Evaluating the Real Determinants of Informal Economy: A Case Study of Pakistan

Ayaz Khan[1], Abrar Khan[2] and Farhatullah[3]

Abstract

One of the most critical issues for developing economies is the ever-increasing size of informal economy. Hence, informal economy and the increase in its size are posing a real economic challenge to developing countries. Pakistan being a developing country is not an exception in this regard where the size of informal economy is expanding which, of course, is a real concern for the country. To tackle and understand the nature of the issue the current study, therefore, attempts to evaluate the determinants of the real size of informal economy from 1972 to 2020 for Pakistan. Real factors of the economy like Employment level, political stability, tax to GDP ratio and cost of working in informal economy are considered in the current research for analysis. Autoregressive Distributed Lagged (ARDL) method is applied to evaluate the informality in short-run and long-run with the help of these real factors. The result shows that in the long run employment level, cost of working and political stability negatively and significantly, while tax variable positively and significantly contribute to informality. In the short run the findings are different as some factors significantly contribute while others do not. Findings of study identify the factors augmenting the size of informal economy in Pakistan. It is suggested to the policy makers in Pakistan that the increasing size of informal economy can be curtailed through political stability, providing more employment opportunities, increasing the cost of working in informal economy and better taxation policy.

Keywords: Developing Country, Informal Economy, Real Factors, Cost of Working, ARDL

JEL Classification: H71, O17

1.   Introduction

The informal economy can be defined as all the economic activities that take place in the premises of a country but are hidden from government agencies. These activities are not included in the official GDP. These transactions are kept

[1]Department of Economics, University of Malakand, Pakistan.
[2]College of Economics and Management, Northeast Forestry University, China
[3]Balochistan University of Information Technology, Engineering and Management Sciences, Pakistan
Corresponding Author's Email: ecoabrarnahk@gmail.com

off the record from the official GDP for the purposes to avoid additional costs. However, the existing literature called such activities in an economy with different names like 'underground, shadow, black, second economy, and so on (Khan and Khalil, 2017). According to Feige (1989) informality in economy happens because people even do legal production and distribution of commodities and services in illegal way in order of tax avoidance and getting benefits through fraud. However, evaluating the size and nature of these activities is an observable issue. Both fields of economics and political literature contain enormous amount of research studies on informal economy. But due to the difficulties in the availability of primary data, national accounts of an economy mostly rely on secondary level data. But still, it is important to have some information about the existence of informality and its nature and size. The informal sector of economy does hurt the formal sector because apparently informal economy is an equal amount that is cutoff from the official or formal economy (Khan and Khalil, 2017). Hence, it can disturb the policy targets of an economy and outcomes of policies.

The Keith Hart (1970, 1973) work is named the pioneered research on the concept of informal economy. In his research, he found two main reasons for initiating informality: namely tax evasion and government regulation. He further said that informal sector labor is self-employed. The structure and implications of tax system determined the motivations to hide economic activities (Tanzi, 1999), other aspects of a country such as the socio-economic and political factors also affect the size of informal economy (Marinov, 2008). However, majority of the developing economies are characterized by a substantial size of informal sector (Chatterjee and Turnovsky, 2018). Therefore, Pakistan, being a developing economy with low per capita income, faces significant challenges in economic progress, especially with low domestic production, exchange rates volatility, foreign debt, altering tax rate policies, negative net exports, high inflation and unemployment, plus now this large size of informal economy (Khan and Khalil, 2017). As Schneider et al. (2010) added that more developed and stable economy means higher formal GDP and less will be the size of informal economy.

Eilat and Zinnes (2000) argued that people involvements in informal means are greater when the judicial system is weak and less efficient. Schneider (2006) points out that while in developed countries, informal sector most often reduces the corruption, however, in developing countries, corruption is increased by the informal sector. A less developed country like Pakistan, therefore, can boost its economic growth if the informal sector of the economy is well controlled by the government (Khan and Khalil, 2017). A report by International Labour

Organization (2018) shows that most of the informal employment exist in the developing economies globally, for instance, in Asia and the Pacific region, this accounts 78% for the countries located in Southern Asia while the countries located in Eastern Asia, this proportion is 49%. Furthermore, Yasmin and Rauf (2003) reported an appreciated increase in informality size about PKRs 12 billion in 1947 to PKRs 1085 billion in 2002. In a recent study by Khuong et al. (2021) showed the size of informal economy, which is 56% of Pakistan's gross domestic product. This means high tax evasion and more fiscal loss to the country which is a great concern for monetary and fiscal departments (Khan and Khalil, 2017).

However, a number of available methods for estimating informality of economy is utilized in case of Pakistan's economy. Such as Kemal (2003) and Kemal and Qasim (2007) used monetary based approach, Yasmin and Rauf (2003) used Ordinary Least Squares (OLS) method to estimate underground economy. Arby et al. (2010) utilized Multiple Indicators and Multiple causes (MIMIC) method and electricity consumption approach. In all the studies, most of variables taken are in monetary terms where actual size of informality estimate is less reliable when estimated through monetary variables. Thus, it is to be noted that diversified results were obtained by these approaches in case of Pakistan's economy and these diversifications make these results less reliable. Moreover, Khan and Khalil (2017) employed the HP Prescott filter and the Feasible Generalized Least Square (FGLS) methods for estimating the size of the informal economy for Pakistan.

Current study extends the findings of the real size of informal economy of Pakistan by Khan and Khalil (2017) and evaluate the least evident scenario of the impact of the determinants on informal economy in case of Pakistan. The selection of factors for evaluating the size of and impact on informal economy in case of Pakistan was carefully done. In addition, the literature is sparse about evaluating informal economy in term of long run and short run responses. The present study in this regard provides fresh insights to evaluate the size and determinants of informal sector in the case of Pakistan by utilizing the Autoregressive Distributed Lagged model to fill the gap. The rest of the study is organized as follows; Section two outlines literature review followed by the proposed methodology for the study in section three. Section four contains results and discussion while the last section concludes the study with some policy recommendations.

2. Literature Review

Both economic and political literature contains a huge amount of research regarding informal activities of an economy around the globe, numerous research studies are conducted to understand the occurrence, nature, size, and dynamics of underground/informal economy not only in the context of developing economies of the world but also in developed economies too. To get guidelines for current study, some relevant studies from existing research literature are reviewed.

2.1. Findings from the World Economies

Evaluating the methods of informal estimates, Andrei and Stancu (2008) studied the economies of Romania and Hungry. They found that same data set of a country under different methods of estimation gives different informal economy estimates as a percent of official GDP. Marinov (2008) found for economy of Bulgaria that tax system, government regulations, the political and socio-economic factors like public sector service providing quality and legislative system quality cause variations in informal economy size. Higher unemployment rate in an economy will force the labor class to go informal (Alm and Yunus, 2009). Kaufmann et al. (2009) studied multi political economies and found that good governance, violence, peace and political stability along with government regulation determined the size of informality in economy. Dell'Anno and Schneider (2006) performed a study on 21 OECD countries containing United States, Austria, Italy and Greece etc. using MIMIC method. They found that USA was at the minimum of about 9 percent and Greece was at maximum of about 28 percent of informality of total economy. Austria was at 10.8 percent. Schneider and Klinglmair (2004) applied MIMIC estimation method of informal economy on Brazil. They found about 39.8 percent of informality of the official economy (GDP) in 2000. Giles et al. (2002) applied Granger's causality to check the causal relationship among informal and formal GDP in Canada. Causality from formal to informal GDP was found significant while the reverse was not true for Canadian economy. Dell'Anno et al. (2004) studied Spain, France and Greece for informality. Findings show the existence of informality (informal economy) in all selected economies. Tax burden was prominent in France and Spain in determining the informal economy while Greece shows no such evidence. Political instability was also found as a cause of informal economic activities in these countries. Etim and Daramola (2020) explored that factors that foster the informal economy by using a comparative systematic review approach for Nigeria and South Africa. The data collected from 31 primary studies showed that a number of factors, including income disparity among citizens, unemployment,

excessive tax burdens, poor corruption control, excessive bureaucratic hurdles from government, lack of social protection survival tendencies, low GDP per capita, and inflationary tendencies drive informal economy in both Nigeria and South Africa. In the context of China's economy, Huang et al. (2020) noted that the urban informal employment is high, accounting for 22% of total output of the country. They further added that the key reasons for this urban informal economy in the country are unemployment, globalization of urban economies, the stages of economic development, and rural-to-urban migration.

2.2.   Findings from Pakistan's Economy

Shabsigh (1995) used monetary based approach for estimation of informal economy for Pakistan. The author found that a reduction in size of informality occurred between 1975 to 1990. In 1975 the size of informality was about 21 percent of the official economy and it decrease to 20 percent in 1990. The author used taxes on exports and imports instead of overall taxes collected within the economy. In contrast, Aslam (1998) performed a study and found an enormous increase in underground economy size of about 75 percent in three decades; 1960-1990, and the main reason according to author was increasing tax burden. In 1960 the informality was about 29 percent of official economy while it was about 44 percent in 1990. Yasmin and Rauf (2003) used simple Ordinary Least Squares (OLS) and currency demand methods based on monetary approach to estimate the size informal economy. They found an appreciated in informality size. The informality increases from about PKRs 12 billion in 1947 to PKRs 1085 billion in 2002. Regional disparities, unemployment and tax burden were the main causes according to the study. The study also suggest that this enormous increase shrinks the official GDP, adversely affect the tax revenue and increase in socio-economic problems but transparent and regular tax auditing can correct the situation. Kemal and Qasim (2012) used Discrepancy method (KQ method) and estimated informal economy. About 91 percent of informality was found 2007-08. The study also discussed major weakness in different methods available and used for estimating informality of an economy. Using electricity MIMIC models and consumption approach, Arby et al. (2010) concluded that informal economy accounts for 30% of Pakistan's total output. By calculating the estimated GDP through Feasible Generalized Least Square (FGLS) over a period of 1972 to 2010, Khan and Khalil (2017) utilized HP-Prescott filter method to obtain the series of actual GDP and potential GDP from the estimated GDP to find the informal GDP for Pakistan. The study found that informal economy accounts for 71% of the economy of Pakistan. According to Kemal and Qasim (2012), the informal economy accounts between 74% and 91% of the reported formal economy in 2008. Using currency

demand equation, Khuong et al. (2021) computed the impact of informal economy on economic growth in Pakistan. The result indicated the existence of 56% informal economy in Pakistan as of gross domestic product.

3.   Methodology and Data

3.1.   Model Specification

Feasible Generalized Least Squares (FGLS) and HP Prescott Filter [4] method is used to estimate the real size of informal GDP[5] (informal economy). While setting this informal GDP (time series) as dependent variable and the real factors as explanatory factors, this study utilized ARDL econometric technique. The reasons are: under consideration in an unrestricted equilibrium correction model, this study in ARDL procedure of F-statistic used in a Generalized Dickey-Fuller type regression for testing the variables lagged level significance. This method has a number of advantages over its alternatives such as the reduced-rank system-based regression approach of Johansen (1988, 1995) and Engle and Granger (1987) two step (error term-based) process for testing the null hypothesis of no co-integration (long-run relationship). First, compared to the Engle-Granger method, the UECM statistical procedure and properties are better since it does not push the short-run dynamics into the residual terms as Engle-Granger method pushed it (Banerjee et al., 1993). Second, the results of bound test procedure stay robust and are non-sensitive to the order of integration of explanatory variables. In other words, whether the explanatory variables taken are purely stationary, non-stationary or mutually co-integrated, the bound test results still remain robust. Another advantage of this bound test procedure is that it is applicable for both small and large sample sizes, unlike Johansan and Engle-Granger procedures of co-integration, which are not reliable in small sample size.

Pesaran and Shin (1999) introduced ARDL bound test approach to investigate the short-run and long-run dynamics of al real factors to informal GDP and its impact on the economy's informal sector. This study also uses this approach. The ARDL method has the advantage of keeping both the short run and the long run relationship along with the adjustment process in a single equation. For long run trends of informal economy basic model is given as:

$$GDP_{inf} = \beta_0 + \alpha_1 TGDP + \alpha_2 EMPL + \alpha_3 PST + \alpha_4 CO + \varepsilon_t \qquad (1)$$

---

[4] Details of both method FGLS and HP filter can be found in Khan, A., and Khalil, S. (2017). The real size of underground economy: A case of Pakistan. *Pakistan Journal of Applied Economics*, *27*(1).
[5] The complete set of GDP$_{inf}$ (real size of informality) time series with comparison is given in appendix of the study.

Where the term GDP_{inf} refers to the gross domestic product (GDP) of informal economy, TGDP stands for tax to GDP ratio, EMPL is defined as total employed labor force in the economy, PST is a dummy variable that measures of political stability of the country (a stable government is labeled with value equal to one (1) and zero (0) otherwise) while CO is the cost of working in informal economy which defined as the punishment through government legislation. ε_t is error term. Furthermore, a vector Zt is defined for applying ARDL to obtain robust long-run relationship, which consists of five set of variables as Zt = (GDP_{inf}, TGDP, CO, EMPL, PST) where GDP_{inf} is explained variable and TGDP, CO, EMPL and PST are explanatory variables (real factors). The Zt follows a p-order vector auto-regression for data generating process. The main steps in ARDL modeling are as follows.

$$\Delta(GDP_{inf})_t = \alpha_0 + \theta_1(TGDP)_{t-1} + \theta_2(GDP_{\text{inf}})_{t-1} + \theta_3(EMPL)_{t-1} + \theta_4(PST)_t + \theta_5(CO)_{t-1} + \sum_{i=0}^{n} \beta_i \Delta(TGDP)_{t-i} + \sum_{i=1}^{n} \sigma_i \Delta(GDP_{inf})_{t-i} + \sum_{i=0}^{n} \delta_i \Delta(EMPL)_{t-i} + \sum_{i=0}^{n} \gamma_i \Delta(CO)_{t-i} + \varepsilon_t \tag{2}$$

In the above model α_0 is the drift term and θ_i are the long run coefficients where i= 1, ……, 5, the β_i, σ_i, δ_i, and γ_i are coefficients of the short run of the model and ε_t is the error term and $\varepsilon_t \sim N(0, \Omega)$. PST is a dummy variable that is why there is neither lag nor difference forming the short and long run parts of the Equation 2 respectively. However, absence and existence of the long-run relationship among GDP_{inf} and real factors (TGDP, CO, EMPL, PST), is checked through dropping the lagged level variables $(TGDP_{t-1}, GDP_{\text{inf}_{t-1}}, EMPL_{t-1}, PST_t, CO_{t-1})$ from Equation 2. For existence and absence of long-run relationship hypothesis are formulated as.

Null hypothesis (no co-integration):

$H_1: \sum_{i-1}^{N} \theta_i = 0$

Alternative hypothesis (co-integration):

$H_1: \sum_{i-1}^{N} \theta_i \neq 0$

F-test is used for checking the evidence of long-run relationships. The F-test non-standard distribution rests on (i) whether the model contains trend or/and an intercept term, (ii) the variables taken in ARDL model are I(0) or I(1) and (iii) the number of regressors or real factors are taken. The bound test asymptotic critical value depends on the nature of variables, whether they are I(0) or I(1) or mutually co-integrated. For this purpose two sets of asymptotic values are generated I(0) named as lower bound critical values and I(1) as upper bound

critical values. Now to find out the optimal lag length for each variable, the Autoregressive distributed lag (ARDL) model estimates $(p+1)^k$ number of regressions. Here, p denoted the maximum number of lags to be used while k means the number of variables in equation.

The existence of co-integration is checked by the F-statistic value, when the estimated F-statistic value is greater than respective upper critical value then null hypothesis of no co-integration is rejected while alternate hypothesis is accepted. If an F-statistic value is less than the lower critical values, then we do not reject null hypothesis in favor of alternate hypothesis and when the calculated F-test value falls in between the values of upper and lower boundaries we cannot reject or accept both null and alternate hypothesis. Establishing the long-run relationship next stage is to find out the ARDL lag order for the model by noting the changing values of either AIC (Akaike Information Criteria) or SBC (Schwartz Bayesian Criterion) and then the selected model is estimated through Ordinary Least Squares (OLS) method. The mathematical approach for ARDL model can be found in Pesran and Pesran (1997) and Pesaran et al. (2001). After finding co-integration among the variables in the model (2) the following long-run model (3) will be estimated

$$GDP_{inf_t} = \alpha_0 + \sum_{i=0}^{n} \beta_i\, TGDP_{t-i} + \sum_{i=1}^{n} \theta_i\, (GDP)_{inf_{t-i}} + \sum_{i=0}^{n} \delta_i\, (EMPL)_{t-i} + \sum_{i=0}^{n} \gamma_i\, (CO)_{t-i} + \lambda_i (PST)_t + \varepsilon_t \qquad (3)$$

For the long run adjustment and short run dynamics the following model will be estimated.

$$\Delta(GDP_{inf})_t = \sum_{i=0}^{n} \beta_i\, \Delta(TGDP)_{t-i} + \sum_{i=1}^{n} \theta_i\, \Delta(GDP_{inf})_{t-i} + \sum_{i=0}^{n} \delta_i\, \Delta(EMPL)_{t-i} + \sum_{i=0}^{n} \gamma_i\, \Delta(CO)_{t-i} + \alpha_4 PST_t + \theta(ECM)_{t-1} + \varepsilon_t \qquad (4)$$

Equation 4 estimates the short run response of real factors and the error correction term $(\theta(ECM)_{t-1})$ denotes long run equilibrium adjustment, where the value of θ is theoretically said to be negative which favor the speed of adjustment toward equilibrium. If the sign of θ is positive, it shows divergence from long run equilibrium.

3.2. Data

The study aims to evaluate the size of unofficial economy in Pakistan over a period extending from 1972 to 2020. For this purpose, the data on variables of interest, i.e. gross domestic product (GDP), cost of working in informal economy (CO), employment labor force (EMPL)) is collected from Economic Survey of Pakistan while data on Tax to GDP Ratio (TGDP) is taken from Statistical Year

Book of the Federal Bureau of Statistics. Political Stability (PST) is used as a dummy variable, calculated by authors. A legislative government which complete its tenure is named political stability and given the value of 1 and 0 otherwise.

4.   Empirical Results and Discussion

The given Table 1 contains information on descriptive statistics of the variables. The table shows that on average GDP is 1665 while on average, TGDP is 0.131. Similarly, on average CO is 2844 where the maximum value is 6178 and minimum values 3922. On average, EMPL is 31242809 and PST is 0.641. The table also shows that all variables are normally distributed except the TGDP and PST. Moreover, variables TGDP, PST and EMPL are skewed toward their respective maximum values. Standard Deviation among all variables is high for EMPL which shows high volatility and dispersion toward GDP. As analyzing time series data required so prerequisite one of which is checking the integration order of the variables taken. For these purposes two well-known tests of stationarity Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) were applied. The results in Table 2 show that none of the variables is stationary at corresponding percent level of significance, respectively, with intercept and trend and intercept. However, TGGP is stationary at intercept only at 5 percent level of significance in Phillips-Perron test. The rest of variables are non-stationary at level. To avoid over estimation in the results, all variables become stationary after taking the first difference at 5% significance level, respectively, with intercept and trend and intercept.

Table 1. Descriptive Statistics of the Variables

Statistics\Variables	Mean	Max	Min	Std. Dev.	Skew	Kurt	JB
GDP	1665	3719	3985	9960	0.644	2.421	3.239
TGDP	0.131	0.246	0.097	0.027	2.763	11.81	176*
CO	2844	6178	3922	1545	0.236	2.557	0.679
EMPL	3124	3304	2947	1144437	0.611	2.405	3.001
PST	0.641	1.000	0.000	0.486	-0.58	1.346	6.69*

Source: Authors' calculations (* shows the level of significance at 5%).

After finding the order of integration of all factors, on Equation 2 two step ARDL estimation method is applied for estimation of long-run relationship in the basic model. In first step appropriate lag length (m) is found out for the model. Bahmani and Bohl (2000) found the sensitivity of long run estimated to appropriate lag length. An Akaike information criterion (AIC) is used in current study. In the second step the existence of long-run relationship is checked. According to AIC results in Table 3, the optimum lag to be used is 2 and in Equation 5 ARDL estimates $(2+1)5 = 243$ total number of regression. However, the given value of F-statistic in Table 3 is compared with Pesaran et al. (2001)

table value. The critical bond value at 5% significance level for F-statistic is 2.62 to 3.79. * Indicates that computed F-statistic falls above the upper bond critical value.

Table 2. Results of Unit Root Analysis

	Augmented Dickey-Fuller Test for Stationarity			
	Level		1st Difference	
Variables	With intercept	Intercept and trend	With intercept	Intercept and trend
GDP$_{inf}$	-1.034767	-2.921449	-5.881147*	-5.948652*
TGDP	0.223041	0.433687	-4.628415*	-4.995787*
CO	-0.080424	-2.428624	-4.449628*	-7.274144*
EMPL	2.285079	-0.123404	-5.805118*	-6.687759*
PST	-2.142569	-2.066400	-7.637626*	-7.647106*
	The Phillips-Perron (PP) Test for Stationarity			
	Level	1st Difference	Level	1st Difference
Variables	With intercept	Intercept and trend	With intercept	Intercept and trend
GDP$_{inf}$	-0.605762	-2.569814	-7.783222*	-11.52192*
TGDP	-0.478358	0.024905*	-4.635815*	-4.998556*
CO	1.551195	-2.150883	-7.588602*	-11.73165*
EMPL	2.765715	-0.123404	-0.5838509*	-6.960427*
PST	-2.045031	-1.939198	-7.652586*	-7.810842*

Source: Authors' calculations (*, ** shows the level of significance at 1% and 5% respectively).

Table 3. Results of AIC and Bound test with its critical values for existence of co-integration

Akaike Information Criteria (AIC)	Bound test and its critical values for existence of co-integration		
		F-test Statistics value = 3.82*	
26.11(0)	Level of significance	Lower bond I(0)	Lower bond I(0)
24.61(1)	1% level	3.41	3.41
24.46(2)	5% level	2.62	2.62
24.66(3)	10% level	2.26	2.26

Source: Authors' calculations (The values of AIC are obtained by using EViews through Unrestricted VAR (Vector Auto-regressive) procedure).

Estimating Equation 2 on selected lag length and applying Wald test for coefficient restrictions the following F-statistic value 3.82 is obtained which is larger than the upper bond value 3.79 and result that null hypothesis of no co-integration is not accepted at 5% level of significance. In other words, in Equation 2 long-run relationships do exist. Table 4 gives the long-run coefficient of selected factors. The sign of coefficients is according to the expectations. The most significant variable is CO (cost of working in informal economy), and the least significant variable is that of TGDP (Tax to GDP ratio).

The selected factors do contribute to both formal and informal economy. All the factors related negatively with informal GDP except that of Tax to GDP ratio (TGDP). The in-depth explanation of each long-run factor is given in conclusion of the study. In the short run results given in Table 5, the Error

Correction Mechanism (ECM) term is highly significant with negative sign. The significance of ECM also supports the existence of long-run relationship between the factors taken in Equation 1.

Table 4. Results of Long-Run Response Model

Regressors		Coefficient	
C		-160166.1	
CO		-10.5824*	
EMPL		-0.0405*	
TGDP		439543.06***	
PST		-23372.54***	
Goodness of Fit Statistics			
R- squared	0.91	AIC	24.12
Adjusted R2	0.89	SIC	24.56
F-statistic	33.5	Durbin-Watson	2.29

Source: Authors' calculations (*, ** & *** represent level of significance at 1%, 5% and 10% respectively & ARDL Technique order (2,1,1,1,0); Dependent variable GDP_{inf}).

Table 5. Results of Short Run Disequilibrium Response Model

Regressors		Coefficients	
C		-73818.26	
ΔCO		-10.67*	
ΔEMPL		-0.01**	
PST		-202.14	
ΔTGDP		-109434.1	
ECM(-1)		-0.08*	
Goodness of Fit Statistics			
R-squared	0.47	SIC	24.84
Adjusted R2	0.36	Durbin-Watson	1.73
AIC	24.53		

Source: Authors' calculations (*, ** & *** represent level of significance at 1%, 5% and 10% respectively and ARDL Technique Order (2,1,1,0,1); Dependent variable ΔGDP_{inf}).

The value ECM term says that about 8 percent of correction in disequilibrium toward equilibrium is done in the current year as the data taken is annual. Further the short-run coefficient of CO is negatively significant and same is that of EMPL coefficient while both TGDP and PST are insignificant in short-run.

Table 6. Results of Sensitivity Analysis

Short run Diagnostic Test	
Serial Correlation LM Test	0.4995(0.612125)
ARCH Test	0.3445(0.841734)
White Heto Test	9.4075(0.584330)
Ramsey RESET	0.8444(0.481602)
Jarque-Bera	2.2358(0.326963)

Source: Authors' calculations (Where p-values are in the parenthesis)

The coefficient stability and robustness are one of important determinants of reliable results and prediction. For this purpose, sensitivity analysis of the factors in the model is done through different tests available. The tests label in Table 6 gives satisfactory results. The evidence of no heteroskedasticity is confirmed by LM (Lagrange Multiplier) test results. The value of Ramsey RESET test favors that there is no miss-specification in the model and the probability value of Jarque-Bera (JB) test evidence that the error term of the model is normally distributed.

5.   Conclusion and Policy Recommendations

Today, the increasing size of informal economy is a vital concern for economies, particularly for developing economies like Pakistan. In efforts to reduce informal economic practices, countries have adopted many preventative strategies for reducing the volume of informal economic activities. Therefore, the current study selects some real factors like tax to GDP ratio, Employment level in economy, cost of working in informal economy, and Political Stability to examine the response of these factors toward the size of informal economy over a period of 1972 to 2020. Hence, the theme of current study is different because for the first-time real factors are used in evaluating the size and dimensions of informal economy by applying the Bound test procedure of Auto-regressive Distributed Lagged Model (ARDL).

The findings of the study revealed that in the short run analysis, the ECM term is significant and shows convergence toward the equilibrium. The Employment (EMPL) and Cost of working in informal economy (CO) significantly affect the informal economic activities while both Tax to GDP Ratio (TGDP) and Political Stability (PST) are insignificant in the short run. In the long-run analysis, the sign of coefficients of the selected variables are according to the expectation. The most significant variable is CO (cost of working in informal economy), and the least significant variable is TGDP (Tax to GDP ratio). The selected factors do contribute to both formal and informal economy. The CO which presents the legislative system of country, and its short-run significance reveals the fact that daily basis working of legislative system. Mean more of peoples are punished regularly and quickly on involving in informal means less will be the size of informal economy and more sustained increase in the official

GDP. The real factor cost of working in informal economy (CO) further explain that when government legislative system like courts, law and order became efficient so informal economy size can be controlled, As the number of punishments for illegal activities (both economic and social) increases greater number of people will come towards the formal activities in the economy see (Friedman et.al (1999); Risteski (2009). Previous studies also evidenced that tax is one of the key determinants of informal activities, whether an individual or an organization goes informal due to tax avoidance. The Tax to GDP ratio (TGDP) shows how much of tax is paid by individuals in the economy. Avoidance of tax starts when the government takes steps to increase the tax rate because tax is a real cost on individuals in the economy. The increase in tax rate leads to tax avoidance (evasion) which leads to an increase in informal activities and shrinks the formal sector of economy see (Schneider (2006); Iqbal et al. (1999); Anderberg et al. 2003). Full employment is the main objective of policy makers. The negative sign of Employment (EMPL) says that whenever the government takes steps to increase employment in the formal economy the involvement of people in informal activities will decrease and hence the formal economy can grow see (Startiene and Trimonis (2010); Kolm and Larsen 2003). The sign of Political Stability (PST) is negative and consistent. Stability of government especially in developing country like Pakistan leads to stable governance and smooth policies (Eilet, Y. and Zinnes, C. 2002; Anoop et al., 2012).

Hence, this study recommends that the real factors under consideration must be properly targeted in combating the existence of informality in system. Policy makers must scrutinize the tax structure carefully and the burden of tax should be equalized. The tax structure must be made more diversified so that the tax base (tax network) enlarges and equality in tax collection can be brought. The corruption in tax collections and tax auditing should be controlled. Income tax needs to be progressive type and consumption tax rate needs to be set at a rate that does not hurt the labor class, so that shipment of labor to informal means is controlled. The legislative system needs proper attention, for smooth working and the system must be made free from any political pressure plus bribing and threats must be controlled. The employment level in the economy also determined the

informal means. The employment process should be free of corruption and new job opportunities must be created.

References

Alm, J. & Yunus, M. (2009). Spatiality and persistence in US individual income tax compliance. *National Tax Journal*, 62(1), 101-124.

Anderberg, D. (2000). A dynamic model of the black economy. *Public finance analysis,* 61(4), 447-461

Andrei, T. & Stancu, S. (2008). Econometric methods used to study the informal economy and regional development. *Informatica Economica* Bulletin, 3, 120-128.

Arby, M. F. Malik, M. J. & Hanif, M. N. (2010). The size of informal economy in Pakistan. *State Bank of Pakistan Working Paper Series, 33,* 1-20.

Aslam, S. (1998). The underground economy and tax evasion in Pakistan: Annual estimates (1960-1998) and the impact of dollarisation of the economy. *The Pakistan Development Review*, 37(4), 621-631.

Bahmani, O. M. & Bohl, M. T. (2000). German monetary unification and the stability of long-run German money demand function. *Economic Letters*, 66(2), 203-208.

Banerjee, A. Dolado, J. J. Galbraith, J. W. & Hendry, D. (1993). Co-integration, error correction, and the econometric analysis of non-stationary data. *Oxford University Press*.

Chatterjee, S. & Turnovsky, S. J. (2018). Remittances and the informal economy. *Journal of Development Economics*, 133, 66-83.

Dell'Anno, R.. Gómez-Antonio, M. & Pardo, A. (2004). The shadow economy in three Mediterranean countries: France, Spain and Greece. A MIMIC approach. *Empirical Economics*, 33, 51-84.

Dell'Anno, R. & Schneider, F. G. (2006). Estimating the underground economy by using MIMIC models: A Response to T. Breusch's critique. *Johannes Kepler University of Linz Working Paper, 0607.*1-30.

Eilat, Y. & Zinnes, C. (2000). The evolution of the shadow economy in transition countries: consequences for economic growth and donor assistance. *Harvard Institute for International Development, CAER II Discussion Paper*, 83, 1-70.

Engle, R. F. & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*, 55(2), 251-276.

Etim, E. & Daramola, O. (2020). The informal sector and economic growth of South Africa and Nigeria: A comparative systematic review. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 134.

Eilet, Y. & Zinnes, C. (2002).The Shadow Economy in Transition Countries: Friend or Foe? A Policy Perspective. *World Development,* 30(7), 1233-1254.

Feige, E. L. (1989). The meaning and measurement of the underground economy. *Feige, EL (1989a)*, 13-56.

Friedman, E., Johnson, S., Kaufmann, D., & Zoido-Lobaton, P. (2000). Dodging the grabbing hand: the determinants of unofficial activity in 69 countries. *Journal of Public Economics*, 76(3), 459-493.

Giles, D. E. A. Tedds, L. M, & Werkneh, G. (2002). The Canadian shadow and measured economies. *Applied Economics*, 34, 2347-2352.

Gulzar, A. Junaid, N. & Haider, A. (2010). What is hidden, in the hidden economy of Pakistan? Size, causes, issues, and implications. *The Pakistan Development Review*, 49(4), 665-704.

Hart, K. (1970). 8. Small-scale entrepreneurs in Ghana and development planning. *The Journal of Development Studies*, 6(4), 104-120.

Hart, K. (1973). Informal income opportunities and urban employment in Ghana. *The Journal of Modern African Studies*, 11(1), 61-89.

Huang, G., Xue, D., & Wang, B. (2020). Integrating theories on informal economies: An examination of causes of urban informal economies in China. *Sustainability*, 12(7), 1-24.

Iqbal, Z., Qureshi, S. K., & Mahmood, R. (1998). The underground economy and tax evasion in Pakistan: A fresh assessment. *Reserch Report Pakistan Institute of Development Economics*, 158, 1-14.

International Labour Organization. (2018). *Women and men in the informal economy: A statistical picture* (3rd edition). ILO, Geneva.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3), 231-254.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press.

Kaufmann, D. Kraay, A. & Mastruzzi, M. (2009). Governance matters VIII, aggregate and individual governance indicators, 1996–2008. *World Bank Policy Research Working Paper, 4978,* 1-24.

Kemal, M. A. (2003). Underground economy and tax evasion in Pakistan: A critical evaluation. *Working Papers & Research Reports*, 184, 1-20.

Kemal, M. A. & Qasim. (2007). Fresh assessment of the underground economy and tax evasion in Pakistan: causes, consequences, and linkages with the formal economy. *PIDE-Working Papers*, 13, 1-19.

Kemal, M. A., & Qasim, A. W. (2012). Precise estimates of the unrecorded economy. *The Pakistan Development Review*, 51(4), 505-516.

Ayaz, K. H. A. N., & Khalil, S. (2017). The real size of underground economy: A case of Pakistan. *Pakistan Journal of Applied Economics*, 27(1), 89-100.

Khuong, N. V. Shabbir, M. S. Sial, M. S. & Khanh, T. H. T. (2021). Does informal economy impede economic growth? Evidence from an emerging economy. *Journal of Sustainable Finance & Investment*, 11(2), 103-122.

Kolm, A. S.& Larsen, B. (2003). Wages, unemployment, and the underground economy, *CE sinfo Working Paper* 1086,1-24.

Marinov, A. (2008). Hidden economy in the rural regions of Bulgaria. *International Review on Public and Nonprofit Marketing*, 5, 71-80.

Pesaran, M. H. Shin, Y. & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics*, 16(3), 289-326.

Schneider, F. (2006). Shadow Economies and Corruption All Over the World: What Do We Really Know? *The Institute for the Study of Labor Discussion Paper,* 2315, 1-23.

Schneider, F. Buehn, A. & Montenegro, C. E. (2010). New estimates for the shadow economies all over the world. *International Economic Journal*, 24(4), 443-461.

Schneider, F., & Klinglmair, R. (2004). Shadow economies around the world: what do we know? *Johannes Kepler University of Linz, Working Paper,* 0403, 1-24.

Schneider, F. (2006) Shadow economies and corruption all over the world: what do we really know? *The Institute for the Study of Labor Discussion* Paper, 2315, 1-26.

Shabsigh, M. G. (1995). The underground economy: Estimation, and economic and policy implications: the case of Pakistan. *International Monetary Fund, Working Papers, 1995/101,* 1-24.

Singh, M. A., Jain-Chandra, M. S., & Mohommad, M. A. (2012). *Inclusive growth, institutions, and the underground economy.* International Monetary Fund.

Startienė, G. & Trimonis, K. (2010). Causes and consequences of non-observed economy. *Economic and Management*. 15, 275-80.

Tanzi, V. (1999). Uses and abuses of estimates of the underground economy. *The Economic Journal*, 109(456), 338-347.

Yasmin, B. & Rauf, H. (2003). Measuring the underground economy and its impact on the economy of Pakistan. *The Lahore Journal of Economics*, 9(2), 93-103.

Yiar, G. (2000). Reforming motivation: how the structure of instruction affects students' learning experiences. *British-Educational-Research-Journal,* 26(2), 191-210.

Appendix A

Table 1. Estimated Informal Economy through Feasible Generalized Least Squares (FGLS) and HP Prescott Filter method using Real Factors approach (values are in percentage)

Year	Current Study	Kemal and Qasim (2012)	Kemal and Qasim (2007)	Arby et al. (2010)	Gulzar et al. (2010)	
	Real Factors	Discrepancy Method	Monetary	ARDL	Electricity Consumption	MIMIC
1974	56.67	38	24.01	27.1	30.7	31.6
1975	74.36	33.1	22.18	25.9	38.3	32
1976	58.22	31.6	24.03	28.4	43.4	32.5
1977	65.52	30.9	23.69	27.9	46.3	32.1
1978	51.54	34.9	28.11	29.2	54.8	32.1
1979	68.94	39.2	30.95	31.1	56.5	31.8
1980	78.46	45.6	33.47	33.3	50.1	31.3
1981	53.74	43	31.6	33.1	47.8	31.2
1982	78.84	47.8	38.95	31.6	51.5	31.4
1983	53.01	42	38.71	32.8	56.9	31.3
1984	71.14	49.3	38.22	32.1	53	31.1
1985	55.86	39.3	35.77	29.6	57.1	31.1
1986	51.92	44.7	36.85	35.2	62.2	31.2
1987	67.54	50.5	36.22	35.4	57.7	31.1
1988	51.42	45.5	35.47	32.7	52.5	30.9
1989	60.53	42.7	37.26	32.5	51.4	30.9
1990	36.34	39.2	39.15	30	55.5	30.8
1991	32.28	36.1	33.73	26.1	46.7	30.2
1992	37.59	44.4	37.35	27.7	46.5	30
1993	72.02	45.5	34.93	30.1	56.7	30
1994	49.16	56.6	33.97	33.3	44.1	29.5
1995	42.93	60.6	38.65	34.8	43.4	29
1996	56.73	68.7	41.64	36.8	51	29
1997	34.89	74.9	35.24	36.4	47.6	28.7
1998	79.81	69	33.23	36.4	54.1	28.8
1999	45.69	46.1	32.01	35.2	49.7	28.7
2000	50.43	56.5	33.78	26	58.4	28.6
2001	42.85	65.7	34.07	26.3	56.6	28.4
2002	49.98	64.3	33.23	27	61	28.1
2003	61.34	68.2	35.65	29	55.3	28.5
2004	44.38	66.6	35.45	24.9	50.8	28.1
2005	45.67	64.8	35.17	18.7	49.6	28.1
2006	41.82	…	35.56	18.3	50.1	28.6
2007	23.73		38.03	18.9	51	28.6
2008	57.84	74.3	37.27	19.6	36.1	27.6
2009	68.63	…	32.81	…	37.2	25.9
2010	70.79	91	31.8	…	47.6	26.6
2011	78.2	…	…	…	…	…
2012	85.79	…	…	…	…	…
2013	87.59	…	…	…	…	…
2014	89.3	…	…	…	…	…
2015	84.44	…	…	…	…	…
2016	88.09	…	…	…	…	…
2017	68.35	…	…	…	…	…
2018	62.26	…	…	…	…	…
2019	60.72	…	…	…	…	…
2020	69.07	…	…	…	…	…