; 83-87.
37. Ram, R. Population increase, economic growth, educational inequality, and income distribution: Some recent evidence. Journal of Development Economics. 1984; 14(3): 419-428.
38. Gregorio, J. D., Lee, J. W. Education and income inequality: new evidence from cross‐country data. Review of

income and wealth. 2002; 48(3): 395-416.
39. Tinbergen, J. The impact of education on income distribution. Review of Income and wealth. 1972; 18(3): 255-265.
40. Chani, M. I., Jan, *et al.* Human capital inequality and income inequality: testing for causality. Quality Quantity. 2014; 48(1): 149-156.
41. Jenkins, H. Education and production in the United Kingdom. Paper No.101 Oxford: Nuffield College; 1995.
42. Moretti, E. Estimating the social return to education: evidence from repeated cross-sectional and longitudinal data. Center for Labor Economics Working Paper, 22; 1999.
43. Freeman, R., Schettkat, R. Skill compression, wage differentials, and employment: Germany vs the US. Oxford Economic Papers.2001; 53 (3), 582-603.
44. Lopez, R., Thomas, V., Wang, Y. Addressing the education puzzle: the distribution of education and economic reform. The World Bank.; 1999.
45. Bhargava, A., Jamison, D. T., Lau, *et al*. Modeling the effects of health on economic growth. Journal of health economics. 2001; 20(3): 423-440.
46. Appiah, E. N., McMahon, W. W. The social outcomes of education and feedbacks on growth in Africa. Journal of Development Studies. 2002; 38(4): 27-68.
47. Yang, D. T. What has caused regional inequality in China? China Economic Review. 2002; 13(4): 331-334.
48. Jamal, H., A. J. Khan. The changing profile of regional inequality. The Pakistan Development Review. 2003; 113-123.
49. Mogues, T. and Carter, M. R. Social Capital and the reproduction of economic inequality in polarized societies. Journal of Economic Inequality. 2005; 3: 193–219.
50. Földvári P, Leeuwen B.V. Educational inequality in Europe, 1870-2000. Presented in: Workshop on human capital in economic history: measurement, determinants and implications, Tuebingen, Germany; 27–29 August, 2010.
51. Checchi, Daniele, Education, Inequality and Income Inequality Research Paper No. 52. 2001. Retrieved from SSRN: https://ssrn.com/abstract=1094825.
52. Checchi, Daniele, van de Werfhorst, Herman. Educational Policies and Income Inequality. IZA Discussion Paper No. 8222; 2014. Retrieved from SSRN: https://ssrn.com/abstract=2450413.
53. Galor, O. The demographic transition: causes and consequences. Cliometrica. 2012; 6(1): 1-28.
54. Blau, F. D., Kahn, L. M. Do cognitive test scores explain higher US wage inequality? Review of Economics and statistics. 2005; 87(1): 184-193.