Tax Revenue and Economic Growth:
An Empirical Analysis for Pakistan

1Masood Mashkoor, 2Sohaib Yahya and 2Syed Ammad Ali

1Department of Commerce, Federal Urdu University, Karachi, Karachi, Pakistan
2Applied Economics Research Centre, Karachi, University of Karachi, Pakistan

Abstract: This study inspects the relationship between tax revenues and the rate of economic growth for Pakistan. The study will focus on the perception that the low ratio of direct to total taxation promotes high economic growth. For the testing of this hypothesis the study consist on annual data from 1973 to 2008. This study analysed the relationship between total tax revenues, Direct tax, gross domestic saving and the rate of economic growth. The results show all the coefficients are statistically significant and the coefficient of error correction term shows low rate of convergence in the long-run. In the short run Saving to GDP ratio causes; the Real GDP Growth, Direct Tax to GDP ratio and Direct tax to Total tax ratio, Direct tax to GDP ratio causes the Real GDP growth, similarly Direct tax to total tax ratio Granger causes Real GDP growth and Direct Tax to GDP ratio but in the long run direct tax to total tax ratio do not cause GDP growth.

Key words:

INTRODUCTION

Taxes play a vital role in economic planning and development, they are the main source of public revenue even economic policies are based on expected tax revenue and the tax policy is a fundamental component of economic policies for every country, in order to maintain their global competitiveness and growth, it is desirable for every government to generate tax revenue to finance essential expenditures without recourse to excessive public sector borrowing, not only this, but also it should minimize disincentive effects on economics activities and inequality. These days the world has become a global village and become more competitive, so to keep the economy competitive any government has to maintain a competitive tax structure in order to attract capital, specialized work and technology which are essential elements for maximizing economic growth.

Theoretical framework allows us to catalogue the different channels through which taxes might cause output growth. First, a higher taxes can depress the investment rate, or the net growth in the capital stock, through high statutory tax rates on corporate and individual income, high effective capital gain tax rates and low depreciation allowances. Second, Tax policy can also discourage productivity growth by reducing research and development (R&D) and the development; if there would be any subsidy (negative tax) it will boost the research activities whose spillover effects can potentially enhance the productivity of existing labor and capital. Third, taxes may reduce the work incentive which will reduce the labor force participation and hours of work, or it may also create biased occupational choice or the acquisition of education, skills and training. Fourth, heavy taxation on labor supply can distort the efficient use of human capital by discouraging workers from employment in sectors with high social productivity but a heavy tax burden and fifth, tax policy can also affect the marginal productivity of capital by distorting investment from high taxed sectors to low taxed sectors, Harberger [1, 2].

Tax structure varies all around the world with the prime motive of attaining maximum revenue with minimum distortion different countries have different philosophies about taxation and have different methods for collection, in the same manner countries have different uses of their revenue which affect the growth differently. Atkinson [3], Castles and Dowrick [4], Agell et al [5], all argue that the different uses of total

Corresponding Author: Masood Mashkoor, Department of Commerce, Federal Urdu University, Karachi, Karachi, Pakistan.
government expenditure affect growth differently and a similar argument applies to the way tax revenue is raised. During the last few decades, many countries have increased taxation quite dramatically, while the others are following the suit. Some countries have incorporated value-added taxation and some are going to incorporate such as Pakistan.

A pioneer study in this regard was conducted by Solow [6], this Neo-Classical growth model implies that taxes do not affect the steady state growth. In other word, tax policy though distortional, yet has no impact on long-term economic growth rates and total factor productivity, on the other hand Endogenous growth theory by Romer [7] emphasizes factors such as “spillover” effect and “learning by doing” by which firms specific decision to invest in capital and R&D, or individual investment in human capital, can yield positive external effects that benefit the rest of the economy, in this model government spending and tax policies can have a long-run of permanent growth effects.


Anastassiou and Dritsaki [12] examined the relationship of the rate of economic growth to the ratio of gross savings to GDP, to the marginal direct tax rate and tax revenues using annual data for the period 1965-2002, found a one-way causal relationship between the marginal direct tax rate and the rate of economic growth with direction from the marginal direct tax rate to the rate of economic growth, as well as between tax revenues and the rate of economic growth going in the same direction as before. Moreover, there is a one-way causal relationship between the ratio of gross savings to GDP and the marginal direct tax rate as well as to the tax revenues, with direction from gross savings to the above variables, while there is no causal relationship between the rate of economic growth and the ratio of gross savings to GDP. On the contrary, there is a bilateral causal relationship between the marginal direct tax rate and tax revenues.

Arnold [13] examines a set of panel growth regressions for 21 OECD countries, found a significant affect of taxes on growth. Johansson et. al. [14] also found that taxes cause growth. According to his finding Corporate taxes are found to be most harmful for growth, followed by personal income taxes and then consumption taxes. Recurrent taxes on immovable property appear to have the least impact. A revenue neutral growth-oriented tax reform would, therefore, be to shift part of the revenue base from income taxes to less distortive taxes such as recurrent taxes on immovable property or consumption.

This study aims to investigate the relationship between taxation mix and the rate of economic growth in Pakistan looking in particular if there is any evidence that taxation variables have a causal role in the process of economic growth. In the empirical analysis of this study we use annual data for the period 1973 till 2008 for the examined variables.

The remainder of the study is as follows: the next section presents the data of the study together with the ARDL. The section 2.1 deals with Dickey-Fuller tests and examines stationarity of the data. In the next section 2.2 we discuss the ARDL model specification. Section 3 gives details of the results including the Granger causality tests. The fourth section is the final section of conclusion.

**Model Specification and Data:** All series examined in the study, Real and Nominal GDP, Gross domestic savings, Direct tax revenues and Total tax revenues, are collected from SBP 50 Years Statistics of Pakistan and latest economics survey of Pakistan 2008-09. The data used for the analysis of this investigation are annual of 36 years, covering the period from 1973 to 2008. All the series except the real GDP are taken on the current prices to inspect the role tax ratios and saving ratio to the current GDP on the economic growth.

For the analysis of causal relationship between Economic Growth, Savings to GDP ratio, Direct tax to GDP ratio and the proportion of Direct taxes to total tax revenue we use the following general multivariate VAR model:
\[
\ln(GDP_{2000})_t - \beta_0 + \sum_{i=1}^{K} \beta_i \ln\left(\frac{GDS}{GDP}_{t-i}\right) + \sum_{i=1}^{K} \delta_i \ln\left(\frac{DTAX}{GDP}_{t-i}\right) + \sum_{i=1}^{K} \alpha_i \ln\left(\frac{DTAX}{ITAX}_{t-i}\right) + \sum_{i=1}^{K} \theta_i \ln(GDP_{2000})_{t-i} + v_t
\]  

(1)

Where:
\[
\ln(GDP_{2000})_t = \text{Logarithm of the real GDP for t time period (LGDP2000)}
\]
\[
\ln\left(\frac{GDS}{GDP}\right)_{t-i} = \text{Logarithm of ratio of the gross domestic saving to gross domestic product at current prices (LSGR)}.
\]
\[
\ln\left(\frac{DTAX}{GDP}\right)_{t-i} = \text{Logarithm of ratio of the direct tax (Income tax + Corporate Tax + Wealth Tax + Gift tax and Estate duty) (DTAX) to the gross domestic product (GDP) in spot prices (LDTGR).}
\]
\[
\ln\left(\frac{DTAX}{ITAX}\right)_{t-i} = \text{Logarithm of ratio of direct tax to indirect taxes (Custom duties + Federal Excise + Sales Tax + Surcharge + Stamp) (LDT\_TTR).}
\]

All data are expressed in logarithms in order to reduce heteroscedasticity and non-stationary effects of time series variables. For evaluating the influences of savings to GDP ratio, Direct tax to GDP ratio and the proportion of direct taxes to total tax revenue on economic growth, suitable diversification is important as in the analysis of time series data the analysis fails when time series are non-stationary and their linear combination is stationary, if this is the case the variables may be co-integrated, for checking the stationarity of our time series data, we used an Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin unit root tests on individual time series to provide an evidence whether the variables are integrated.

**Unit Root Test:** The results of the unit root test as appear in Table 1. The minimum values of the Schwarz (SC) statistics have provided the better structure of the ADF equations as well as the relative numbers of time lags, under the indication “Lag”. The results of Table 1 suggest that according to the results of Augmented Dickey-Fuller (ADF) test, the null hypothesis of unit root can not be rejected at 5% of level of significance for all the variables except logarithm of direct tax to tax ratio (LDT\_TTR) But the Kwiatkowski-Phillips-Schmidt-Shin suggests that null hypothesis of stationarity can be accepted for all the variables except Logarithm of GDP\_2000.

**Auto Regressive Distributed Lag (ARDL) Augmented by Level Variables Model:** After determining that the logarithms of the model variables are not integrated of order 1 and may also be cointegrated we can analyse the results using Auto Regressive Distributed Lag augmented by level variables model suggested by Pesaran and Shin [15], Pesaran [16] and Pesaran et al. [17] to ascertain the direction of causation between variables. There are advantages of using this approach as an alternative of the conventional Johansen [18] and Johansen and Juselius [19]. While the conventional cointegration method estimates the long run relationships within a context of a system of equations, the ARDL method employs only a single reduced form equation Pesaran and Shin [20].

The ARDL approach does not involve pre-testing variables, which means that the test on the existence relationship between variables in levels is applicable irrespective of whether the underlying regressors are purely I(0), purely I(1) or mixture of both. This feature alone, given the characteristics of the cyclical components of the data, makes the standard of cointegration technique unsuitable and even the existing unit root tests to identify the order of integration are still highly questionable. Furthermore, the ARDL method avoids the larger number of specification to be made in the standard cointegration test. These include decisions regarding the number of endogenous and exogenous variables (if any) to be included, the treatment of deterministic elements, as well as the optimal number of lags to be specified. The empirical results are generally very sensitive to the method and various alternative choices available in the estimation procedure Pesaran and Smith [21]. With the ARDL, it is possible that different variables have different optimal lags, which is impossible with the standard cointegration test. Most importantly, the model could be used with limited sample data (30 observations to 80 observations in which the set of critical values were developed originally by Narayan [22] by using GAUSS. Basically, the ARDL approach to cointegration involves estimating the conditional error correction (EC) version of the ARDL model for Economic Growth and its determinants.
\[
\Delta \ln(GDP_{2000})_t = \beta_0 + \sum_{i=1}^{K_1} \beta_i \Delta \ln \left( \frac{GDS}{GDP} \right)_{t-i} + \sum_{i=1}^{K_2} \delta_i \Delta \ln \left( \frac{DTAX}{GDP} \right)_{t-i} + \sum_{i=1}^{K_3} \eta_i \Delta \ln \left( \frac{ITAX}{GDP} \right)_{t-i} + \\
\sum_{i=1}^{K_4} \theta_i \Delta \ln(GDP_{2000})_{t-i} + \lambda_1 \ln(GDP_{2000})_{t-1} + \lambda_2 \Delta \ln \left( \frac{GDS}{GDP} \right)_{t-1} + \lambda_3 \Delta \ln \left( \frac{DTAX}{GDP} \right)_{t-1} + \lambda_4 \ln \left( \frac{ITAX}{GDP} \right)_{t-1} + v_t
\]

Where \( \Delta \) is first-difference operator and \( K_1, K_2, K_3 \) and \( K_4 \) are the optimal lag length. The F test is used for testing the existence of long-run relationship. When long-run relationship exist, F test indicates which variable should be normalized. The null hypothesis for no cointegration among variables in Equation 2 is \( H^0, \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0 \) against the alternative hypothesis \( H^*, \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \neq 0 \). The F-test has a non-standard distribution which depends on:

- Whether variables included in the model are I(0) or I(1),
- The number of regressors and
- Whether the model contains an intercept and/or a trend.

Given a relatively small sample size in this study of 36 observations, the critical values used are as reported by Narayan [22] which based on small sample size between 30 and 80. The test involves asymptotic critical value bounds, depending whether the variables are I(0) or I(1) or a mixture of both. Two sets of critical values are generated which one set refers to the I(1) series and the other for the I(0) series. Critical values for the I(1) series are referred to as upper bound critical values, while the critical values for I(0) series are referred to as the lower bound critical values. If the F test statistic exceeds their respective upper critical values, we can conclude that there is evidence of a long-run relationship between the variables regardless of the order of integration of the variables. If the test statistic is below the upper critical value, we cannot reject the null hypothesis of no cointegration and if it lies between the bounds, a conclusive inference cannot be made without knowing the order of integration of the underlying regressors. If there is evidence of long-run relationship (cointegration) of the variables, the following long-run model is estimated.

\[
\ln(GDP_{2000})_t = \lambda_0 + \lambda_1 \ln(GDP_{2000})_{t-1} + \lambda_2 \Delta \ln \left( \frac{GDS}{GDP} \right)_{t-1} + \lambda_3 \Delta \ln \left( \frac{DTAX}{GDP} \right)_{t-1} + \lambda_4 \ln \left( \frac{ITAX}{GDP} \right)_{t-1} + v_t
\]

Pesaran and Shin [33]-recommended choosing a maximum of 2 lags. And the selection criterion is based on either Akaike Information Criterion (AIC) or Schwarz Bayesian Criterion (SBC).

The ARDL model of short-run dynamics can be derived by using the error correction model (ECM) of the following form.

\[
\Delta \ln(GDP_{2000})_t = \beta_0 + \sum_{i=1}^{K_1} \beta_i \Delta \ln \left( \frac{GDS}{GDP} \right)_{t-i} + \sum_{i=1}^{K_2} \delta_i \Delta \ln \left( \frac{DTAX}{GDP} \right)_{t-i} + \sum_{i=1}^{K_3} \eta_i \Delta \ln \left( \frac{ITAX}{GDP} \right)_{t-i} + \\
\sum_{i=1}^{K_4} \theta_i \Delta \ln(GDP_{2000})_{t-i} + \gamma_i ECM_{t-i} + v_t
\]

Where \( ECM_{t-i} \) is the error correction term generated using the Equation 3 and can be defined as:

\[
ECM_t = \ln(GDP_{2000})_t - \lambda_0 - \lambda_1 \ln(GDP_{2000})_{t-1} - \lambda_2 \Delta \ln \left( \frac{GDS}{GDP} \right)_{t-1} - \lambda_3 \Delta \ln \left( \frac{DTAX}{GDP} \right)_{t-1} - \lambda_4 \ln \left( \frac{ITAX}{GDP} \right)_{t-1}
\]

The coefficients of Equation 4 shows short-run dynamics except the \( \lambda \) which shows the speed of adjustment.
RESULTS

To examine the long run relationship among variables using the Equation 2, as suggested by Pesaran et al. [17] and Narayan [22], the order of an ARDL(k1,k2,k3,k4) model in 4 variables (LSGR, LDTGR, LDT_TFR, GDP2000) were selected by searching across the 4^4 = 256 ARDL Models, spanned by K = 1, 2 ... 4 and i=1, 2, ... 4 using the AIC criterion as suggested by Pesaran et al. [17] the resulted choice of an ARDL(1,3,3,2) specification. The calculated F-statistics for the cointegration test as displayed in Table 2, suggest that the F-statistics are significant at 1% level of significance in restricted intercept with no trend and significant at 5% level of significance in restricted intercept with trend.

The long run coefficients obtained from the ARDL model by normalizing it on the GDP2000 as shown in Equation 6

\[
\ln(GDP_{2000})_t = 17.67 + 1.377 + \ln\left(\frac{GDS}{GDP}\right)_t - 0.623 \ln\left(\frac{DTAX}{GDP}\right)_t + 0.985 \ln\left(\frac{DTAX}{ITAX}\right)_t
\]

After removing the insignificant coefficients from the ARDL(1,3,3,2) model with level variables to get the parsimonious results as suggested by Hendry[24] (General to Specific Approach) except the level variables, the calculated F-statistics(7.979) for the cointegration using the wald test on coefficients of level variables shows stronger results as compare to previous model, This confirms the existence of long-term relationship among variable used.

The coefficients of long run relationship and their statistics is given in Table 3, we generated the error correction term using the normalized coefficients by LGDP2000 given in

Table 4. The error correction term for the long run impact is generated using the Equation 7

\[
ECM_t = \ln(GDP_{2000})_t - 2.218\ln\left(\frac{GDS}{GDP}\right)_{t-1} + 0.032\ln\left(\frac{DTAX}{GDP}\right)_{t-1} - 0.144\ln\left(\frac{DTAX}{ITAX}\right)_{t-1}
\]

With the help of Table 4 we define the error correction term (ECM) and finally found the Equation 7

The final error correction model given in Table 5 shows that in the long run series will converge but at slow rate.

The results for error correction model for growth causality is presented in table 5. All the coefficient are statistically significant. The coefficient of ect (-0.0389) shows low rate of convergence in the long-run.

The Table 6 shows that in the long run all the right hand side variables granger cause real GDP growth and Direct to Total Tax ratio. We can further conclude the following causality relationships in the short run:

| Saving to GDP ratio - Real GDP Growth |
| Saving to GDP ratio - Direct Tax to GDP ratio |
| Saving to GDP ratio - Direct Tax to Total Tax Ratio |
| Direct Tax to GDP ratio - Real GDP Growth |
| Direct to Total Tax Ratio - Real GDP Growth |
| Direct to Total Tax Ratio - Direct Tax to GDP ratio |

Table 1: Unit root tests

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lags</th>
<th>Test statistics (ADF)*</th>
<th>(KPSS)**</th>
<th>Lags</th>
<th>Test statistics (ADF)*</th>
<th>(KPSS)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGDP2000</td>
<td>1</td>
<td>-1.47589</td>
<td>0.15876</td>
<td>0</td>
<td>-0.31225</td>
<td>0.108473</td>
</tr>
<tr>
<td>LSGR</td>
<td>0</td>
<td>-2.84973</td>
<td>0.10535</td>
<td>0</td>
<td>-1.34044</td>
<td>0.075779</td>
</tr>
<tr>
<td>LDTGR</td>
<td>1</td>
<td>-3.83236</td>
<td>0.07056</td>
<td>0</td>
<td>-0.26147</td>
<td>0.089002</td>
</tr>
<tr>
<td>LDT_TFR</td>
<td>8</td>
<td>-5.14271</td>
<td>0.07029</td>
<td>8</td>
<td>-5.04917</td>
<td>0.089412</td>
</tr>
</tbody>
</table>

* Critical Value: -3.587322, Null Hypothesis: Variable has a unit root

** Critical Value: 0.146, Null Hypothesis: Variable is stationary
Table 2: F-statistics for co-integration relationship

<table>
<thead>
<tr>
<th>Value</th>
<th>Lag</th>
<th>Bound Critical Values* (Restricted Intercept with no trend)</th>
<th>Bound Critical Values* (Restricted Intercept with trend)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I(0)</td>
<td>I(1)</td>
</tr>
<tr>
<td>F-statistics</td>
<td>6.51897</td>
<td>1%</td>
<td>4.614</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5%</td>
<td>3.272</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
<td>2.676</td>
</tr>
</tbody>
</table>

* Note: Based on Nayan (2004)

Table 3: The coefficients of long-run relationship

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-statistics</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGDP2000(-1)</td>
<td>-0.038059</td>
<td>-3.540314</td>
<td>0.0017</td>
</tr>
<tr>
<td>LSGR(-1)</td>
<td>0.086413</td>
<td>4.821892</td>
<td>0.0001</td>
</tr>
<tr>
<td>LDTG(-1)</td>
<td>0.001266</td>
<td>-0.66039</td>
<td>0.9479</td>
</tr>
<tr>
<td>LDT_TTR(-1)</td>
<td>0.005614</td>
<td>0.2925</td>
<td>0.7724</td>
</tr>
</tbody>
</table>

Table 4: Normalized Long-run coefficients

<table>
<thead>
<tr>
<th>Dependant LGDP2000</th>
<th>LGDP2000</th>
<th>LSGR</th>
<th>LDTG</th>
<th>LDT_TTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.218</td>
<td>0.0325</td>
<td>-0.1441</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: The error correction model

<table>
<thead>
<tr>
<th>Dependant Variable LGDP2000</th>
<th>Independent Variable</th>
<th>Coefficients (t-statistics)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.796274* (6.401635)</td>
<td></td>
</tr>
<tr>
<td>ECT(1)</td>
<td>-0.038039* (3.906585)</td>
<td></td>
</tr>
<tr>
<td>DLSGR(-1)</td>
<td>-0.053655* (-3.453934)</td>
<td></td>
</tr>
<tr>
<td>DLDTG(-1)</td>
<td>0.031664* (3.142004)</td>
<td></td>
</tr>
<tr>
<td>DLDTG(-2)</td>
<td>0.021327* (2.324044)</td>
<td></td>
</tr>
<tr>
<td>DLDT_TTR(-2)</td>
<td>-0.029345* (-3.09007)</td>
<td></td>
</tr>
</tbody>
</table>

Diagnostics test:

| LM Test | 2.153472 [0.09] |
| Normality (Jarque Bera) | 0.5329 |
| White Heteroskedasticity | 1.999622 [0.133182] |

Notes:

1. t-Statistics in () parenthesis
2. Probability in [] square brackets
3. * significant at 5% level of significance

Table 6: Granger Causality using Pesaran’s approach to ARDL error correction model

<table>
<thead>
<tr>
<th>Dependent</th>
<th>LGDP2000</th>
<th>LSGR</th>
<th>LDTG</th>
<th>DLDT_TTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lags ΔLGDP2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lags ΔLSGR</td>
<td>11.3800 (0.002)</td>
<td>0.348 (0.56)</td>
<td>7.433 (0.0113)</td>
<td>13.309 (0.0006)</td>
</tr>
<tr>
<td>Lags ΔLDTG</td>
<td>5.181 (0.0128)</td>
<td>0.167 (0.8471)</td>
<td>1.689 (0.2043)</td>
<td>1.452 (0.239)</td>
</tr>
<tr>
<td>Lags ΔLDT_TTR</td>
<td>9.545 (0.0047)</td>
<td>0.996 (0.3275)</td>
<td>7.254 (0.0122)</td>
<td>4.705 (0.0653)</td>
</tr>
<tr>
<td>ECT(1)</td>
<td>-0.035 (0.00)</td>
<td>0.097 (0.2246)</td>
<td>0.179 (0.1198)</td>
<td>0.162 (0.0742)</td>
</tr>
</tbody>
</table>

CONCLUSION

The main idea of this paper is to examine the impact causal relation among taxation mix, saving and real economics growth under ARDL system. The results show that saving causes the real GDP growth unidirectional because we did not find the any statistical evidence regarding the causality of saving by GDP. The direct tax to GDP ratio granger causes the real GDP growth significantly; which implies that a higher level of direct tax will foster the real growth. This finding is very important for policy implication since Pakistan have a heavy reliance on indirect taxes.

REFERENCES