ISSN (online) 2664-9535 (print) 2664-9527 https://atcr.kra.go.ke

Forecasting Tax Revenues, Frequency of Observation Matter. A case Study of Tanzania

Masoud Mohammed Albimana [1]
Issa Moh'd Hemedb[2]

[1] Institute of Tax Administration (ITA), Dar-es-Salaam, Tanzania
[2] Zanzibar University, Zanzibar, Tanzania

Received 12 August 2022
Accepted for publication 12 August 2022
Published 11 August 2022

Abstract

This paper intends to examine whether using higher frequency data has more power in forecasting than low frequency data. The sample size ranges from 1996 to 2016 and 2000 to 2015. Ordinary Least Square (OLS) method was used to forecast three components of tax revenues including total revenue (TR), Pay As You Earn (PAYE) and Value-added Tax (VAT). The results show that, both TR and PAYE forecasts are slightly better when using low frequency data. However, for VAT, forecasting power is slightly better when using higher frequency data. Also, the nature of the tax can have different implications in selection of data frequency.
Keywords: Forecasting, Tax Revenue

1. Background

Forecasting models using time series have frequently been used in financial markets and economic growth models, but it has not been well-articulated within tax revenue studies. The selection of sample size is a key component in forecasting and should carefully be conducted prior to analysis, but this aspect has been very often overlooked in tax studies. Theoretical studies of time series insisted on using large sample size which would capture a "long run" period in order to ensure accuracy of estimation and forecasting (Hakkio and Rush, 1991; Hawley et al, 2019). However, Hakkio and Rush (1991) failed to have consensus to answer this question "How long is a 'long run'?'. They did add that, the length of the 'long run' may differ depending on the field. For some academic fields, the long run can be a decade while for others it can be a month (Hawley et al, 2019) . It might be true that in some fields, adding some independent variables increase additional observations on long-run hence shorter sample size might be acceptable.

In the empirical literature, researchers often face limitations of using relatively short span of data due to lack of longer span data. On one front, too long time series data leads to structural breaks that we saw in exchange rate system for African countries and trade openness, which all emerged effectively in the early 1980s. On the other, estimation using short time series are subjected to different claims from different studies (e.g.; DeCarlo& Tryon, 1993; Huitema & McKean, 1991, 1994; DeCarlo & Tryon, 1993; Solanas et al., 2010; Krone et al., 2017)

The minimum number of sample size recommended in time series forecasting differs. However, a considerable consensus ranges from 30 to 50 observations (Hakkio and Rush, 1991; Poole et al., 2002; McCleary et al., 1980; Warner, 1998). The general conclusion from some literature is that, the quality of estimation coefficients increases with an increasing number of sample size.

In turn, some researchers tend to choose higher frequency data for forecast purposes (Su Zhou, 2001; Hakkio and Rush, 1991; Lahiri and Mamingi, 1995; Choi and Chung, 1995; Ng, 1995). Su Zhou (2001) suggested that, using fixed sample size of 20 to 50 years, moving from low frequency to higher frequency data, may either double or even triple the power of the tests. The validity of this suggestion is still doubtful as it was based on small annual data (See examples; Bahmani-Oskooee, 1996; Masih and Masih, 1996; Taylor, 1995).

The empirical testing of previous discussions is still limited especially regarding tax and fiscal studies. Henceforth, the present paper is motivated to examine whether power of forecasting when higher frequency (quarterly) data and low frequency (annual) data were used. This objective is driven by debates in previous literature including Su Zhou (2001), Hakkio and Rush (1991), Lahiri and Mamingi (1995), Choi and Chung (1995) and Ng and perron (1995).

2. Literature Review

2.1 Discussion of high and low frequency data.

Previous studies by Shiller and Perron (1985), Perron (1989) and Hakkio and Rush (1991) have shown that when data are sampled at discrete points, increasing frequency of observations while data span is fixed does not increase the power of unit root. However, their suggestion has been occasionally misinterpreted due to their support for using small annual data (Bahmani-Oskooee, 1996; Masih and Masih, 1996; and Taylor, 1995).

Bahmani-Oskooee (1996, p. 481) supported Hakkio and Rush's (1991, p. 572) suggestion which, testing ten annual observations is no difference from 120 monthly observations. He added that, using annual data of 30 years is as good as using quarterly and monthly data over the same period. In contrast, Choi and Chung (1995) found that increasing frequency of data improves power of ADF test but not the PP tests. In addition, Ng and Perron (1995) suggested that varying data span increases power of estimating. Hooker (1993) also suggested that, temporal disaggregation increases the power of ADF cointegration test. Lahiri and Mamingi (1995), on the other hand, argued that when data span varies, the sample length matter more than the number of observations.

Otero and Smith (2000), after using long-term and short-term interests, concluded that, cointegration depends more on the sample length than the number of observations. In his simulation study, Su Zhou (2001) showed that when studies use short sample size of annual data between 30 and 50 years, higher frequency data yield better forecast power and create less distortion. They illustrated that, using fixed sample size of 20 to 50 years, higher frequency data (quarterly or monthly) has the ability to double or even triple the power of forecasting compared to lower frequency data (annual).

Recently studies also analysed performance of different methods and sample size in tax revenue forecasting. For example, Sabaj and Kahveci (2018) examined the forecasting of tax revenues for Albania economy using the annual data from starts from 2005q1 – 2016q1 with a minimum of 55 observations for each models. The selected methods were better with lower errors than official forecasts. Molapo et al (2019) used Bayesian Vector Auto-regression (BVAR) and State Space exponential smoothing (Error, Trend, Seasonal [ETS]) mode to forecast tax revenues by using quarterly data from 1998 to 2012 for South Africa. They suggested that ETS methods outperformed BVAR Method for total tax revenue while BVAR were best for major tax types

Ofori et al (2021) compared forecast power between ARIMA with intervention and Holt linear trend method for VAT monthly (higher frequency) data from 2002-2019 in Ghana. They suggested that, ARIMA with Intervention method outperformed the Holt linear trend model in terms of accuracy and precision. Also, Hecht and Zitzmann (2021). They suggested that using the continuous-time model in the N = 1 scenario was unsatisfactory for up to 100 time points and some parameters showed underperformance on some criteria even for 250 time points. They argued that, the 50-time point rule of thumb from the N = 1 is not satisfactory.

Chung et al (2022). Compared how Machine Learning (ML) methods and traditional methods performs in forecasting revenues for local government. They used 31 local governments for time series data with length varies 10 to 21 years. They found that, traditional methods perform better compared to ML algorithms with the exception of property tax.

3. Methodology

3.1 Description of the variables and Sample Size.

The paper intends to answer the question whether using higher frequency (quarterly) data is more powerful in forecasting (closer forecast) than using low frequency (annual) data. In this objective, we will use data from 1996.Q1-2016.Q4 (88 observations) to forecast the period ranging from 2017Q1-2017Q4. Three common taxes VAT (Value-added tax); PAYE (Pay As You Earn), Total tax revenue (TR) will be used as samples due their great contribution to total Tanzania's tax revenue handled by Tanzania Revenue Authority (TRA).

We selected tax base for each tax, VAT, PAYE and TR, based on previous studies as explained in the next section. Before long run estimation using OLS, all data were tested if they were stationary at first difference using Augmented Dickey Fuller (ADF) test, which is widely used in time series analysis. Then, we tested for cointegration using Johansen and Juselius (1991) multivariate cointegration test to ensure that our forecasting regression is not spurious.

We chose to calculate tax elasticity/buoyancy using Ordinary Least Square (OLS) method, as it is the best method compared to point estimate or average point estimates. OLS use regression techniques to minimize the errors between actual and the forecasted values. The best regression was selected based on several forecasting criteria such as root mean squared error (R.M.S.E), mean average error (M.A.E), mean absolute percentage error (M.A.P.E) and Theil's inequality coefficient (TH.I.C). In some cases, we used Adjusted R-squared and Akaike information criteria (AIC). At the end, we compared the forecast values with actual values to come up with forecast power.

3.2 Model Specification:

The theory of taxation states that tax revenue is collected by various means with respect to different types of tax. In our regression, we have three main tax revenues, total tax revenue (TR), Pay-As-You-Earn (PAYE) tax and Value-added Tax (VAT).

TR is expressed as a function of nominal GDP (Klazer, 2013; Bayer, 2015). It implies that, as economic growth expands, it stimulates the growth of tax base such as house hold consumption, domestic investment and international trade. Also, house hold final consumption (HFC) can act as substitute if there is a degree of distortion in GDP trend as it is one of the main determinants of nominal GDP (Bayer, 2015).…………………..……………………………………
…… (1)

Second model specification is PAYE which is expressed as a function of salaries and wages (SW) and unemployment rate (UEM) (Bayer, 2015). The amount of wage and salary of employees can be linked directly to the amount of tax collected through PAYE. The relationship between PAYE and salaries and wages is expected to be positive while that of PAYE and unemployment rate is expected to be negative. The negative relationship is due to belief that, as unemployment increases, wages and salaries rolled out shrinks, in turn, causing PAYE collection to decline subsequently (Kyobe and Danninger, 2005). In some cases, we can also use expansion of GDP as explanatory variable for PAYE since it has a positive correlation with profits gained by individual taxpayers which makes up the tax base for PAYE (Kyobe and Danninger, 2005; Bayer, 2015).

……………………………………………………......... (2)

The third is a VAT Model that uses two different explanatory variables namely total household final consumption (HFC) and nominal GDP. Theoretically, VAT is charged from purchases of goods and services. Thus, household final consumption is its best proxy (See, Jenkins et al, 2000; IMF,1985; Bayer, 2015). In some cases, we can also use expansion of GDP as explanatory variable for VAT due to its positive correlation with total consumption and VAT collection (Kyobe and Danninger, 2005; Bayer, 2015).

………………………………………..……………….(3)

3.3 Data Sources:

Table 4.1 reports the sources of data for each variable. The estimation of these data was done based on the objectives of this study and time scope. For all three models, the data range is from 1996 to 2016, whereby data were grouped into low frequency and high frequency. Three main sources for the data collected include World Bank, International Labour Organization (ILO) and Tanzania Revenue Authority. The variables used in this study are Total Revenue (TR), Value-Added Tax (VAT), Pay-As-You-Earn tax (PAYE), Unemployment (UEM), Wage and Salaries (WAGE), Household Final Consumption Expenditure (HC) and Nominal GDP Per Capita (NGDP).

See annex Table 4. 1 Description of the Variables

4.0 Empirical Results

4.1 Trend analysis of variables

(i)Total Tax Revenue

Although this revenue does not constantly increase over time, it keeps on changing in minor variations in different categories of taxes.

See annex Figure 1: Total Tax Revenue (Bill. TZS)

However, the total tax revenue has progressively increased throughout the studying period. Figure 1 above illustrates the linear trend of total tax revenue. A dip could be seen from 2005 to 2006. The trend then peaks slightly from 2007 to 2008, instigated by extensive reforms passed by parliamentary meeting in 2006/2007 (AfDB, 2010), followed by a drop again from 2009 to 2011 due to the 2008 global financial crisis. Then there was a significant economic growth between 2013 and 2015 followed by a dip in 2016 due to complications surrounding property tax collection.

In general, total tax collection shows a simple linear trend through the period between 2003 and 2017, with slight variations secondary to external factors. Therefore, estimation using linear regression would be well fitted for this analysis.

4.2 Empirical Estimation Results

Here, we report the empirical findings, including unit root test, cointegration, estimation results and the performance of forecast evaluation technique of time series data for tax revenues in various length and frequency.

4.2.1 Unit root and Cointegration Results

See annex Table 1: Unit root test

Before estimation of OLS, we have to verify whether our variables are free from unit root problem and has long run relationships. To determine whether the series has unit root problem, we applied the Augmented Dickey-Fuller (1979) tests, the results of all four sub-samples (higher frequency and lower frequency time series) indicated that the series become stationary at first difference (Table 1).

The cointegration results are presented in Tables 2A and 2B for lower and higher frequency time series data respectively. After considering both trace statistics and Max-Eigen, the results confirm the existence of a long-run relationship. Generally, this concludes that we can estimate and forecast using OLS method, as expected.

See annex Table 2: High Frequency Vs Low Frequency; Cointegration Results

4.2.2 Objective One: OLS Regression Estimation Results for higher and low frequency data

The regression results from Table 3 shows that, household consumption (HC) has negative and statistically significant effect on total revenue (TR). In contrast, the value of nominal GDP has positive and statistically significant effect on both TR and VAT. The impact of unemployment (UEM) and waged and salaried workers (WAGE) has negative and positive effect on PAYE respectively with statistical significance, as expected.

See annex Table 3: OLS regression results for high frequency sample

Table 3 and Table 4 are the regression results of three types of tax components TR, PAYE and VAT. The values Adjusted R-squared, as a measurement of goodness of fit, are above 90 percent for all tax types of both low frequency and high frequency regressions. The higher frequency time series models register higher adjusted R-squared than that of lower frequency time series models. Similarly, Akaike information criteria (AIC) are lower for the higher frequency models than the lower frequency models.

See annex Table 4: OLS regression results for low frequency sample

4.2.3 Results of Forecasting for higher and low frequency data

4.2.3.1 Forecasting Evaluation Criteria

Before embarking on forecasting, we looked into the four evaluation criteria for forecasting, R.M.S.E., M.A.E., M.A.P.E and TH.I.C. Generally, after considering all criteria, the results remain ambiguous (Table 5). However, the evaluation criteria generally show better support on the use of lower frequency data (TR and PAYE) for forecast. However, for the VAT, the criteria support the use of higher frequency data.

See annex Table 5: Revenue Categories Forecasting Error (High Frequency Vs Low Frequency)

4.2.3.2 Actual and forecast tax revenue

Table 6 compares forecast values with actual values between 2016Q1 to 2017Q4 for higher frequency data. Also, we present the actual and forecasted value between 2016 and 2017 for lower frequency data. To simplify the presentation of the information, all data in Table 6 are expressed in percentage form. The real figures can be obtained in Appendix A.

See annex Table 6: Actual and forecasted values for the three taxes in two frequencies using OLS method

The forecasted values for 2016 and 2017 are presented quarterly, for higher frequency data and annual manner for lower frequency data. For TR and PAYE higher frequency data, the difference between forecasted and actual values for the first two quarters of 2016 (2016q1 and 2016q2) are slightly lower compared to those of the lower frequency data. For example, for TR tax components in higher frequency data, we found that, the difference between forecasted and actual values in 2016Q1 and 2016Q2 are 0.27 and 0.319 respectively, lower than that in the lower frequency data which is 0.337.

However, for the next two quarters (2016Q3 and 2016Q4) for TR and PAYE, the differences are now higher for higher frequency data compared to the lower frequency. For example, TR differences in 2016Q3 and 2016Q4 are 0.361 and 0.398, higher than 0.337 recorded in lower frequency data of the same year. For the year 2016, the TR forecasted value for low frequency data is 22.937 (TZS. 9163 billion) while the actual value is 22.6 percent (TZS6532.54 billion), a difference of 0.337 percent (TZS 2630.46 billion). That is a huge difference, when translated to currency value. The actual, forecasted and difference values can be found in Appendix A and the graphs for TR in Appendix B, PAYE in Appendix C and VAT in Appendix D.

Since forecasting power was mixed between lower and higher frequency data, the uncertainty is resolved by the forecasting criteria above, which supported lower frequency data as a better forecaster. Having said the above, the comparison of actual and forecasted values shows better accuracy in low frequency data. This is agreed by several literature which found that using small annual data makes no difference to quarterly data for forecasting purposes (Bahmani-Oskooee, 1996; Masih and Masih, 1996; and Taylor, 1995). Bahmani-Oskooee emphasized that testing ten annual observations makes no difference to 120 monthly observations. He added that, using annual data of 30 years is as good as using quarterly and monthly data over the same period.

On the other hand, the forecasting results shows that, VAT model forecast performs better when higher frequency data was used (Table 6). Its forecasting errors are the slightest, when compared to the actual values. This is in line with Su Zhou (2001) who claimed that, higher frequency data yield power gain and less size distortion when conducting contegration analysis. Ng and Perron (1995) on the other hand, supports using longer data span, and not higher frequency, to yield power gain. Also, Hooker (1993) suggested that, temporal disaggregation increases the power of ADF cointegration test.

To conclude this section, we found that, using low frequency data for forecasting purpose is good for TR and PAYE but the higher frequency data suits VAT better.

5. Conclusion and recommendation

Generally, for the first objective, after considering the four evaluation criteria (R.M.S.E., M.A.E., M.A.P.E and TH.I.C), we found that models with lower frequency time series are better and accurate in forecasting than the models with higher frequency, with the exception of VAT model, which showed otherwise. This means that an annual data is able to give better forecast than the quarterly or monthly data if we are looking at total tax revenue (TR) or Pay-As-You-Earn tax (PAYE). Forecasting VAT, on the other hand, is better looked at in quarterly intervals.

Therefore, we suggest that forecasting tax revenue be done using lower frequency data is better and more accurate forecasts. However, for VAT tax forecast, higher frequency data make for better and more accurate forecast. This, we support Hakkio and Rush (1991) in the sense that sample size is subjective to a field. The nature of the tax can have implication in selection of sample size and data frequency, as being seen in the case of VAT.

Having said all that, our experiment had only tested annual and quarterly for the period 1996.Q1-2016.Q4 (88 observations) to forecast the period ranging from 2017Q1-2017Q4. This limits our result to just comparisons between two conditions. Future studies should divide the time series into more variation to see the extent how frequency over the long horizon and with different tupe of taxes can give more significant results.

6. References

[1.] Bahmani-Oskooee, Mohsen. 1996. Decline of the Iranian rial during the postrevelutionaryp eriod: A productivitya p-proach. Journal of Developing Areas 30:477-92.

[2.] Bayer, O. (2015). Relevance of Input Data Time Series for Tax Revenue Forecasting. Procedia Economics and Finance, 25: 518-529.

[3.] Choi, I., Chung, B.S., (1995). Sampling frequency and the power of tests for a unit root: a simulation study. Economics Letters

[4.] Chung, I. H., Williams, D. W., & Do, M. R. (2022). For Better or Worse? Revenue Forecasting with Machine Learning Approaches. Public Performance & Management Review, 1-21.

[5.] DeCarlo, L. T., & Tryon, W. W. (1993). Estimating and testing autocorrelation with small samples: A comparison of the c-statistic to a modified estimator. Behaviour Research and Therapy, 31, 781–788. https://doi.org/10.1016/0005-7967(93)90009-J

[6.] Hakkio, Craig S., and Mark Rush. (1991). Cointegration:H ow shorti s the long-run? Journal of InternationalM oney and Finance 10:571-81.

[7.] Hawley, S., Ali, M. S., Berencsi, K., Judge, A., & Prieto-Alhambra, D. (2019). Sample size and power considerations for ordinary least squares interrupted time series analysis: a simulation study. Clinical epidemiology, 11, 197.

[8.] Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29(5), 773-785.

[9.] Hooker, M.A., 1993. Testing for cointegration: power versus frequency of observation. Economics Letters 41, 359–362.

[10.] Huitema, B. E., & McKean, J. W. (1991). Autocorrelation estimation and inference with small samples. Psychological Bulletin, 110, 291–304. https://doi.org/10.1037/0033-2909.110.2.291

[11.] Hecht, M., & Zitzmann, S. (2021). Sample size recommendations for continuous-time models: Compensating shorter time series with larger numbers of persons and vice versa. Structural Equation Modeling: A Multidisciplinary Journal, 28(2), 229-236.

[12.] Johansen S, Juselius K (1990) Maximum likelihood estimation and inference on cointegration-with applications to the demand for money. Oxford Bull Econ Stat 52:169–210

[13.] Klazar, S. (2013): Efektivnost predikcí daňových příjmů v ČR. Praha, University of Economics, Prague, 2003. Doctoral thesis.

[14.] Krone, T., Albers, C. J., & Timmerman, M. E. (2017). A comparative simulation study of AR(1) estimators in short time series. Quality & Quantity, 51:1–21. https://doi.org/10.1007/s11135-015-0290-1

[15.] Kyobe, A. J., & Danninger, S. (2005). Revenue Forecasting—How is it done? Results from a Survey of Low-Income Countries. IMF Working Papers, 2005(024).

[16.] Lahiri, K., Mamingi, N., 1995. Testing for cointegration: power versus frequency of observation — another view. Economics Letters 18:381–386

[17.] Masih, Abul M., and Rumi Masih. 1996. Empirical tests to discern the dynamic causal chain in macroeconomic activity: New evidence fromT hailand and Malaysia based on a multivariatec ointegration/vectorer ror-correction modeling approach. Journal of Policy Modeling 18:531-60.

[18.] McCleary, R., Hay, R. A., Meidinger, E. E., & McDowall, D. (1980). Applied time series analysis for the social sciences. SAGE.

[19.] Molapo, M. A., Olaomi, J. O., & Ama, N. O. (2019). Bayesian vector auto-regression method as an alternative technique for forecasting South African tax revenue. Southern African Business Review, 23(1).

[20.] Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. Journal of the American Statistical Association, 90(429), 268-281.

[21.] Otero , J and Smith J, (2000), Testing for cointegration: power versus frequency of observation -- further Monte Carlo results, Economics Letters, 67, (1), 5-9

[22.] Ofori, M. S., Fumey, A., & Nketiah-Amponsah, E. (2021). Forecasting Value Added Tax Revenue in Ghana. Journal of Economics and Financial Analysis, 4(2), 63-99.

[23.] Poole, M. S., McPhee, R. D., & Canary, D. J. (2002). Hypothesis testing and modeling perspectives on inquiry. In M. L. Knapp & J. A. Daly (Eds.), Handbook of interpersonal communication (3rd ed., pp. 23–72). SAGE

[24.] Solanas, A., Manolov, R., & Sierra, V. (2010). Lag-one autocorrelation in short series: Estimation and hypotheses testing. Psicológica, 31, 357-381.https://www.uv.es/psicologica/articulos2.10/9SOLANAS.pdf

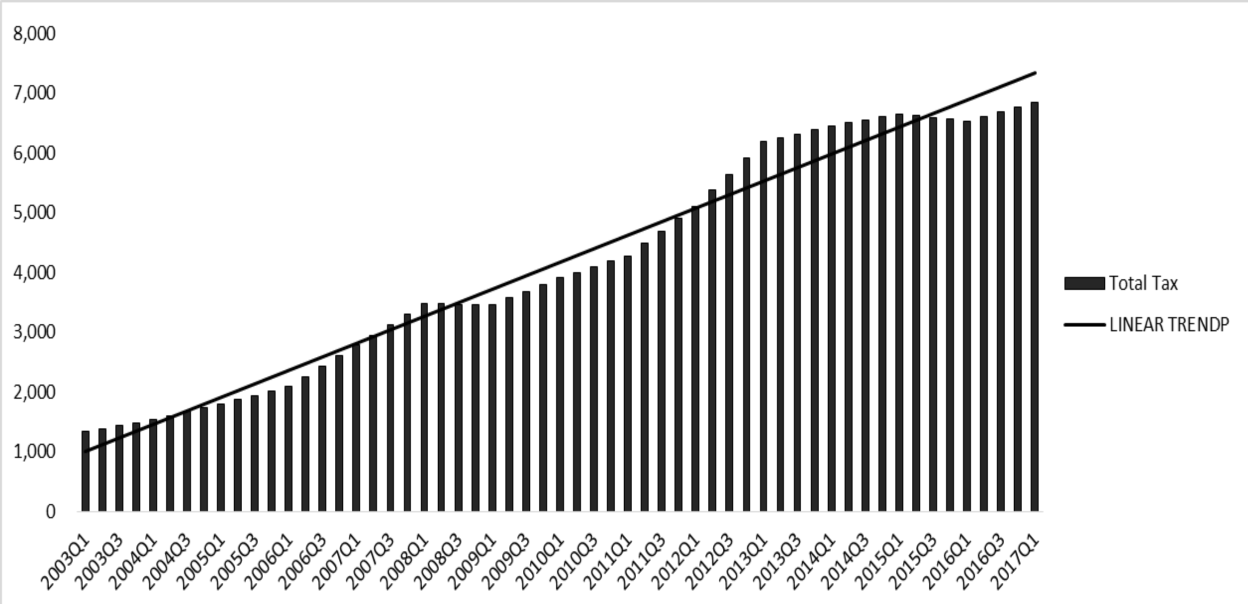
[25.] Sabaj, E., & Kahveci, M. (2018). Forecasting tax revenues in an emerging economy: The case of Albania. University Library of Munich, Germany.
[26.] Taylor, Mark P. 1995. Modeling the demand for U.K. broad money. The Review of Economics and Statistics 75:112-7.
[27.] Warner, R. M. (1998). Spectral analysis of time-series data. Guilford Press.
[28.] Zhou, S. (2001). The power of cointegration tests versus data frequency and time spans. Southern Economic Journal, 906-921.

Annex

Table 4. 1 Description of the Variables

Variable	Description	Measurements	Data source
TR	Total Tax Revenue includes all tax collected by TRA	Total value of tax revenues collected per year expressed in Local currency (TZS) but changed into percentage form.	TRA Website (2021)
VAT	Value-Added Tax is a tax charged at 18 percent from the difference of sales and purchases.	Total VAT expressed in local currency (TZS) but changed into percentage form.	TRA Website (2021)
PAYE	Pay-As-You-Earn is a tax charged from employee's monthly salary and/or wages.	Total PAYE values expressed in local currency (TZS) but changed into percentage form.	TRA Website (2021)
UEM	Unemployment refers to the share of the labour force that is without work but available for and seeking employment.	Unemployment, total (% of total labour force) (modelled ILO estimate)	International Labour Organization
WAGE	Wage and salaried workers (employees) are those workers who hold the type of jobs defined as "paid employment jobs,"	Wage and Salaried workers, total (% of total employment)	World Bank national accounts data, and OECD National Accounts data files.
HC	Household final consumption expenditure (formerly private consumption) is the market value of all goods and services, including durable products purchased by households.	Household final consumption expenditure (constant 2010 US$) but changed into local currency (TZS) for the respective year.	World Bank national accounts data, and OECD National Accounts data files.
NGDP	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.	GDP in local currency	World Bank national accounts data, and OECD National Accounts data files.

Figure 1: Total Tax Revenue (Bill. TZS)



Source: Authors

Table 1: Unit root test

Variable	High Freq.	Low Freq.
TR	-1.471094	-1.19884
VAT	1.371594	-0.4506
PAYE	-0.046728	-0.9848
NGDP	-1.63009	-2.1881
WAGE	-0.041418	0.5582
HC	-0.914652	-1.20601
UEM	-1.727462	0.9248
TR	-3.813435***	-5.1157***
VAT	-7.279520***	-12.7100***
PAYE	-3.866581***	-3.9528***
NGDP	-3.8767***	-3.6118***
WAGE	-4.857253***	-3.1562**
HC	-3.304005**	-5.3948***
UEM	-4.378644***	-3.8434**

Table 2: High Frequency Vs Low Frequency; Cointegration Results

A. Low-frequency equation				
	TR	VAT	PAYE	
Hypothesized No. CE(s)	**Trace Statistic**	**Trace Statistic**	**Trace Statistic**	**Critical Value**
None *	31.01700**	36.82255***	42.22615***	29.7971
At most 1	8.077913	10.39101	11.54128	15.4947
At most 2	1.722718	0.674068	3.367579	3.8411
Hypothesized No. CE(s)	**Max-Eigen Statistic**	**Max-Eigen Statistic**	**Max-Eigen Statistic**	**Critical Value**
None *	22.93908**	26.43155***	30.68487***	21.1316
At most 1	6.355195	9.716939	6.173697	14.2646
At most 2	1.722718	0.674068	3.367579	3.8415
B. High-frequency equation				
	TR	**VAT**	**PAYE**	
Hypothesized No. CE(s)	**Trace Statistic**	**Trace Statistic**	**Trace Statistic**	**Critical Value**
None *	41.62663***	46.53872***	45.71943**	29.7971
At most 1	10.75114	8.681595	20.71205	15.4947
At most 2	1.703075	0.360191	7.655964	3.8411
Hypothesized No. CE(s)	**Max-Eigen Statistic**	**Max-Eigen Statistic**	**Max-Eigen Statistic**	**Critical Value**
None *	30.87549***	37.85713***	25.00738*	21.1316
At most 1	9.048065	8.321403	13.05608	14.2646
At most 2	1.703075	0.360191	7.655964	3.8415

*, **, *** denotes rejection of the hypothesis at the 0.1, 0.05 and 0.01 levels of significance.

Table 3: OLS regression results for high frequency sample

Dependent Variable	**TR**	**PAYE**	**VAT**
Independent variable	**High Frequency**	**High Frequency**	**High Frequency**
Constant	-7.3802***	11.505***	-2.5411***
	[1.7184]	[0.1410]	[0.8964]
HC	-2.2505***		-0.3983
	[0.5326]		[0.2778]
NGDP	3.1551***		1.1149***
	[0.4838]		[0.2524]
UEM		-0.3324***	
		[0.1411]	
WAGE		3.6874**	
		[0.2014]	
Adjust R^2	0.961383	0.961383	0.958074
F-statistic	932.5772	932.5772	857.9310
AIC	-0.766611	-0.766611	-1.150706
BP LM Test	61.737***	61.737***	60.609***
Hetero Test	18.347***	18.347***	11.1303***

Note: *"*","**" and "***"indicate significance levels at 1 percent, 5percent and 10 percent respectively. The values in brackets refer to standard error. All studying variables are presented in percentage form.*

Table 4: OLS regression results for low frequency sample

Dependent variable	TR	PAYE	VAT
	Low Frequency	Low Frequency	Low Frequency
Constant	-7.2912**	11.498***	-2.5982
	[3.3857]	[0.7547]	[1.8784]
HC	-2.2820**		-0.3899
	[1.0564]		[0.5861]
NGDP	3.1835***		1.1085**
	[0.9598]		[0.5325]
UEM		-0.3290	
		[0.3284]	
WAGE		3.6893***	
		[0.2177]	
Adjust R^2	0.936865	0.959204	0.956128
F-statistic	134.5527	212.6118	197.1412
AIC	-0.213684	-0.569393	-0.964616
BP LM Test	4.351460	4.156069	2.717999
Hetero Test	4.334938	4.834098	2.465888

Note: *"*","**" and "***"indicate significance levels at 1 percent, 5percent and 10 percent, respectively. The values in brackets refer to standard error. All studying variables are presented in percentage form.*

Table 5: Revenue Categories Forecasting Error (High Frequency Vs Low Frequency)

Tax form	Observed Freq.	R.M.S.E	M.A.E	M.A.P.E	TH.I.C
TR	**High Freq.**	0.4083	0.4014	1.774%	0.0089
	Low Freq.	0.4078	0.4027	1.772%	0.0088
PAYE	**High Freq.**	0.3952	0.3859	1.868%	0.0098
	Low Freq.	0.3934	0.3865	1.861%	0.0093
VAT	**High Freq.**	0.0883	0.0880	0.4289%	0.0021
	Low Freq.	0.0899	0.0898	0.433%	0.0022

Source: Author's Calculation.

Table 6: Actual and forecasted values for the three taxes in two frequencies using OLS method

	High Freq.				Low Freq.
	2016Q1	**2016Q2**	**2016Q3**	**2016Q4**	**2016**
Actual Value (LNTR)	22.597	22.596	22.6	22.608	22.6
Forecasted Value (LNTRF)	22.867	22.915	22.961	23.006	22.937
Difference	**0.27**	**0.319**	**0.361**	**0.398**	**0.337**
	2017Q1	**2017Q2**	**2017Q3**	**2017Q4**	**2017**
Actual Value (LNTR)	22.619	22.635	22.654	22.677	22.646
Forecasted Value (LNTR)	23.049	23.092	23.133	23.173	23.112
Difference	**0.43**	**0.457**	**0.479**	**0.496**	**0.466**
	2016Q1	**2016Q2**	**2016Q3**	**2016Q4**	**2016**
Actual Value (LNPAYE)	20.785	20.771	20.761	20.758	20.769
Forecasted Value (LNPAYEF)	21.024	21.062	21.101	21.139	21.082
Difference	**0.239**	**0.291**	**0.34**	**0.381**	**0.313**
	2017Q1	**2017Q2**	**2017Q3**	**2017Q4**	**2017**
Actual Value (LNPAYE)	20.759	20.765	20.777	20.794	20.774
Forecasted Value (LNPAYEF)	21.177	21.214	21.252	21.289	21.234
Difference	**0.418**	**0.449**	**0.475**	**0.495**	**0.46**
	2016Q1	**2016Q2**	**2016Q3**	**2016Q4**	**2016**
Actual Value (LNVAT)	20.684	20.703	20.724	20.746	20.714
Forecasted Value (LNVATF)	20.77	20.795	20.819	20.842	20.808
Difference	**0.086**	**0.092**	**0.095**	**0.096**	**0.094**
	2017Q1	**2017Q2**	**2017Q3**	**2017Q4**	**2017**
Actual Value (LNVAT)	20.77	20.794	20.82	20.847	20.808
Forecasted Value (LNVATF)	20.763	20.783	20.901	20.919	20.893
Difference	**0.007**	**0.011**	**0.081**	**0.072**	**0.085**