Revenue and expenditure forecasting techniques for a PER Spending

ANNEX 3:

August 2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

This annex provides a very summary description of a number of methods of forecasting that can be used in a sector PER. Forecasting is the use of various applicable analytical methods to project a variable into the future. In this case, our interest is in the revenues available in the future but expenditure can also be forecasted when necessary using some of the same basic methods. Tax revenue forecasting plays a central role in annual budget formulation. It provides policy makers and fiscal planners with the data needed to guide borrowing, use accumulated reserves, or specify monetary measures to balance the budget. It also informs about what fiscal actions are sustainable and hence how to balance fiscal policy to address the problems in the balance of payments and hence foreign debt.

The annex covers (i) the budgeting context; (ii) the nature and process of revenue forecasting; the steps in forecasting; and the forecasting methods. The methods described in this annex are: (i) qualitative forecasting; (ii) judgement forecasting; (iii) micro-simulation models; (iv) moving average methods, including ARIMA; (v) exponential smoothing and Holt-Winters methods; (vi) single equation regression forecasting; and (vii) macroeconometric and GDP-based forecasting models. Expenditure forecasting is considered briefly in the Annex when macroeconomic revenue forecasting is addressed. The forecasting of the expenditure ceiling will be illustrated.

The numerical methods of the Annex will be illustrated with Stata code. The note on microsimulation is designed largely for awareness of the existence of a forecasting method that is being increasingly used in revenue forecasting. In a setting where data are scarce, such as in the partner countries, it will normally be necessary to rely heavily on qualitative and judgemental forecasting. Successful use of these methods in turn relies on the existence of inclusive arrangements for multi-stakeholder participation in budgeting that encourage dialogue and transparent information-sharing. Forecasting is a specialist exercise that is prone to error and requires the build-up of expertise over time. Dialogue and information sharing helps to minimize error in the forecasting exercise.

1.1 The Budgeting Context

Sector-specific public expenditure reviews are integral components of the national budget cycle when the purpose of budgeting is to maximize the net outputs and outcomes of government expenditure. Such ‘performance-based’ budgeting relies on good forecasts of revenues from all sources. It also involves forecasting and prioritizing of the resulting expenditures of all sectors and of the economy as a whole, with particular reference to expenditures on:

1. Infrastructure for social and economic development, including poverty reduction.
2. Social protection services for poverty reduction.

Forecasting is needed because the start of the process of annual budgeting precedes the actual expenditures of budgeted fund by almost a year. The need applies with greater force to expenditures associated with medium-term and long term planning, which look ahead for 3-5 years.

The need for forecasting applies to sector budgeting, local government budgeting, and national budgeting. Further, in the budgeting process, sector-specific forecasts are also needed to ensure that sector budgeting is mutually consistent with budgeting at the level of the local communities and budgeting to meet commitments on internationally-agreed development goals. In all cases, performance-based budgeting requires routine measurement of the economy, efficiency and effectiveness of the expenditures on infrastructure and social services, as part of the process of prioritizing them.

There is no commonly-accepted standard method of producing macroeconomic and revenue forecasts for a sector PER. The PER Team will have to choose from, and give its own weight to, a set of modelling techniques, consumer and business surveys. Apart from microsimulation, the list of methods described in this manual can be treated as a basic set that can be used with the data available to any PER in the partner countries. Important among these is the use of expert opinion and judgement. Fiscal policy in the partner countries must be concerned with improving the consistency of resource allocation within a medium-term fiscal framework that promotes economic development. Thus, the PER should also be guided by a tractable small structural model such as illustrated in Annex 2. Such a model allows focus on long time horizons (greater than a year) rather than business cycles and treats the economy as an evolving system.[1] For the budget year itself, the PER should supplement this with single structural econometric equations and individual variable forecasts that preserve the underlying consistency of the macroeconomic accounting supply and demand identities. Preservation of the identities allow the forecasts to capture the short-run adjustment of the economy to the long run development path. The identity-preserving model can also be used for 'what if' scenarios, such as are usually reported in the IMF Article IV reports. In this case, the model is shocked and the resulting output compared with a baseline scenario used for budgeting.

Use of judgement developed from in-depth analysis of the major sectors of the economy is an important aspect of this effort. This judgement should be developed through dialogue with the business community and other informed stakeholders along with expert opinion and judgement to produce macroeconomic and revenue forecasts. Indeed, a regular dialogue with the business community can help identify emerging trends in the economy with respect to variables such as investment and exports. Dedicated resources should be allocated to such regular engagement with the business community. In the USA, the results of this dialogue are published as the US Federal Reserve's Beige Book; in the UK, the Bank of England's Agency Report; in Canada, the Bank of Canada's Business Outlook Survey.

[1] Business cycles are defined as fluctuations or oscillations in economic data that recur periodically, for periods between 1.5 and 8 years. Periods longer than 8 years are parts of long waves or long trends. The main reference for this view is Burns, A. F., and Mitchell, W. C. (1946). *Measuring Business Cycles*. New York: National Bureau of Economic Research.

2 The Nature and Process of Revenue Forecasting

Accurate forecasting of revenues and expenditures is important for avoiding both underfunding and excessive funding of the government, and related consequences of associated surpluses or deficits. Forecasting uses available data and methods of analysis to estimate the value of a variable in the future. Here we are concerned with revenues and expenditures.

2.1 What Revenues to Forecast

Revenue forecasting seeks to estimate inflows from the following sources:

1. Tax revenues identified from the taxable capacity of the sectors and aggregated to the national level, including fees, permits, and licences.
2. Sales by sector agencies from productive business activities.
3. Intergovernmental transfers received by the sectors from the national pool.
4. Sector funding from international sources, including earmarked donor support, and other loans and grants.
5. Hidden industrial sector transfers and ‘off-budget’ funding, whereby the sectors finance activities and institutions that benefit government, such as schools, clinics, libraries, and the like.

Revenue forecasts can apply to aggregate total revenue or to single revenue sources such as sales tax revenues or property tax revenues. The forecasting methods seek to specify and identify the substantive and numerical relationships between the factors that determine taxable capacity and the amounts of revenues actually collected.

1. The factors that determine the taxable capacity are defined in terms of
    a. Productive capacity - value-added - by the industrial sectors
    b. Tax rates, including value-added taxes.
    c. User fees and costs of permits issued.
    d. Draw-downs rates from sales of government-owned business enterprises.

2. The actual amounts of revenues government collects in the form of
    a. Income, profits taxes – PAYE
    b. Domestic taxes on goods and services
    c. Property taxes
    d. Other taxes
    e. Draw-downs from sales of government-owned business enterprises
    f. Receipts of social security contributions

2.2 What Expenditures to Forecast

Expenditure forecasts can apply to aggregate total expenditure or to single expenditure categories. Here too, the forecasting methods seek to specify and identify the substantive and numerical relationships that determine the government’s spending program as classified under COFOG.

Within each COFOG category, expenditure forecasting seeks to estimate outflows of the following:

1. Intermediate Consumption
2. Compensation of Employees
    a. Wages and salaries
    b. Employers’ social contributions
    c. Employers' imputed social contributions
3. Tax incentives and allowances on Production and Imports
4. Subsidies
5. Tax incentives and allowances on
    a. Property Income
    b. Income and wealth
6. Social benefits other than Social Transfers in-Kind
    a. Pensions
    b. Unemployment benefits
    c. Long-term care benefits
    d. Family benefits
    e. Benefits n.e.c.
7. Social Transfers in-kind
8. Other Current Transfers
    a. To Households
    b. To CARICOM, etc.; To OECS
    c. Current transfers n.e.c.
9. Capital Transfers
    a. Bank support package
    b. Capital transfers n.e.c.
10. Gross Capital Formation
    a. General infrastructure to support enterprise
    b. Other general services
    c. Health
    d. Education
    e. Research and development
    f. Road Transport and Utilities
    g. National Security
    h. Depreciation/Maintenance
11. Contingencies

2.3 The Forecasting Steps

2.3.1 Step 1: Selection of forecast period

This step involves selection of a period over which budget data are examined, which depends on:

a. The availability and quality of data,
b. The type (and sustainability) of the revenues or expenditures to be forecasted,
c. The degree of accuracy sought.

There are many appropriate time-frames over which the forecasts should be prepared. Much depends on who is forecasting and what is forecasted. The central government might look at forecasts for one year ahead to forecast gross revenues or expenditures. To invest appropriately, the managers of the electricity supply, the Ministry of Education and the Ministry of Health must look ahead for 20 years to forecast demand and thus must forecast revenues and expenditures in the same time frame.

2.3.2 Step 2: Examination of data for stationarity

This step involves examination of the data for symmetries, such as trends and rates of change. Here, the main concern is to identify evidence of:

a. Stability
b. Nonstationary nonlinear paths, such as
    i. Exponential growth
    ii. Quadratic paths
    iii. Structural breaks and seasonal or cyclical variation.

The predictability of the budget category must be assessed, given the evidence on the trends. This is based on the characteristics of the category, such as

a. The rate structures approved for revenue collection.
b. The level demand and rate of change of demand for revenues from the relevant source.
c. Significant seasonal or cyclical variation in expenditure.

2.3.3 Step 3: Adoption of assumptions

Government fiscal policy is affected by economic, social and political forces. Accordingly, forecasting requires explicit assumptions and processes. This step involves adoption of applicable assumptions about the revenue source that affect the methods used, including how the revenues are affected by:

a. Changing economic conditions
b. Changing population size and citizen demand
c. Changing government policies.
d. Changing administrative procedures

2.3.4 Step 4: Selection of forecasting methods

This step involves actual selection and application of the methods to estimate or forecast revenue collections in future years. The method selected depend on the nature and type of revenue.

1. Qualitative and judgemental methods are needed for revenue sources that are highly uncertain, including:
   i. New revenues sources
   ii. Grants
   iii. Asset sales or sales from government-owned business activity.
2. Quantitative methods are needed when the revenues are based on greater certainty, such as
   i. Revenues from income.
   ii. Revenues from habits, such as the ‘sin taxes’ – taxes on alcohol and cigarette sales.
3. When in doubt, the method selected should be the simplest of the options available, subject to the evaluations in steps 5 and 6 below.

Following the selection, the methods are applied to obtain the estimates sought. More than one method might be used and the results averaged.

2.3.5 Step 5: Evaluation of estimates

At this step, evaluation of the estimates is done to ascertain their reliability and validity.

a. Evaluation of validity requires validation of the assumptions about the revenue source. Valid estimates require sound assumptions about the existing environment:
   i. Economic.
   ii. Population (demand).
   iii. Administrative.
   iv. Political.
b. Evaluation of reliability is based on sensitivity analysis. This involves:
   i. Varying key parameters used to create the estimates.
   ii. Assessing if small parameter changes result in large changes in the estimates or small.
      A. If large, the projection is given a low degree of reliability.
      B. If small, the projection is given a high degree of reliability.

2.3.6 Step 6: Monitoring of outcome and comparison with forecasts

Actual revenue collections are observed and compared against the estimates. The extent of deviation or 'error' is measured and used to assess the accuracy of the forecasts. As a general rule, a forecast should be 'unbiased' in the sense that the expected value of the deviation of the actual estimates and the forecast should be zero. Frankel (2011)[2] has warned that bias is typical of government forecasting. The degree of accuracy is a measure of the likelihood that there will be revenue and hence budget shortfalls or surpluses. From this standpoint, forecast accuracy means that the actual forecast errors should be minimised to the greatest extent possible. The PER Team should be aware that forecast accuracy is related to the considerations in Step 2. Forecasting errors will tend to be larger when the data is non-stationary and smaller when the data is stationary. Non-stationary behaviour arises near turning points in the economic seasons and cycles, when the economy is farthest above or below its trend. Stationary behaviour will be observed when the economy is growing near to its trend growth path.

2.3.7 Step 7: Updating the Forecasts

In this final step, the forecasts are updated if the assumptions must be changed. Specifically, updating is needed when the conditions in the economy are changing. These relate to:

a. Economic forces.
b. Population (demand).
c. Administrative arrangements.
d. Political developments.

[2] Frankel, J. (2011). Over-optimism in Forecasts by Official Budget Agencies and its Implications, *Oxford Review of Economic Policy* 27(4): 536-562.

3 The Forecasting Methods

Many forecasting methods are available, ranging from relatively informal qualitative methods to quantitative methods that are highly sophisticated. They are widely used in revenue forecasting (Kyobe and Danninger, 2005).[3] Classic references are Bowerman and O'Connell (1993)[4], Box and Jenkins (1976)[5], Brockwell and Davis (2002)[6], Chatfield (2004)[7], and Hamilton (1994).[8]

As indicated above, in forecasting, the choice of methods depends on what is being forecasted and the time-horizons of the forecasts. Sophistication is not the same as accuracy and simplicity is an asset when many categories of revenues or expenditure are being projected. In the case of revenue forecasting especially, since many categories of tax are involved, the volume of work might be very large. This recommends less sophisticated methods that are quite reasonable in the light of the data challenges and the costs of using more sophisticated methods.

Whatever the choice, all revenue and expenditure forecasting assume that the past combines with current discretionary policy interventions to predict the future. That is, let X_t be the category of revenue or expenditure, $g_{x,t}$ the *forecasted rate of change of the category* from $t-1$ to t, and DP_t government's discretionary policy intervention in t. Then, the general budget forecast equation is:

1. $X_t = (1 + g_{x,t})X_{t-1} + DP_t$

All of the forecasting methods described below are applications of equation (1). It will be referenced repeatedly in this Annex.

3.1 Qualitative and Judgement Forecasting Methods

Qualitative forecasting methods are based on judgements or 'guesses' about the trends in the revenues that can be expected from each category. The judgements can be provided by experts or by consensus among interested stakeholders who are involved in some way with the tax source and can make reasonable assessments of what is likely to happen in the future. This is the method that is perhaps most appropriate under two conditions:

a. When there is little applicable data available.

b. When there is rapid change in the environment and hence the assumptions – economic conditions, population growth, technological conditions, administrative arrangements, political conditions, and the like.

---

[3] Kyobe A, and S. Danninger. (2005). Revenue Forecasting-How is it Done? Results from a Survey of Low-Income Countries. *IMF working paper* No.WP/05/24, IMF, Washington D.C., USA.

[4] Bowerman, B. L. and O'Connell, R.T. (1993). *Forecasting and Time Series: An Applied Approach,* 3rd ed. Pacific Grove, CA: Duxbury.

[5] Box, G.E.P and Jenkins, G.M. (1976). *Time Series Analysis, Forecasting, and Control*. San Francisco: Holden-Day.

[6] Brockwell, P.J. and Davis, R.A. (2002). *Introduction to Time Series and Forecasting*. New York: Springer.

[7] Chatfield, C. (2004). *The Analysis of Time Series,* 6th ed. Boca Raton, FL: Chapman & Hall/CRC.

[8] Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, NJ: Princeton University Press.

c. When the revenue and expenditure categories being forecasted depends primarily on the policy choices of government, and are not linked directly to macroeconomic developments or do not follow a stable trend. Examples of these are government spending to support the financial sector, revenues obtained from public sector entities, and expenditure on large infrastructure projects. In equation (1), this means that the forecast is shaped mainly by DP_t. The previous year's level does not matter much.

The qualitative estimates can also be based on a study of the trends in available data, but will tend to be heavily influenced by what is assumed about the environment and how changes will affect evident trends.

Qualitative forecasting is highly susceptible to political influence, especially information about what government plans to do and information about the revenue imperatives of the budget. Notwithstanding its value when data are limited, excessive reliance on qualitative judgements tends to lead to error. There is no consistency of the assumptions of different experts and the method provides no protection against spurious correlations and assertions about causal relationships. Dialogue for mutual consistency is also needed to get the best out of this approach, but even in this case the consensus might be dominated by strong and influential experts in the group. The method is perhaps most useful as a complement to strong quantitative methods that can be produce and used by some of the experts available to government. In this case, the final estimates should be produced by adjusting for the qualitative assessments by experts from different fields or ministries who have some significant qualitative information about, and experience with, the revenue sources.

3.2 Quantitative Methods

Quantitative methods rely heavily on available data that are closely related to the revenue source. There are two general types of quantitative forecasting methods: (i) time series forecasting based on identification and projection of past trends; and (ii) construction of causal models using specification of a model that relates the particular revenue type to the variables that are assumed to cause it or to associate strongly with it.

These methods make and transmit explicit assumptions to the forecasts, using very specific numerical procedures the precision of each of which can be specified. That is, they assign an estimate of the 'margin of error' or degree of possible error to the forecasts. Time series methods usually provide better forecasts but, when well-done, the causal models have the advantage of providing more reliable information about the interactions among the factors that inform the ultimate forecasts. Annex 2 provides an example of how interactions are analysed with causal models. In general, quantitative methods are better than qualitative methods when forecasting revenues.

Below is annotated Stata code to allow repeated numerical experimentation with the data. Used as written, the code will allow any data to be loaded and used without modification. Other code must be added to generate practical output. In this Annex, time series data will be used, so the code indicates that Stata must be told with "**tsset**'.

```
>capture program drop NAME1 // this line allows repeated use of  the program name
>program NAME1 // names the program to be run
>capture log close // keeps a running log of the work for review
>log using NAME2, replace // names the log of the work
>clear // used to allow repeated runs with original data in memory, original dataset not modified
>set matsize 1000 // optional, only useful if running large models
>set more off // prevents repeated halting of the reporting of the results generated
>use C:\NAME3\name4.dta // path code to the Stata dataset called name4.dta
>tsset month, monthly // tells the program that monthly time series are being used
>end // indicates that the computations have ended
>NAME1 // tells Stata run this program
>log close // closes the log of the work being done
>exit
```

3.2.1 Time Series Forecasting – Basic Concepts

Time series forecasting of revenues is based on the assumption that patterns in the historical data of a data series can be used to project future revenues. The method relies on the concepts of trend, cycle, season and random change.

3.2.1.1 Distinguishing Trend, Cycle, Season and Noise

In general, a time series is usually decomposed into four components: (i) a trend, which is normally of greatest interest to the forecaster; (ii) a seasonal component; (iii) a cyclical component, which reflects the effects of business cycles; and (iv) a random (noise) component.

Trend

Trend is the general (average) direction of change of a data over time. If the time series increases over time, then it is said to have a positive trend. If it decreases over time, it is said to have a negative trend. In either case, *the mean of the series is changing over time* and the series is called *non-stationary*. If direction does not change in either direction, then the series has no trend. Its mean is constant over time and the series is call *stationary*. This is a simple test to apply:

a. Divide the data in several parts – say 5 (sequential) parts.
b. Compute the mean and the variance of each part.
c. If the mean or variance of each part is about the same as the mean or variance of all the other parts, then the time series is stationary.
d. If the mean or variance of the series is different from those of any of the other parts, then the time series is non-stationary.

Identification of trend requires a series that is long enough for the patterns to be discernible. The general rule for the length of the series is that if the number of observations is less than 50 observations, it is ‘short’ and requires special methods. Most time series methods can be used if a series is equal to or longer than 50 observations. The trend of a time series may have any of several mathematical characteristics, for example: (i) linear, if it increases or decreases at a constant rate over time; (ii) exponential, if the level of the series increases by a constant *percentage* over time.

Cyclical

The cyclical fluctuations of a revenue time series result from the business cycles of the economy. They are mainly the fully worked out medium term results of fluctuations in capacity utilization and are the focus of medium-term expenditure frameworks of countries. Introduction of value-added taxes (VAT) and even user fees in health and education in the partner countries have made the revenue flows significantly dependent on the business cycles of their economies. So, these effects are now relatively important to a PER.

Seasonal

Seasonality of revenues refers to revenue cycles that change within a year. If a time series tends to vary over the course of a year in response to some phenomena, including variations in output with a fixed capital stock, or the effects of moving from the dry season to the wet season, or the times when schools are on holidays, then it is said to have a ***seasonal*** component. These components are of great interest to revenue forecasters. An interesting example is the spike in customs duties or air-travel taxes, or even VAT during the winter season of the North Atlantic, when the Caribbean has balmy weather and visitor arrivals spike.

The PER Team should examine carefully how the seasonal component and the trend component relate, because this has a direct bearing on the methods used to smooth the series and construct forecasts. The algebra is simple and intuitive. Consider vat_t, with a trend, T_t, and a seasonal component, S_t. Trend and season could be additive, in which case, we would check with:

2. $vat_t = T_t + S_t$

If we suspect that they are multiplicative, then we would use:

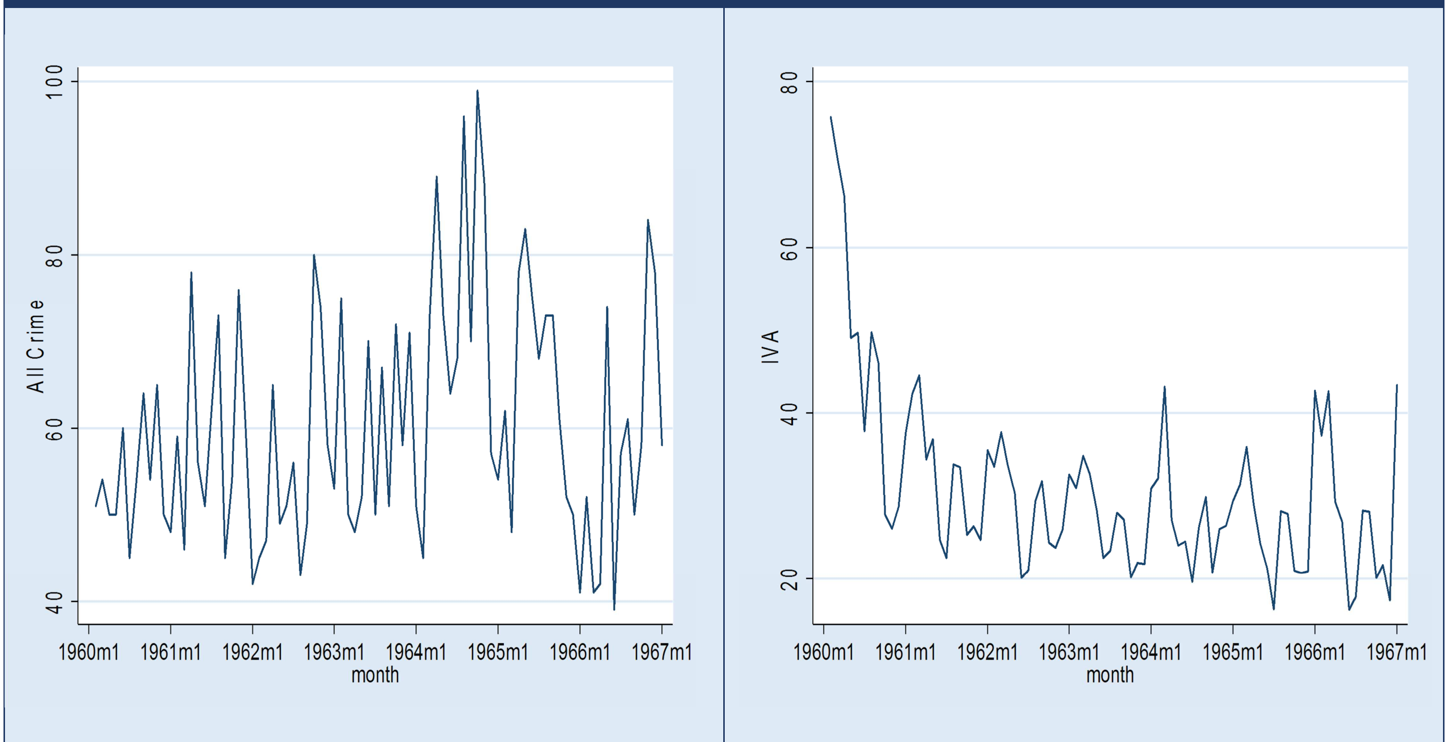
3. $vat_t = T_t * S_t$

How can the PER Team tell? The rule is simple. Check the ***size*** of the seasonal fluctuations and how they change along the trend:

a. If the size is not related in any way to the level of the trend of the time series, then T_t and S_t are additive.
b. If the size is related to the trend level of the series, then T_t and S_t are multiplicative effects.

Consider Figure 1, Panel A. It portrays monthly crime in one Caribbean country. It is clear from examination of the data that serious crime fluctuates more widely in the later months of the series when the trend seems to be higher. The right approach here is to treat the trend and seasonal effects as multiplicative. On the other hand, the data in Panel B seems to fluctuate independently of the level of the series. The series shows additive trend and seasonal effects. With a little experience and a lot of graphing, the analyst can tell quickly which of these scenarios applies. Multiplicative effects in time series are far more common than additive effects.

Figure 3-1: Monthly Data on Crime and Visitor Arrivals in a Caribbean Country



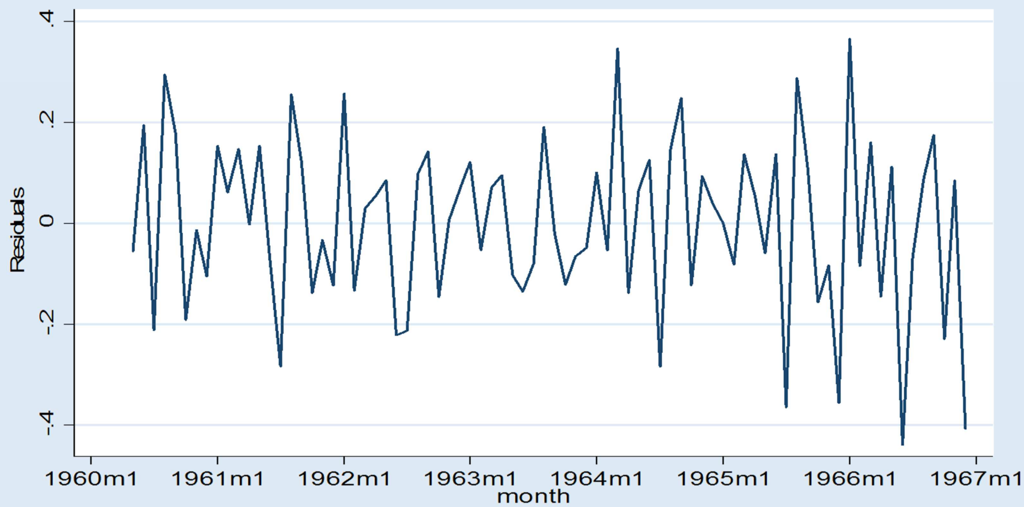
Panel A: Crime data in a Caribbean Country

Panel B: International Visitor Arrivals in a Caribbean Country

3.2.1.2 Random Noise

Randomness refers to unexpected events that may distort trends that otherwise exist over the long-term. These are often called ‘noise’ or ‘shocks’. For example, natural disasters might affect the economy and lower tax flows. Time series forecasting turns on the ability to remove from the data the random component. Removal of noise is called ‘smoothing’ or ‘filtering’ in the professional literature. Random noise is like the residuals of a regression model – the deviations of actual data from a fitted model. Figure 2 illustrates. The random fluctuations represent the effects of all sorts of factors that analysts assign to randomness because they cannot explain them with the time series model of trend, cyclical and seasonal effects. When noise is removed from a series, the other components of the series are made more visible and can be identified with well-known techniques and modern statistical software, and can also be used for forecasting.

Figure 3-2: An example of random noise

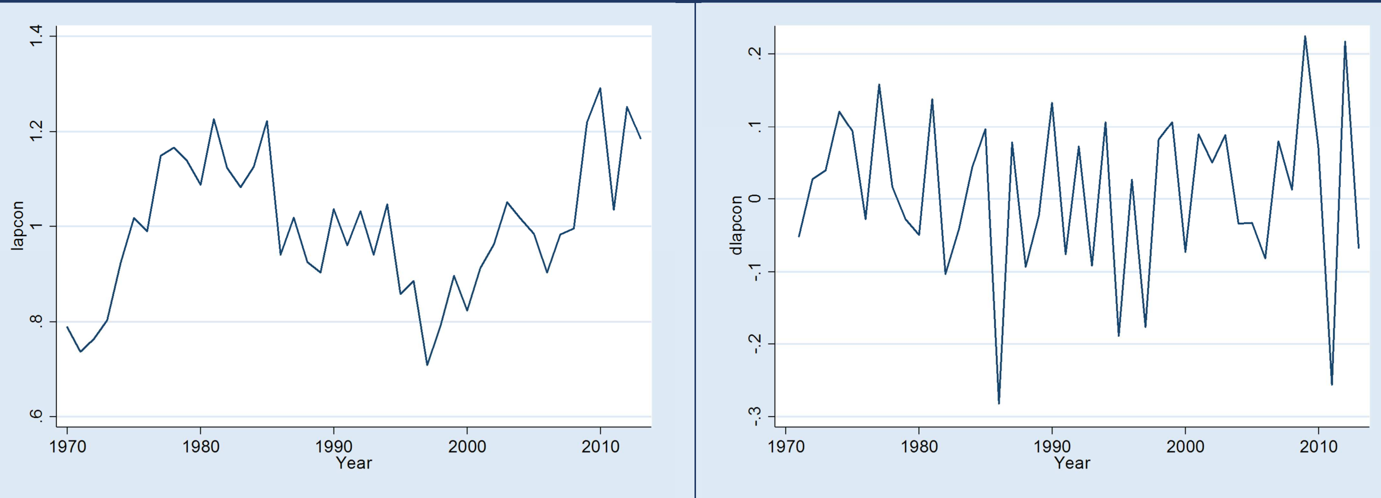


3.2.1.3 The Simplest Forecast – the Naïve Forecast

The simplest forecast of a budget item (revenue source or an expenditure category) is often called a 'random walk' or naïve forecast. The method is very useful when forecasting small, erratic budget categories for which the sign or direction of movement of the previous trend is not robust indicator when choosing the forecast period. The method assumes that the revenue available now, time t, is the same as the revenues that will be available in the immediate future, time $t+1$. In relation to equation (1), this means that $g_{x,t} = 0$. The naïve forecast is therefore widely used in combination with judgement forecasting of revenues.

For example, suppose the data in Panel A of Figure 3 represent the VAT collections of Trinidad and Tobago since 1970. Panel B is the first difference of that data. The difference is computed by subtracting VAT for the last period from VAT for the current period.

Figure 3-3: Data illustrating a random walk



Panel A: Effective Consumption Capacity for T&T

Panel B: First Differencing of Effective Consumption in T&T

Then, Panel B suggests that a reasonable model for the VAT data might be:

4. $vat_t - vat_{t-1} = \varepsilon_t$

where ε_t is white noise. White noise is a special form of randomness. It is central to the analysis of time-series data. A variable ε_t is *white noise* if, for all t,

a. Its expected value equals 0.
b. Its variance is a constant σ^2.
c. For any variable such as vat_t, the draws are independent. This means that the correlation between ε_t and ε_s is zero for s not equal to t.

If ε_t has a normal distribution, then it is called Gaussian white noise. This concept is mentioned here because it is one of the most common benchmarks in all forecasting and it will be needed for the rest of the Annex.

The model in Equation (4) can be written as:

5. $vat_t = vat_{t-1} + \varepsilon_t$

This is a random walk model. It is widely used for non-stationary data. Random walks usually have long periods in which there appears to be a trend upwards or downwards, which can change unpredictably and move in the other direction. The *forecast* of VAT for the next year would then be:

6. $forecast\ vat_{t+1} = actual\ vat_t$

There are several variants of such a forecast:

a. The past two periods can be averaged to make the forecast.
b. The current observations plus a guesstimate for seasonal effects could be used as the forecast.

An alternative to equation (5) is for the budget item to have a path defined by

7. $x_t = \delta + x_{t-1} + \varepsilon_t$

where δ is a constant term and ε_t is white noise. This is a case of a *random walk with drift*, where the drift term is δ. It means that the current value of the budget item is equal to its value last period $(t-1)$ plus some fixed amount δ some white noise, ε_t. Notice two things:

a. The mean of the small budget item depends on t, so the item is not stationary.
b. If equation (7) applies, then the best prediction of x_t is simply $\delta + x_{t-1}$.

There is one additional type of random walk that the PER Team should consider. That is one with both a time trend and drift. It generates the budget item according to the following equation (8):

8. $x_t = \alpha + \delta t + x_{t-1} + \varepsilon_t$

Here, α is a constant term, δ is the coefficient of the time trend indicating how fast x_t changes with time. The best predictor in this case is $\alpha + \delta t + x_{t-1}$.

In addition to their use in naïve forecasting, these concepts will become clearly later in the manual when doing more sophisticated forecasting.

3.2.1.4 The Moving Average Forecast

The moving average forecast is perhaps the most commonly used forecast of budget categories. It makes the average from arbitrarily selected recent data in the series. Moving averages are simple linear filters that do a good job of forecasting and is often the standard against which other methods of budget forecasting are judged.

Like any average, the moving average eliminates the random elements and smooths the data by using an average of parts of the data. To understand a moving average of revenues one must think about 'leads', which are periods ahead of time t, and 'lags', which are past periods. The conventional label for the number of periods ahead is F and for the number of lags is L. Then, for x_t the observations of a time series, the future value forecasted is the moving average $M_{x,t}$ of x_t. This is defined generally as:

$$9. \quad vat_{t+1} = M_{x,t} = \frac{\sum_{i=-L}^{F} w_i x_{t+i}}{\sum_{i=-L}^{F} w_i}$$

In this equation:

- $w_i, i = -L \ldots 0 \ldots F$, is the set of weights to be assigned to the measure x_{t+i}.
- x_t is the variable to be forecasted (in this case vat)
- L is the longest lag in the span of the filter
- F is the longest lead in the span of the filter

3.2.1.5 Equally Weighted Moving Average

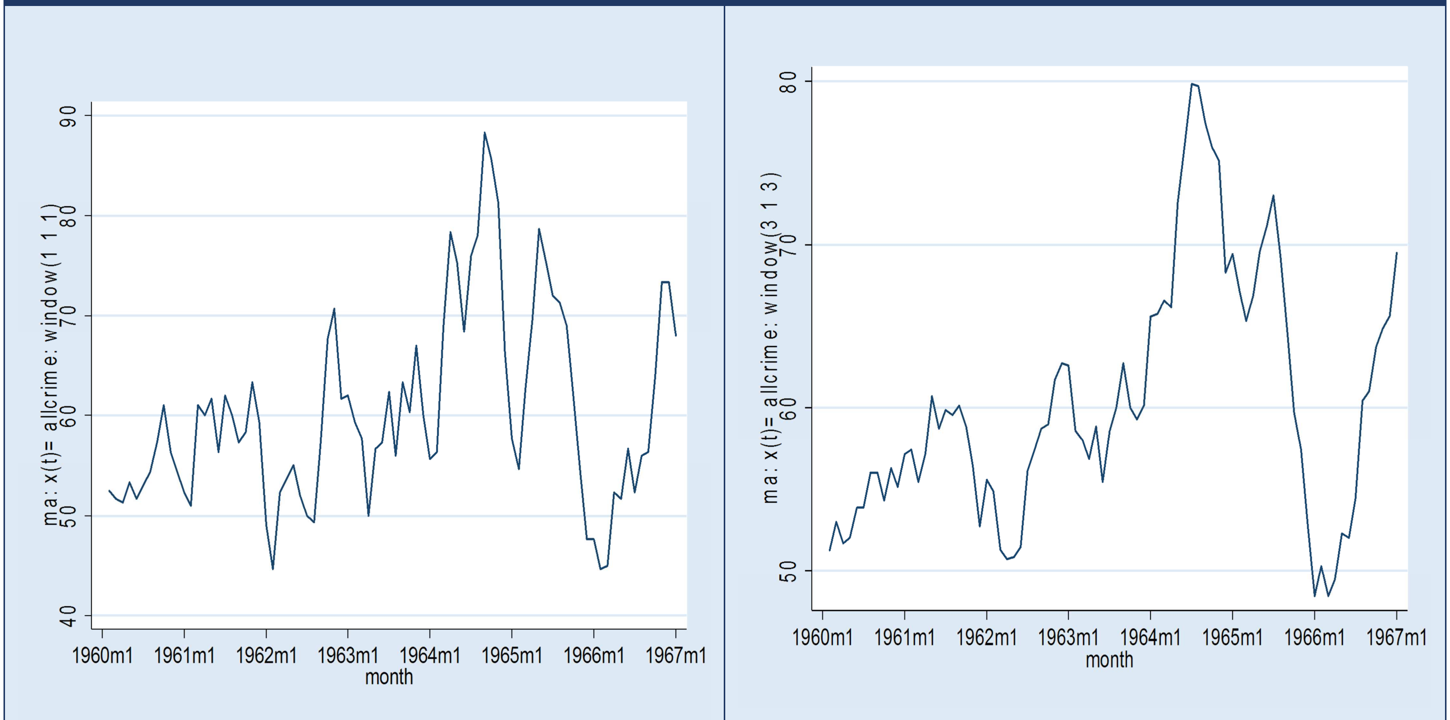
When all the weights are the same and conventionally set at 1, a special case is the equally weighted moving average of the form. The forecast is:

$$10. \quad vat_{t+1} = M_{x,t} = \frac{1}{F+L+1} \sum_{i=-L}^{F} x_{t+i}$$

The definition assigns equal weights to all observations. For a "symmetric" moving average, it is necessary to set $F = L$. So, for a three-period moving average, one sets $F = L = 1$. This means the average is computed with the first lagged term, the current term, and the first lead term. Common practice is determined by the number of time periods of tax collection in a year. Thus, for quarterly tax payments only four terms are used, and for monthly tax data twelve terms.

The graph in Panel A of Figure 4 is a 3-period symmetric moving average of the data in Figure 1. The graph in Panel B is a 7-period moving average, so it uses the first 3 lags, the current period, and the first 3 leads. The averages have eliminated some of the randomness in the data, leaving a smooth trend (trend-cycle) component. The graph in Panel B is smoother than that in Panel A, but the true trend-cycle might be smoother still.

Figure 3-4: 3-period and 7-period Symmetric Moving Averages



Panel A: 3-period moving average | **Panel B:** 7-period moving average

Using the program written in the Introduction, the above graphs can be generated with additional Stata **'tssmooth'** code, as follows:

```
>tssmooth ma svat3=vat, window(1 1 1)
>tsline vat svat3 – to see the original and the smooth data
>tssmooth ma svat7=vat, window(3 1 3)
>tsline vat svat7- to see the original and the smooth data
```

The general form of the code is

```
>tssmooth ma newvar=exp [if] [in], window(#1 #2 #3)
```

The type of forecast made with a moving average depends on the type of variation in the data. Reconsider the multiplicative representation of the VAT data above (equation 3). If we only concentrate on the seasonal effects, and if there is good graphical reason to believe that the size of the shocks is related to the trend level of the series, then the random shocks ε can be incorporated multiplicatively as

11. $vat_t = T_t * S_t * \varepsilon$

This is the Bowerman and O'Connell (1993: 355)[9] *multiplicative decomposition model* of a time series. Since a moving average creates an average over several periods, it always tends to remove the noise, especially if the periods are well-chosen. If the periods F and L are long enough, for example $F + L + 1 \geq 12$ for monthly data, then the seasonal component might be removed as well. In that case, all that remains is the trend-cycle component. If we choose $F = 5$ and $L = 6$, or for $F = 6$ and $L = 6$, the forecast is

12. $vat_{t+1} = M_{vat,t} = \frac{1}{F+L+1} \sum_{i=-L}^{F} vat_{t+i}$

It is also possible to forecast the seasonal behavior of the revenues. Since $M_{vat,t}$ is just the trend-cycle, it must also follow that

13. $S_t * \varepsilon = \frac{vat_t}{M_{vat,t}} = \frac{T_t * S_t * \varepsilon}{M_{vat,t}}$

This equation isolates the seasonal and noise components of the time series. For example, for the 7-period moving average, the applicable Stata code is:

```
>g seasvat=vat/svat7
```

When the method is applied to the series in Panel A, Figure 1, using the 7-period moving average, the result is the seasonal series in Panel A, Figure 5.

Now, since the seasonal and noise components are multiplied, the mean of each cannot be individually identified in this equation. However, a bit of statistical theory helps here. In linear regression analysis with a constant term, it must be assumed that the mean of the error term is zero if the constant term is to be identified. This same assumption used here amounts to assuming that the mean of ε, is 1, i.e., $E(\varepsilon) = 1$. If that assumption holds, it must also hold that the process of taking the average yields

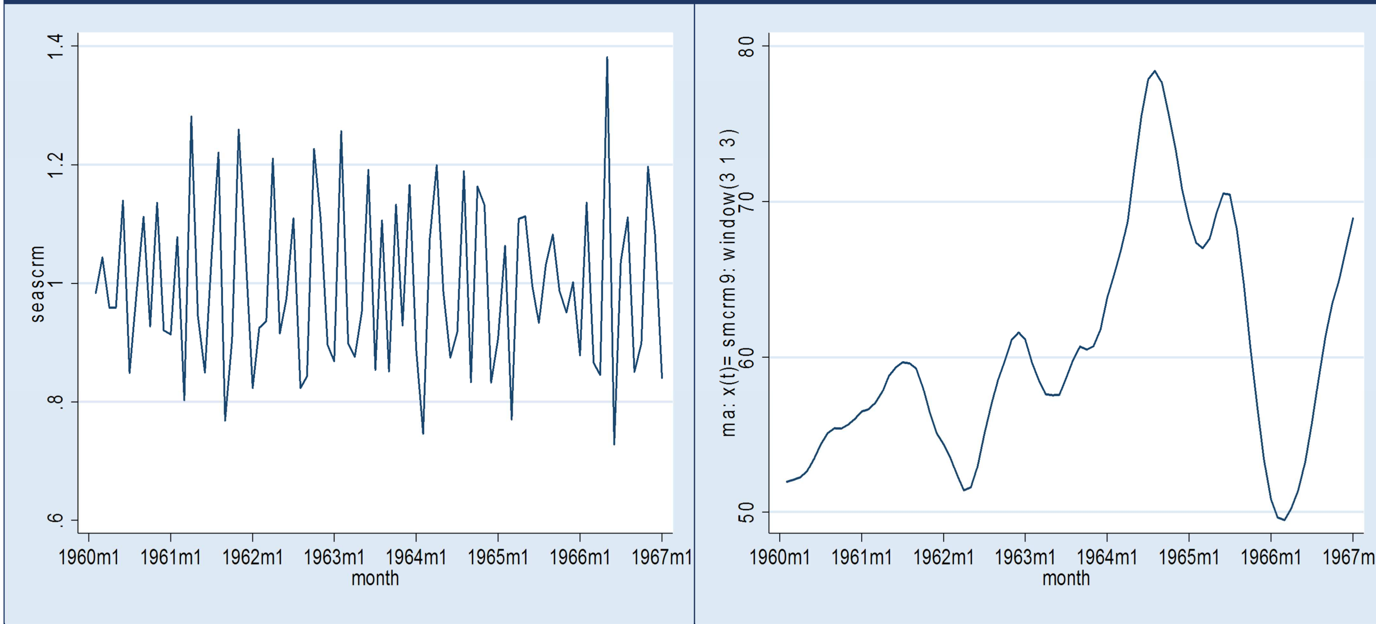
14. $S_t = \frac{vat_t}{M_{vat,t}} = \frac{T_t * S_t * \varepsilon}{M_{vat,t}}$

This can be used as the seasonal forecast for the next period. However, if in doubt about the assumption that the mean of ε, is 1, i.e., $E(\varepsilon) = 1$, it is easy to get the seasonal component alone just by taking another moving average of the result in the equation (11), choosing the values for F and L for that average after careful scrutiny of the graph of $S_t * \varepsilon$. The results of this move are portrayed in Panel B, Figure 5.

---

[9] Bowerman, B. L. and O'Connell, R.T. (1993). *Forecasting and Time Series: An Applied Approach,* 3rd ed. Pacific Grove, CA: Duxbury.

Figure 3-5: Graphs of Seasonal Effects

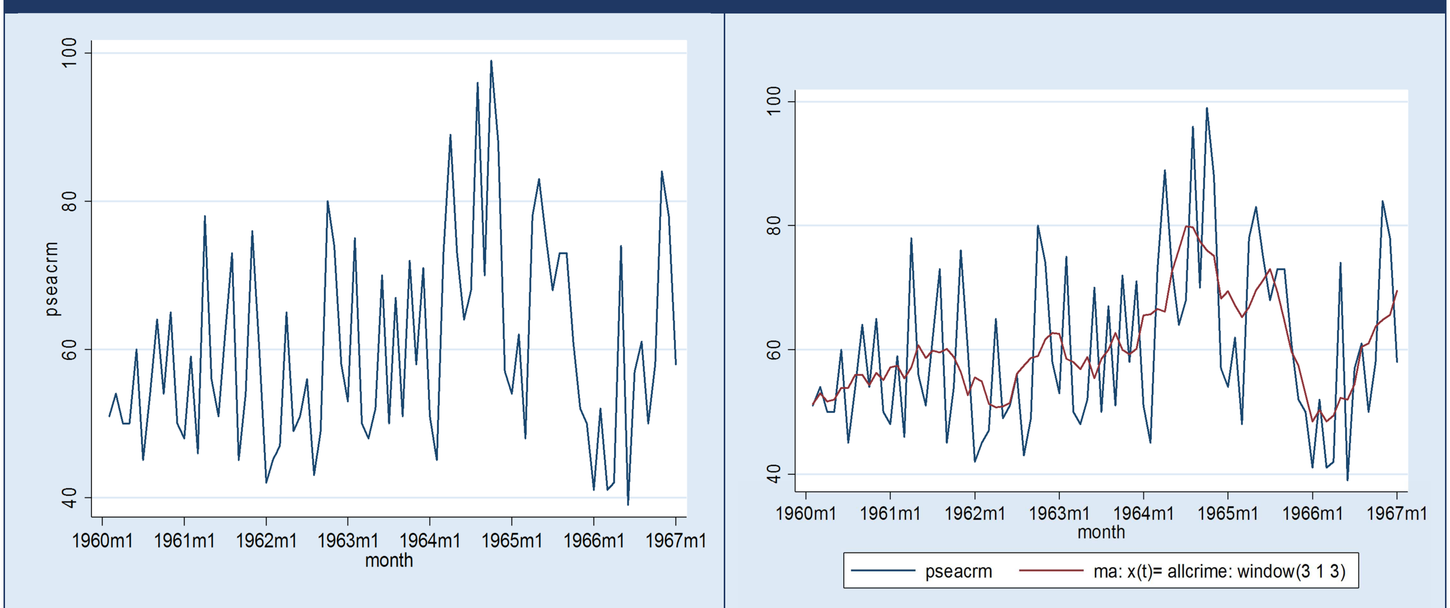


Panel A: Seasonal and Random Effects | **Panel B:** Seasonal Effects Isolated with Moving Average

Finally, using equation (11), the graph of vat_t with seasonal effects considered would be as portrayed in Panel A, Figure 6. Panel B, Figure 6 should also be graphed for a visual understanding of how the predicted seasonal effects compare to the trend-cycle. The forecast for $t + 1$ would be:

15. $vat_{t+1} = M_{vat,t} * S_t$

Figure 3-6: Trend and Seasonal Effects Combined



Moreover, the PER Team could take an average over all the series, M_{vat}, and then measure the size of each $M_{vat,t}$ relative to that average by computing:

16. $R(S_t) = \frac{M_{vat,t}}{M_{vat}}$

This type of calculation is very helpful in understanding what is happening to revenues during a budget year, and what is likely to happen in the next year. In this example, the PER Team would be able to assess the likely amount by which VAT would exceed the average in a given month, and should find that in some months the expected revenue flow is higher than in others.

3.2.1.6 Getting Additive Seasonal Effects - A Regression Method

If the seasonal effects are additive, then regression analysis can be used to isolate and analyze them. Much depends on the type of trend that is evident after graphing the data. Consider a quarterly revenue series, x_t, the graph of which exhibits a linear time trend. Let Q_i be a categorical or dummy variable for quarter i. So $Q_i = 1$ if quarter I and $Q_i = 0$ otherwise. For t an indicator of time, we could run the regression:

17. $\widehat{x_t} = \beta_0 + \beta_1 t + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \varepsilon_t$

With a constant term (β_0) in the equation, one of the Q_i must be omitted to avoid perfect collinearity with β_0. We have simply chosen Q_1 but this is a matter of choice. The coefficient β_1 measures the trend effects, which is to say the slope of the trend. The additive seasonal effects can now be isolated as follows:

1. For quarter 2, $Q_2 = 1$, so the average level in that quarter, *net of the time trend*, is $\beta_0 + \beta_2$.
2. For quarter 3, $Q_3 = 1$, so the average level in that quarter, *net of the time trend*, is $\beta_0 + \beta_3$.
3. For quarter 4, $Q_2 = 1$, so the average level in that quarter, *net of the time trend*, is $\beta_0 + \beta_4$.

The notion of 'net of the time trend' is to be understood as 'after controlling for the time trend'. So β_2 measures the difference between the average level of y_t in the first quarter and the average level of y_t in the second quarter, after controlling for the time trend, and so on for β_3 and β_4.

Of course, the regression could be run without a constant term and include all the dummies. However, the measures of the seasonal effects would be the same. What is more interesting is to use a reference period or a reference level in the model, because that helps with getting to the true meaning of 'controlling for the trend effects'. For example, let $\bar{t}$ be the mean time or the midpoint of the series. Then, we really should estimate the forecasting model as:

18. $\hat{x}_t = \beta_0 + \beta_1(t - \bar{t}) + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \varepsilon_t$

Then, the appropriate interpretation is that around the **midpoint** of the time series, the average value of y_t in Q_1 is β_0. Similarly, around the around the **midpoint** of the time series, the average value of value of y_t in Q_2 is $\beta_0 + \beta_2$, and so on. An alternative reference point can be used, such as the terminal time T, with suitable adaptation of the interpretation.

With Stata, all of this is easier than the algebra might suggest. In Figure 1, the monthly international visitor arrivals show evidence of additive seasonal effects. To estimate them with regression analysis, run a regression with a constant and eleven dummies, since the data is monthly. Stata provides a very powerful way of creating the dummies quickly with its **xi** command. The applicable code is:

```
>gen moy = month(dofm(month))
>summarize month, meanonly
>gen month2 = month - r(mean)
>xi: reg visits month2 i.moy
```

1. **The moy** variable that **generate**d records the month of the year – month 1 to month 12. The **dofm()** function converts the monthly variable **month** to a daily variable.
2. The second command generates the mean of month.
3. The third command produces the $(t - \bar{t})$ or month-meanmonth variable
4. The fourth command runs the regression with the dummies, telling Stata to create the dummies on the fly.

Table 1 is the type of Stata output you should expect:

Table 3-1: Stata output of Additive Seasonal Model

Source	SS	df	MS		
				Number of obs	= 84
				F(12, 71)	= 9.45
Model	6295.96845	12	524.664037	Prob > F	= 0.0000
Residual	3941.81712	71	55.518551	R-squared	= 0.6150
				Adj R-squared	= 0.5499
Total	10237.7856	83	123.346814	Root MSE	= 7.4511

iva	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
month2	-.2117039	.0338741	-6.25	0.000	-.279247	-.1441608
_Imoy_2	2.142686	4.000161	0.54	0.594	-5.833407	10.11878
_Imoy_3	6.132962	3.997148	1.53	0.129	-1.837124	14.10305
_Imoy_4	-1.871049	3.994419	-0.47	0.641	-9.835695	6.093598
_Imoy_5	-6.30506	3.991977	-1.58	0.119	-14.26484	1.654716
_Imoy_6	-11.90478	3.989821	-2.98	0.004	-19.86026	-3.949308
_Imoy_7	-14.62022	3.987951	-3.67	0.000	-22.57197	-6.668475
_Imoy_8	-5.141376	3.986368	-1.29	0.201	-13.08997	2.807215
_Imoy_9	-4.811101	3.985072	-1.21	0.231	-12.75711	3.134907
_Imoy_10	-13.8494	3.984065	-3.48	0.001	-21.7934	-5.905398
_Imoy_11	-12.64769	3.983344	-3.18	0.002	-20.59026	-4.70513
_Imoy_12	-12.51456	3.982912	-3.14	0.002	-20.45626	-4.57286
_cons	37.15009	2.822398	13.16	0.000	31.52238	42.77779

The results indicate that IVA really has two seasons in this country: October to February and June/July. The high point is January with about 3715 visitors and every other month in the seasons has significantly fewer visitors. For example, in June, the month of the mid-point in the second season, the number of visitors is expected to be 3715-1190 or about 2525 visitors. So, tax revenues from this tax base should be expected to behave similarly.

3.2.1.7 Unequally Weighted Moving Averages –Forecasting total taxes

When the weights assigned are not equal, the resulting forecast is an unequally-weighted moving average. Suppose the PER Team wants to use the moving average to forecast total taxes without adding up the individual tax forecasts. Then, the forecast is best computed with a moving average as

$$19.\ total\ taxes_{t+1} = M_{x,t} = \frac{\sum_{i=-L}^{F} w_i x_{t+i}}{\sum_{i=-L}^{F} w_i}$$

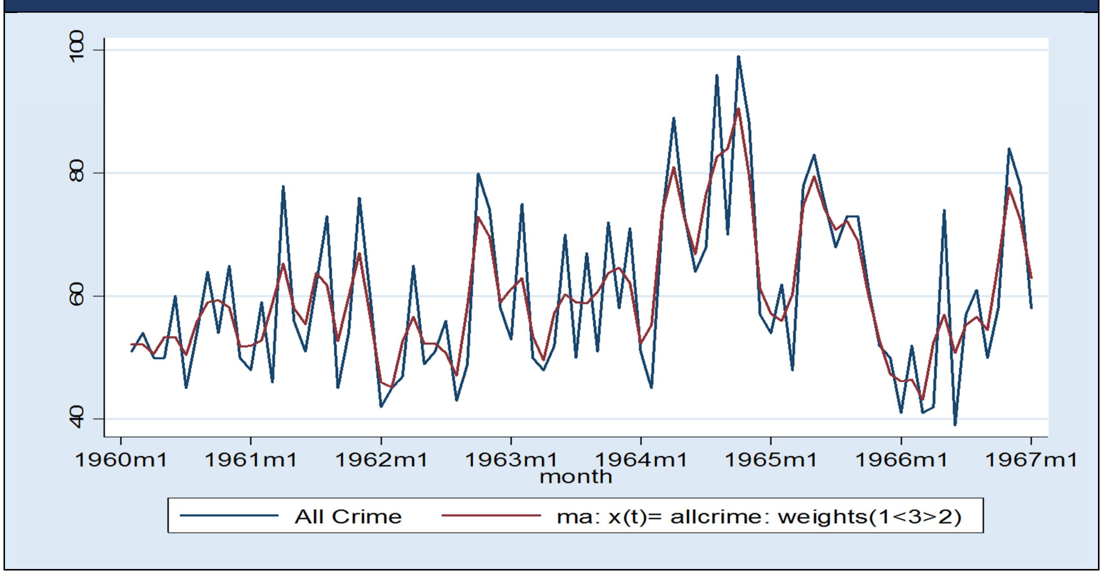
where $w_i, i = -L \ldots 0 \ldots F$ is the set of weights to be assigned to the past total taxes x_{t+i}. Suppose that available information indicated that in a 3-period moving average, the applicable weights should be 1 for x_{t-1}, 3 for x_t and 2 for x_{t+1}. Then, the definition yields:

$$20.\ total\ taxes_{t+1} = M_{x,t} = \frac{\sum_{i=-L}^{F} w_i x_{t+i}}{\sum_{i=-L}^{F} w_i} = \frac{1}{1+3+2}[x_{t-1} + 3x_t + 2x_{t+1}]$$

The famous Census-II X11 and X12 seasonal adjustment procedures use this method. This is the method of choice in computing moving averages of total taxes since total revenues vary seasonally during the year. For example, using the crime data in Figure 1, the weighted moving average graphed in Figure 7 can be generated with the following Stata code:

```
>tssmooth ma wmcrm3=allcrime, weights(1<3>2)
>tsline allcrime wmcrm3
```

Figure 3-7: Results of Weighted Moving Average



3.2.1.8 Exponential Smoothing

Exponential smoothing is a common forecasting technique widely used in the private sector. It is a corrected moving average of forecasts, with the correction being an adjustment for the error observed in the preceding forecasts. There are two types of exponential smoothing in the literature: simple exponential smoothing and double exponential smoothing.

Simple Exponential Smoothing

Simple exponential smoothing is a popular method that is suitable for smoothing a revenue series X_t that has no trend but has a mean**,** β_{0t}, that changes over time.[10] That is, we are considering the tax generating process as:

21. $X_t = \beta_{0t} + \varepsilon_t$

The simple exponential smoother produces smooth budget forecasts S_t using the definition:

22. $S_t = \alpha X_t + (1-\alpha)S_{t-1} = S_{t-1} + \alpha(X_t - S_{t-1})$

Here, α is called the smoothing parameter and takes values between 0 and 1. It is sufficient to observe that since α and $(1-\alpha)$ are weights (probabilities for example) that sum to 1, it must also hold that the smoother removes noise in the revenue series X_t. Part of the reason a method such as this is often treated as not 'rigorous' is that α has to be chosen by the analyst, for example after suitable dialogue among members of the PER Team. This makes it vulnerable to political influence. S_{t-1} can be treated as the mean value of the revenue series X_t from the initial period up to $t-1$. So, the simple exponential smoother takes a fraction $(1-\alpha)$ and uses it to update this mean, S_{t-1}, at each *t* by adding a fraction, α, of the revenues, X_t, collected at time *t*. What results is that as the actual revenues change over time, the smoother will also change, but at a slower pace. This helps to remove the noise in the tax data. S_t is the model's forecast of revenues X_{t+1}, if we have an initial value S_0 at time t_0. Similarly, S_{t-1} was its forecast of X_t.

To see the advantages of simple exponential smoothing over the simple moving averages discussed so far, consider a slight adjustment of the smoother that is widely used to forecast consumer spending. Assume that equation (22) forecasted revenues X_t with some error. So, it forecasted S_{t-1} but X_t was realized. Then, the forecast error is $X_t - S_{t-1}$. The next forecast should take account of this error. Suppose that the smoother is written as a forecast of X_{t+1} by giving the earlier forecast of X_t a weight of $(1-\alpha)$ rather than just α and then adjusting it by adding the α-weighted difference $X_t - S_{t-1}$ to compensate for the earlier error. The result is:

23. $S_t = \alpha(X_t - S_{t-1}) + (1-\alpha)S_{t-1}$

[10] So, by the definitions provided above, the series is not stationary.

This is actually the adaptive expectations model that has been widely used to forecast consumer spending based on expected income, when those expectations are not fully realized. This, in turn, is an application of the general partial adjustment model, so-called because only a fraction of the error is used to adjust the earlier forecast.

Now, to see where the idea of 'exponential' came from, observe that application of the definition in (22) to substitute for S_{t-1} gives

24. $S_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + (1-\alpha)(1-\alpha)S_{t-2}$

If this is done repeatedly for t periods, the revenue-generating process can be generalized to:

25. $S_t = (1-\alpha)^t S_0 + \alpha \sum_{k=0}^{k=t-1}(1-\alpha)^k X_{t-k}$

Thus, the simple exponential smoother is actually a special case of a weighted moving average, and in fact is an exponentially-weighted moving average because the weights applied to past observations decline geometrically. So, the method tends to give more weight to recent revenue data. Its advantage is that, rather than just using some recent revenue data arbitrarily selected, all past revenue data are incorporated into the current period's exponential moving average forecast.

Initializing and Choosing α

A good way to get S_0 is to use an average of the first set of observations in the revenue series X_t, for example the first half of the series. It is a matter of choice, which adds to doubt about rigor. To choose α, apart from just experimenting and seeing what gives a good fit to the revenue data, some statistical methods can help. This parameter determines the extent of smoothing and the general expectation is that as α approaches 1, the degree of smoothing falls. This happens because less weight is placed on the previous estimate of the mean of the unfiltered series and more weight is placed on the most recently observed value of the unfiltered series.

Some degree of smoothing is required to eliminate the noise in the data. One approach, generally used by most statistical software, is to choose α in order to minimize the 'root mean square error' of the forecasts – the square root of squared deviations between the forecasts and the realized values. To make forecasts beyond S_t, which is the forecast of X_{t+1}, S_t itself is used because X_{t+1} is not yet realized. This is a consequence of the absence of a trend in the data. Thus, a forecast for X_{t+2} applies the definition and gives:

26. $\widehat{X_{t+2}} = \alpha\widehat{X_{t+1}} + (1-\alpha)S_t = \alpha S_t + (1-\alpha)S_t = S_t$

The general Stata syntax for this simple exponential smoother is:

```
>tssmooth exponential [type] <newvar> = <exp> [if] [in],
                     [ parms(#) [ samp0(#) | s0(#) ] forecast(#) ]
```

Here is the Stata code used to generate the exponentially weighted moving average of the data in Figure 1:

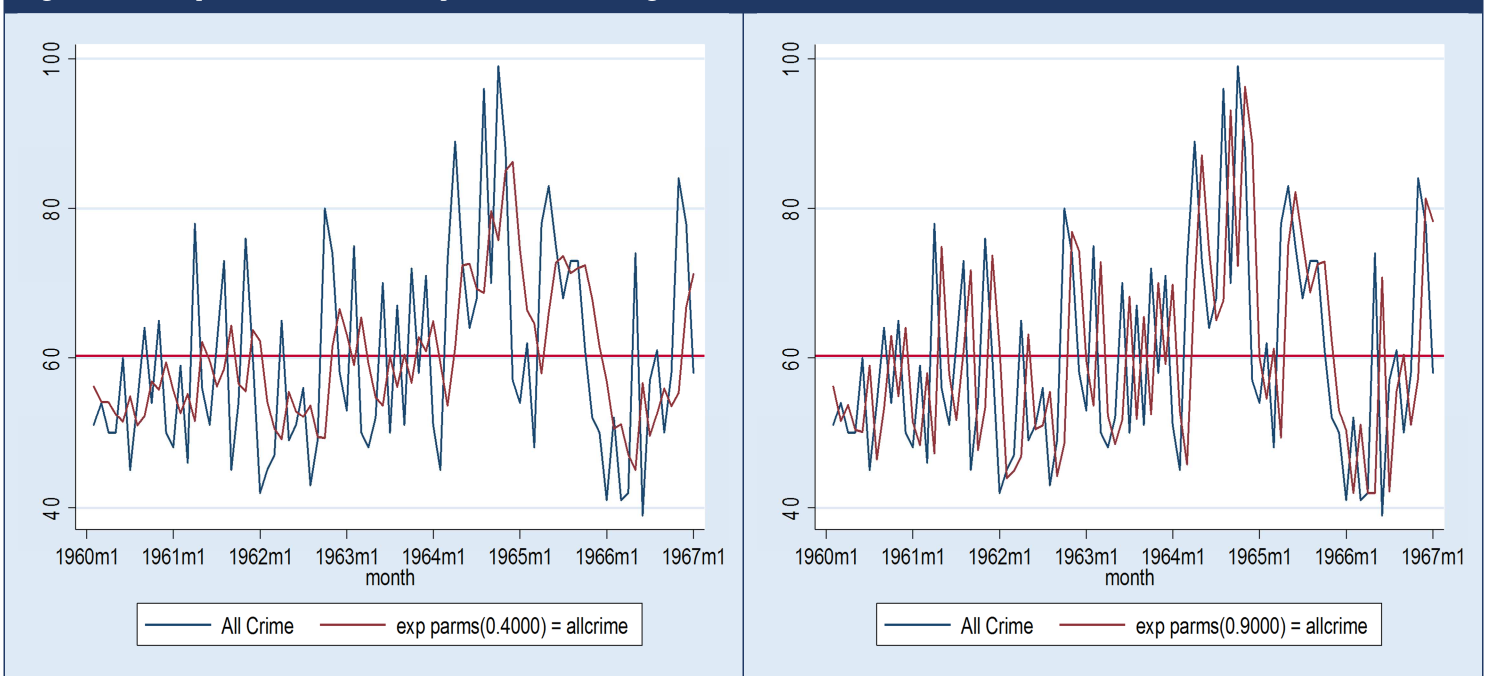
```
>tssmooth exp sallcri4=allcrime, parms(0.4)
>summarize allcrime
>tsline allcrime sallcri4, yline(`=r(mean)')
```

And,

```
>tssmooth exp sallcri9=allcrime, parms(0.9)
>summarize allcrime
>tsline allcrime sallcri9, yline(`=r(mean)')
```

The results are graphed in Figure 8, with the original series for comparison. Notice that the smoothing parameter 0.4 has smoothed the series (Panel A) much more than the parameter 0.9 (Panel B) and does not capture the movement of the series as well. This is because in Panel A the forecast for each period is adjusted by only 40% of the previous period's forecast error, as compared to 90% for the smoothing in Panel B. In the latter case, the forecast tracks the actual series very closely, with a one-period lag. The one-period lag occurs because the forecasted value for period t is the smoothed value for t-1, which is mostly just that period's actual value.

Figure 3-8: Comparison of results of exponential smoothing



Panel A: Smoothing Parameter 0.4 **Panel B:** Smoothing Parameter 0.9

Double Exponential Smoothing

Double exponential smoothing is what it says – application of the simple exponential smoother twice to the same series. It is also one way to take account of trends in the revenue series. It is appropriate for forecasting a revenue series when the graph of the series has a mean that evolves over time and *also has a time trend*. An important distinction must be made between a global trend and a local trend. If the trend moves ***in the same direction for all time*** periods, it is called *global*. In that case, the model of the revenue generating process is:

27. $X_t = \beta_{0t} + \beta_1 t + \varepsilon_t$

This equation is similar to the one used above to demonstrate how regression methods can be used to address additive seasonality. However, trends are not usually nicely behaved. If the trend changes direction in some periods, the distinct trends are called *local trends*. If each local trend is linear, we consider the revenue generating process as:

28. $X_t = \beta_{0t} + \beta_{1t} t + \varepsilon_t$

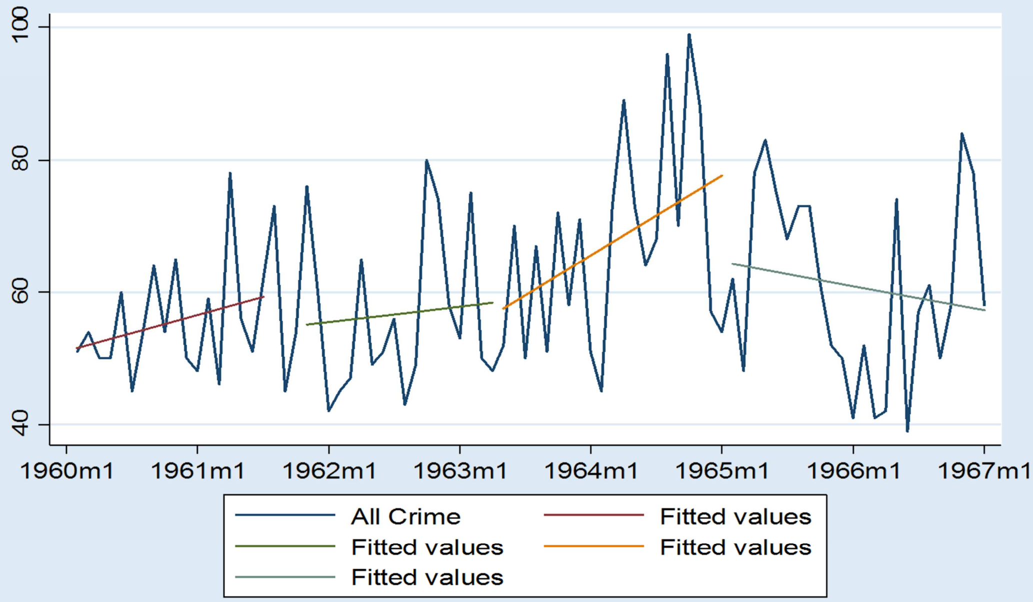
So, both the intercept (β_{0t}) and the slope (β_{1t}) evolve slowly with time, perhaps rising in some cases and falling in others. *Linear* local trends are special. More general mathematical forms tend to characterize local trends.

Practically, use the following Stata code to take a look at the graphs of the data in Panel A, Figure 1:

```
> tsline allcrime || lfit allcrime t in 1/18 || ///
                  lfit allcrime t in 22/39 || ///
                  lfit allcrime t in 40/60 || ///
            lfit allcrime t in 61/84
```

The output is reported in Figure 9. The distinct local linear trends are readily visible. From *t* = 1 to *t* = 18, there is an upward trend; from *t* = 22 to *t* = 39, there is an upward trend with a lower slope than in the previous period; from *t* = 40 to *t* = 60 there is the strongest upward trend in the set of blocks; and from t=61 to t=84, there is a downward trend.

Figure 3-9: Local Trends in Data (Panel A, Figure 1)



Operationally, it is appropriate to forecast revenue series with this type of behaviour by applying the simple exponential smoother twice to the series. A good reference for this method is Montgomery, Johnson, and Gardiner (1990, Ch.4; Appendix 4A)[11]. For the first round of smoothing, the applicable equation is, as before:

29. $S_t = \alpha X_t + (1-\alpha)S_{t-1}$

Thus, for the second round, the forecast equation is

30. $S_t^{(2)} = \alpha S_t + (1-\alpha)S_{t-1}^{(2)}$

Clearly, this forecast can be written as:

31. $S_t^{(2)} = \alpha(S_t - S_{t-1}^{(2)}) + S_{t-1}^{(2)}$

The term $S_t - S_{t-1}^{(2)}$ is measure of the *effect of the time trend* at time t. Thus, equation (31) essentially forecasts the effect that the time trend has on the revenue series X_t.

To isolate the trend doing all of this in Stata, use the following code:

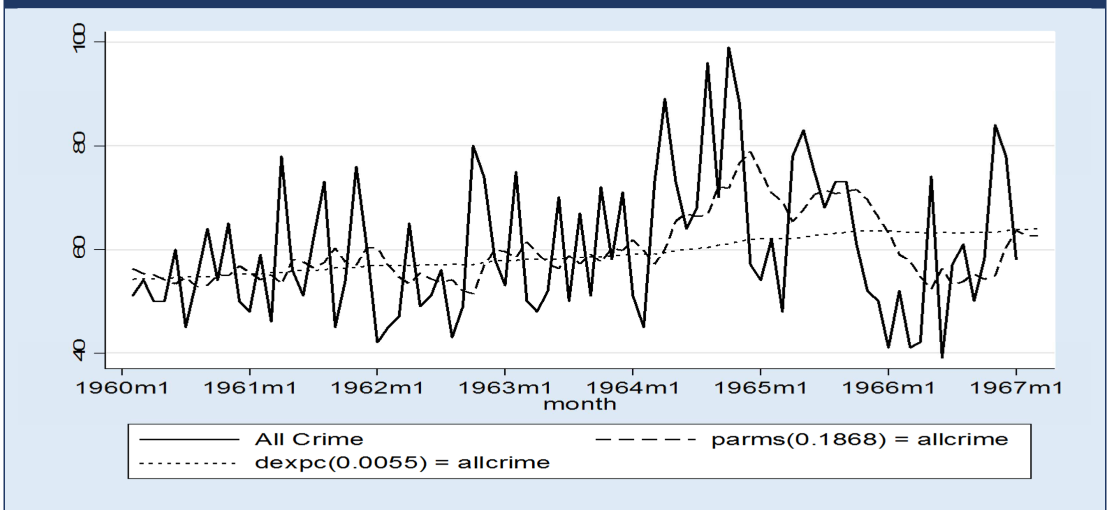
```
>tssmooth exp sallcri=allcrime, forecast(2)
>tssmooth dexp dsallcri=allcrime, forecast(2)
>tsline allcrime sallcri dsallcri, scheme(s2mono)
```

[11] Montgomery, D. C., Johnson, L.A., and Gardiner, J.S. (1990). *Forecasting and Time Series Analysis,* 2nd ed. New York: McGraw-Hill.

The difference from the earlier code is deliberate. The analyst could try different values for α and then pick the one that appears to best fit the data. However, we have left out the **parms()** option and let **tssmooth exponential** and **tssmooth dexponential** find the respective values of α that minimize the root mean squared error (RMSE) automatically for each method. That is, for each chosen α the method computes the sum of the squared deviations from the forecasted trend, and takes the square root. Then, it picks the one that gives the smallest RMSE. If you run the code with the data in Figure 1, you will find that the simple exponential smoother used $\alpha = 0.18$ and the double exponential smoother used $\alpha = 0.006$. The graphing code **scheme(s2mono)** forces the program to use dashed lines rather than colour to distinguish the various graphs. Figure 10 shows the comparative results generated, against the actual data.

The two smoothers have produced different fits and neither does a great job of fitting the data overall. The simple exponential smoother has a smaller RMSE (13.09 versus 13.26). The forecast made by the double-exponential smoother grows steadily over time, whereas the simple exponential smoother's forecast rises up to t=60 and then declines after that. The forecasts differ because the double exponential smoother includes a time trend, which here is predicted to be positive overall, whereas the simple exponential smoother does not attempt to account for the time trends.

Figure 3-10: Comparative results of simple and double exponential smoothing



3.2.1.9 Holt-Winters Forecasting

Like double-exponential smoothing, Holt-Winters smoothing can be used when forecasting a series that can be modelled as a time trend, with the constant term and the time-slope varying over time and exhibiting local trends. The methods are based on the work of Winters (1960)[12] and Holt (2004)[13].

[12] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science* 6: 324–342.

[13] Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* 20: 5–10.

Equations (29) and (30) use the same smoothing parameter, α. Holt-Winters forecasting allows α to change in equation (30), the second updating of the forecast. So, it is essentially a generalisation of the double exponential smoother with two smoothing parameters rather than one. It produces a budget forecast of the form:

32. $\widehat{X_{t+1}} = a_t + b_t t + \varepsilon_t$

where, $\widehat{X_t}$ is the forecast of the original series X_t using simple exponential smoothing and an optimal smoothing parameter, α; where both the intercept (a_t) and the slope (b_t) evolve slowly with time; and where the smoothing and forecasting with equation (31) uses a smoothing parameter coefficients that might differ from α.

The smoothing equations are therefore:

33. $S_t = \alpha X_t + (1-\alpha)S_{t-1}$

34. $S_t^{(2)} = \beta S_t + (1-\beta)S_{t-1}^{(2)}$

Now, consider a time series X_t that exhibits a time trend T_t, which describes how the series changes in general over time, and a local linear trend L_t, which describes the average value of the time series for time periods in the neighborhood of time t. Since L_t is a mean that drifts over time, it is nothing but the model-generated value of a_t in equation (32). And, since T_t is the general description of how the series changes over time, it can only be a description of the evolving slope b_t in equation (32). Once the smoothing parameters α and β are chosen optimally from equations (33) and (34), in particular to minimize the root mean-square-error, then Bowerman, O'Connell, and Koehler (2005)[14], and the classic work of Harvey (1989)[15], provide nice proofs that the updating or forecasting equations for the local and global trends are as Holt (2004) proposed:

35. $\widehat{X_{t+1}} = a_t + b_t$
36. $a_t = \alpha X_t + (1-\alpha)(a_{t-1} + b_{t-1})$
37. $b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$

Observe that these (recursive) equations take the form of simple exponential smoothers, except that $a_{t-1} + b_{t-1}$ is used in equation (36) instead of only a_{t-1}. The trend at $t-1$, i.e., b_{t-1}, indicates how much the time series can be expected rise (or fall depending on direction) between $t-1$ and t because of the trend effect. The term $a_{t-1} + b_{t-1}$ in equation (36) accounts for the total effects of both the trend and the local effects between $t-1$ and t. The change in the local level between $t-1$ and t, i.e., $a_t - a_{t-1}$, is used to update the forecast of the trend component at time t. It should already be clear that the updating equations for double-exponential smoothing are special cases of these equations, with $\alpha = \beta$.

[14] Bowerman, B. L., O'Connell, R.T., and Koehler, A.B. (2005). *Forecasting, Time Series, and Regression: An Applied Approach*. 4th ed. Pacific Grove, CA: Brooks/Cole.
[15] Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

To use the two equations (35 and 36), we are going to need initial values a_0 and b_0, much as we needed them for simple exponential smoothing. Because of the presence of a time trend, there is an intuitively attractive way to get them. Run a simple OLS regression of X_t on time, t, and, having generated the coefficients, set a_0 to the value of the estimated constant term of the regression and set b_0 as the estimated slope coefficient.

Then, after computing the series a_t and b_t with equations (35) and (36), it only remains to apply the forecasting component of equation (32) to get the forecast for X_{t+1}, called a one-step-ahead forecast, as:

38. $\widehat{X_{t+1}} = a_t + b_t$

In general, to get the h-step-ahead revenue forecast at time t, that is, the forecast for X_{t+h}, use:

39. $\widehat{X_{t+h}} = a_t + b_t h$

As usual, we use Stata's powerful routines to generate the forecasts. The general syntax to use is: The most important aspects of the Stata syntax are:

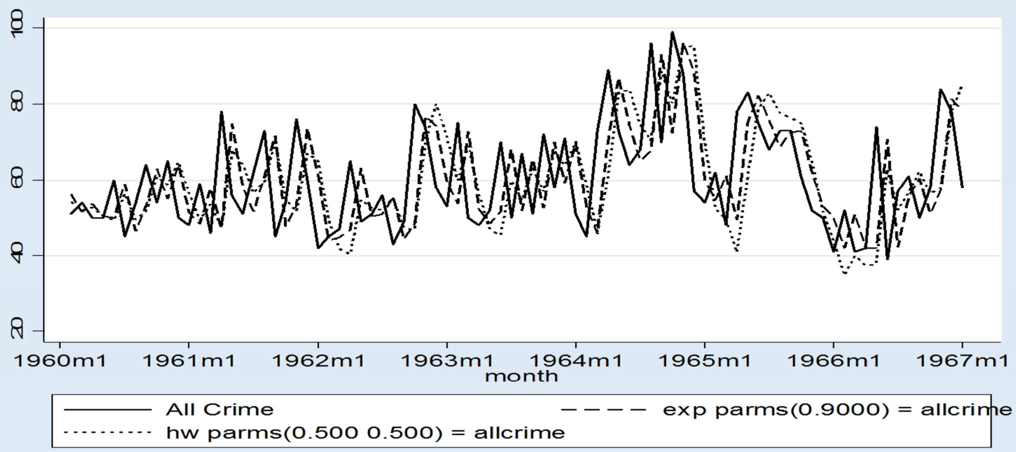
```
>tssmooth hwinters [type] <newvar> = <exp> [if] [in], [parms(#a #b) s0(#L #T)
forecast(#)]
```

As before, the analyst can supply parameters to the code using **parms(#a #b)** to specify the smoothing parameters α and β. The **forecast(#)** element in the code will produce forecasts for # periods beyond the end of the sample. The element **s0(#L #T)** allows the analyst to provide the starting values a_0 and b_0. Practically, the results in Figure 12 were generated with the following code:

```
>tssmooth hwinters hwalcrm = allcrime, parms(.5 .5)
>tsline allcrime sallcri9 hwalcrm,scheme(s2mono)
```

The Holt-Winters forecasts with **parms(.5 .5)** are mostly similar to those produced by the simple exponential smoother with **parms(.9)**, and the graph in Figure 11 demonstrates that.

Figure 3-11: Holt-Winters and Simple Exponential Smoothing Compared



3.2.1.10 Seasonal Effects with Holt-Winters

There are fairly straightforward ways to adjust the Holt-Winters Methods to cater for the seasonal effects evident in the revenue data, especially as that relates to their effects on the timing of budget releases. The earlier distinction between trend and seasonal effects should be recalled, along with the distinction between multiplicative and additive representations of the data.

Additive Seasonal Effects with Holt-Winters

If there is graphical justification for treating the seasonal effects as additive effects, then the methods used extend the above analysis quite easily. Let S_t be a variable that measures the seasonal effects at time t. The first thing to observe is that the additive method is applicable to data whose seasonal effects can forecasted with a model such as:

40. $\widehat{X_{t+j}} = a_t + b_t j + S_{t+j} + \varepsilon_{t+j}$

Comparing with equation (32), in addition to the non-seasonal a_t, the local level component, and b_t, the global time trend component, the forecasting model now has S_t, the seasonal effect.

Let τ define the period of the seasonal effects. The updating equations are adjusted accordingly to:

41. $a_t = \alpha(X_t - s_{t-\tau}) + (1-\alpha)(a_{t-1} + b_{t-1})$
42. $b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$
43. $S_t = \gamma(X_t - a_{t-1}) + (1-\gamma)S_{t-\tau}$

In this case, a_t represents the local level of revenues after removing any seasonal effects in the data. That is, when data on X_t becomes available, a_t is updated using X_t after subtracting the seasonal effects. The evident strategy is that since no estimates are available (*ex-post*) for the seasonal effect of period *t*, the seasonal effect for the same period one year ago, i.e., $s_{t-\tau}$, is used in its place. In the equation for S_t, the estimated seasonal effect is updated using $X_t - a_{t-1}$, which is to say the difference between the realized X_t and the deseasonalized local trend a_{t-1}. Again, seasonality is captured in the equation for S_t using $S_{t-\tau}$, which is to say the seasonal effect for the same period one year ago.

As before, after computing the series a_t, b_t and S_t with equations (41), (42) and (43), it only remains to apply equation (40) to get as the forecast for X_{t+j}, to get the j-step ahead at time t or if you like the forecast for X_{t+j} as:

44. $\widehat{X_{t+j}} = a_t + b_t j + S_{t+j-\tau}$

Note that the period of the seasonality is taken into account in the forecast through $S_{t+j-\tau}$.

Multiplicative Seasonal Effects with Holt-Winters

If there is graphical justification for treating the seasonal effects as multiplicative effects, with a slowly adjusting trend component, then one can represent the forecasting model as:

45. $\widehat{X_{t+j}} = (a_t + b_t j) * S_{t+j} + \varepsilon_{t+j}$

In generating the updating or current period forecasting equations for the local and global trends and the seasonal effects, the essential adjustment is to normalize rather than subtract in order to remove seasonal effects or local (deseasonalized) trend effects from the realized data. Thus, the corresponding updating equations are

46. $a_t = \alpha(\frac{X_t}{s_{t-\tau}}) + (1-\alpha)(a_{t-1} + b_{t-1})$
47. $b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$
48. $S_t = \gamma\left(\frac{X_t}{a_{t-1}}\right) + (1-\gamma)S_{t-\tau}$

It follows immediately that, after computing the series a_t, b_t and S_t with equations (46), (47) and (48), it only remains to apply equation (45) to get the j-step-ahead forecast at time t, or if you like the forecast for X_{t+j}, as:

49. $\widehat{X_{t+j}} = (a_t + b_t j) * S_{t+j-\tau}$

It is clear that incorporating seasonality increases the complexity of the forecasting model and the data requirements. Typically, about 4 years of monthly revenue data would be needed for its use. Global and local trends are estimated along with seasonal parameters. However, with available computer software, the challenge is not in the estimation itself. It is in the understanding of the method.

3.2.1.11 Box-Jenkins ARIMA Forecasting (Box and Jenkins, 1976)[16]

Box-Jenkins autoregressive integrated moving average forecasting (ARIMA) is included in this manual for completeness. ARIMA is one of the most widely used methods of forecasting. The main reason is that statistical confidence intervals can be constructed around its forecasts, so that the precision of the forecasts can be ascertained. Further, however, ARIMA model are very useful when it is necessary to understand how exogenous shocks in one period, t, influence revenue and expenditure outcomes in the future, $t+1$, $t+2$, and so on. There are three concepts in the method:

Autoregressive: This describes the component of the model that represents any correlation found between the value of *the revenue series* at time t and some linear combination of the values at times t - 2, t - 1, t + 1, and so on. If the correlations are found, the data must be transformed by differencing (and maybe by adding a trend) in order to use applicable statistics. Differencing involves subtracting X_{t-1} from X_t for all observations. In general, the autoregressive component forecasts the future value $\widehat{X_{t+1}}$ using p past observations X_{t-p} and the following forecast model of a process of order p:

50. $\widehat{X_t} = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p}$

The $\alpha_p s$ are the estimated parameters used in the forecasts.

Integrated: This refers to a summation process that translates the calculations of the model, which are based on differences, back to a level variable that can be interpreted intuitively. Suppose that on observation of the tax collections for the last month tax officials warn that collections for the next year will be substantially lower than previously forecast. This is a suggestion that the average around that time will be substantially lower than now and suggests non-stationarity by the tests suggested above. The forecaster must care about this, because it has a direct bearing on how the forecast has to be done. Nonstationary revenue series are much more difficult to analyse and forecast that stationary revenue series.

[16] Box, G.E.P and Jenkins, G.M. (1976). Time Series Analysis, Forecasting, and Control. San Francisco: Holden-Day.

If a revenue series is stationary, then all the standard methods of cross-sectional statistics, such as the central limit theorem and the consistency of estimators, generalize in straightforward ways to the analysis of the revenue series. However, if the revenue series is nonstationary, then the distributions of estimators are much more complex and are not simple generalizations of the familiar distributions of estimators for the stationary series. The most common approach to such non-stationary revenues, and the one adopted by ARIMA, is to first transform the series so that it is stationary. This is usually done by differencing (and perhaps by adding a trend). Then, to understand the calculations, it is necessary to recover the forecasted revenue in level form so that it can be interpreted. This is the purpose of 'integration' or 'summation' and the sum must exist if the method is to work.

Moving average: This refers the component of the forecasting model that is similar to those described in previous sections. That is, the moving average component of ARIMA forecasts based on previous forecasting errors. For ε_t the forecast error at time t, the moving average component of a forecasting model for process of order q is:

51. $\widehat{X_t} = \beta_0 + \beta_1\varepsilon_t + \beta_2\varepsilon_{t-1} + \cdots + \beta_p\varepsilon_{t-q}$

The autoregressive and the moving average components combine to form autoregressive moving average (ARMA) models. So, it should be clear that ARMA models assume a stationary data series ***before*** first differencing or the inclusion of a time trend. If differencing and the addition of a trend is necessary, then the issues of integration and summability arises and the forecast models becomes the Box-Jenkins ARIMA model.

The number of autoregressive and moving average lags in an ARIMA model is referred to as ARIMA(*p*,*d*,*q*), for p the length of the autoregressive lags, d the number of times the data must be differenced in order to make the series stationary, and q the number of lags in the moving average component. So, if d=0, then the ARIMA forecasting model is an ARMA(*p*,*q*) forecasting model.

If high-frequency revenue data are available, such as monthly or quarterly data, then the forecasting model should take account of seasonality. The method can address seasonality by taking into account the autoregressive or moving average trends that occur at different points in time. In the literature, the language that will be encountered when incorporating seasonality into the ARIMA model is expressed as ARIMA(*p*,*d*,*q*)(*P*,*Q*), for *P the number of seasonal autoregressive lags* and *Q the number of seasonal moving average lags*.

There are three steps in Box-Jenkins ARIMA forecasting:

> **Step 1: *Model identification*.** Here, the revenue forecaster must decide whether the time series is described by an autoregressive process, a moving average process, or both. Graphical representation of the data or several statistical methods are used for this purpose.
>
> **Step 2: *Model estimation and diagnostic checks*.** Here, the forecaster must check that the specified model is correctly identified. There are a wide variety of diagnostic tests used for this purpose.
>
> **Step 3: *Forecasting*.** If the model passes the diagnostic tests, the forecaster then forecast the revenues. The forecasts are produced along with confidence intervals - a measure of their accuracy

The confidence intervals are also a check on the validity of the revenue model and the usefulness of its forecasts. If a forecast is generated that is of dubious value in budgeting, but it is predicted with a high degree of confidence, then the whole exercise should be redone.

3.2.1.12 Assumptions of Box-Jenkins ARIMA

To forecast revenues with ARIMA, two assumptions must be satisfied.

1. The data series must be at least 50 observations long (Newton, 1988).[17] For the partner countries, this is a major obstacle to the use of the method. Most countries report certain key tax and expenditure data on an annual basis and most of the series available date to no earlier than 1970.
2. The data series must be stationary. As indicated above, this means that the revenue series must have a constant mean and constant variance. If this condition is not met, the forecasts can be spurious.

If the series is too short, then the known statistical methods do not apply and the forecasts become unreliable. If the series is non-stationary, the data series needs differencing. It may also be necessary to add a time trend when doing the forecasting. A linear trend can be added to make the series stationary if the non-stationarity in the series is characterized by a mean and variance that change by a constant amount over time. If instead the data is first-difference non-stationary, then first differencing of the data will render the series stationary. These decisions are made during the model identification process. Statistical software are now available to make the decisions straightforward. The constant-variance component of the second assumption of ARIMA models is referred to as *homoscedasticity.* If the variance around the mean is great even after differencing, the assumption fails and the revenue series is heteroscedastic. Several remedies exist for this problem, including data transformation using natural logarithms, square roots, or cubed roots, and the like.

---

[17] Newton, H.J. (1988). *Timeslab: A Time Series Analysis Laboratory*. Pacific Grove, CA: Wadsworth & Brooks/Cole Publishing.

3.2.1.13 Software and Simplicity

It has been emphasized that powerful modern statistical software, such as Stata 14, are available to implement these methods, so we have focused on clarification of the ideas involved. It should also be emphasized that all the methods described so far are simple special cases of more modern time series methods described in the references cited, such as the ARIMA representations of data and related analytical methods. The general lesson from the methods so far is that once the parameters used in a representation of the data can be estimated reliably, the PER Team can proceed to make forecasts. With their greater demand for detail, these modern methods are also far more flexible that the ones mentioned so far, and can be far more accurate forecasters that the exponential smoothers mentioned above.

So why emphasize these simple forecasting methods in the manual? We illustrate the answer by reference to the aforementioned ARIMA methods. A PER Team from the member countries of the project will tend to find that ARIMA methods are data hungry – large-sample- methods. Caribbean countries have major challenges on this front. Just as important, application of ARIMA methods give very accurate forecasts if, given large enough samples, the ARIMA model is well-specified. However, even relatively modest misspecification of an ARIMA model will yield very inaccurate forecasts because the number of parameters that must be correctly estimated is quite large. Exponential smoothing models estimate only a few parameters, which introduce less risk of error and can therefore be adequate for the needs of a PER Team, even though they are not as accurate as a well-specified ARIMA model. They also require significantly less time and much less exploratory analysis to get right than do ARIMA models. For this reason, single variable and single equation methods are the backbone of revenue forecasting as well as general forecasting of macroeconomic variables that are exogenous in a system-wide macroeconomic model (Annex 2).

When doing a PER, quite a few single equation forecasts of time series, sometimes hundreds of them, are needed to understand the trajectories of the system and its performance. This is especially true when dealing with many sectors and many interests and stakeholders, such as in education, health and social welfare. Use of ARIMA to forecast a few variables might be manageable, but this can become quite demanding on the Team’s time if hundreds of series must be forecasted, leaving the use of the exponential smoothers and the only viable strategy. This type of reliance on simplicity is reinforced in the remainder of the Manual.

3.2.1.14 Causal Models

Equation (17) above is an example of a member of a wide class of causal regression models employing a time trend that is widely used for aggregate revenue and expenditure forecasting. These are especially important for forecasting the aggregate taxes. This aggregate could be used in forecasting the debt to GDP ratio in Equation (6) of the illustrative macroeconometric model of Annex 2, incorporating the effective consumption capacity. An aggregate tax forecast can be used to set the overall size and structure of revenues on which the budget will be based.

It provides policy makers and fiscal planners with estimates of the gross resources that are likely to be available from the tax base, the extent of borrowing that will be needed to support any given level of expenditure, the necessary use of accumulated reserves, or the need for monetary policy measures to balance the budget. It also informs about what fiscal actions are sustainable and hence how to balance fiscal policy to address the problems in the balance of payments and hence foreign debt.

In addition to allowing the analysis of data with additive seasonality, the class of forecasting models tend to work well for revenues that are heavily influenced by macro and sector-level economic factors, such as business license fees, income taxes, and VAT, and the like. The macroeconomic factors most commonly used to explain such taxes are population, income, and price.

3.2.1.15 Using Time Trends as Proxy for Technical Change in Revenue Forecasting

First, consider the generalization of the forecast equation (17) by using the time trend as a proxy for technical progress. In partner countries, most of the technological change that affect taxes is truly exogenous, generated by the rest of the world, so the result will also have a strong causal interpretation. The wide class employing a time trend to forecast aggregate taxes is:

52. $\hat{x}_t = \beta_0 + \beta_1 t$
53. $\hat{x}_t = \beta_0 + \beta_1 t + \beta_2 t^2$
54. $\hat{x}_t = \beta_0 + \beta_1 t + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4$
55. $\hat{x}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 Q_2 + \beta_4 Q_3 + \beta_5 Q_4$

Here, β_1 is the parameter of the linear trend. In equations (53) and (55), β_2 the parameter of the quadratic trend, a very plausible option since technical progress is usually expected to be logistic and the quadratic picks up elements of that process. The estimated values of the parameters of the quarterly dummy variables would reveal the average level of the dependent variable during the quarter. Testing the equality of the dummy coefficients can reveal whether there are significant differences in the average level of the revenues across the seasons. All of these models can estimated with ordinary least squares regression (OLS). The forecasts can be done independently and then incorporated into Equation (6) of Annex 2.

Equation (54) has already been estimated above with the data in Panel B, Figure 1. The Stata codes and results for equation (55) are provided below.

```
>g mnthsq=month^2
>xi:reg iva month2 mnthsq i.moy
```

The first line of code creates the variable t^2 from the months listed in sequence. The second line includes this new variable in the estimation model along with t as before. Table 2 shows that the addition of t^2 seems to have improved the representation of the trend and seasonal effects better than equation (55) as reported in Table 1. First, the adjusted R^2 is better at 76% of all the variation in the data. Second, the representation now picks up the importance of March as a month when international visitor arrivals increase. Moreover, March is now picked up as the month with the highest tourism arrivals, essentially linked to the Carnival season. Second, it also picks up the fact that the second tourism season runs from May to July, rather than simply June and July.

Table 3-2: Stata-generated Results with Diagnostics for Equation (51)						

Source	SS	df	MS			
Model	8185.31904	13	629.639926		Number of obs =	84
Residual	2052.46652	70	29.3209503		F(13, 70) =	21.47
					Prob > F =	0.0000
Total	10237.7856	83	123.346814		R-squared =	0.7995
					Adj R-squared =	0.7623
					Root MSE =	5.4149

iva	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
month2	-.9786318	.098661	-9.92	0.000	-1.175405	-.7818587
mnthsq	.0090227	.001124	8.03	0.000	.0067809	.0112644
_Imoy_2	2.142686	2.907016	0.74	0.464	-3.655174	7.940547
_Imoy_3	6.223188	2.904848	2.14	0.036	.429652	12.01672
_Imoy_4	-1.70864	2.902914	-0.59	0.558	-7.49832	4.081039
_Imoy_5	-6.088515	2.901194	-2.10	0.039	-11.87476	-.3022664
_Imoy_6	-11.65215	2.899672	-4.02	0.000	-17.43536	-5.868936
_Imoy_7	-14.34954	2.898339	-4.95	0.000	-20.1301	-8.568988
_Imoy_8	-4.870696	2.897189	-1.68	0.097	-10.64896	.9075644
_Imoy_9	-4.558466	2.896222	-1.57	0.120	-10.3348	1.217866
_Imoy_10	-13.63285	2.895444	-4.71	0.000	-19.40763	-7.858072
_Imoy_11	-12.48529	2.894866	-4.31	0.000	-18.25891	-6.711658
_Imoy_12	-12.42433	2.894503	-4.29	0.000	-18.19724	-6.65143
_cons	15.38287	3.400023	4.52	0.000	8.601735	22.164

3.2.1.16 GDP-based Forecasting of Aggregate Tax Revenue

The main flaw in the approach through time and exogenous technology is that the size and structure of the tax base have not been considered. Policy must be concerned with all three variables. Generally, one would expect to find a close relationship between taxes and their bases in a revenue forecast. For example, the amount of income tax should depend on the amount of taxable income and the tax rate, with the taxable income dependent on the GDP or its components. Thus, instead of using a time-trend, the GDP or its components can serve as the independent variable in a gross revenue model, since the GDP and its components comprise the tax base.

If a sufficiently long time series data are available, then for Y the aggregate value-added and X the total taxes, the aggregate form of the forecast model of gross taxes is simply:

56. $\widehat{x_t} = \beta_0 + \beta_1 y_t$

Here both variables have been deflated and are in natural logarithms. Then, the estimated parameter β_1 is also a measure of the *elasticity* of response of total taxes to changes in gross value-added. Elasticity measures the responsiveness of budget variable X_t to changes in Y_t. For d the change (or differential) operator, it is normally defined as: $\varepsilon_{X,Y} = \frac{dx_t}{dy_t}$. If $dx_t = dy_t$, so that the budget category changes at the same rate as the underlying variable Y_t, then elasticity is equal to 1. If $dx_t > dy_t$, then the elasticity will be higher than 1 and if $dx_t < dy_t$, then the elasticity will be lower than 1. Elasticities other than 1 reflect sustained **structural changes** within either the budget category or the underlying variable.

In addition to the availability of time series data, this approach using elasticity (equation 56) must take into account two characteristics that are typical of the stationary x_t with y_t data used: (i) non-stationarity, and (ii) cointegration between gross revenues and the GDP. Nonstationary processes have been characterized above. Forecasting with non-stationary x_t with y_t might work just because the two variables grow over time and are highly correlated; the more highly correlated the better. However, the forecast equation cannot be used to explain *why* it is good policy to grow y_t as a way to grow x_t. Such a use in the budget process requires that x_t and y_t be cointegrated. Two variables are said to be cointegrated when the following conditions apply:

a. The two variables are non-stationary in the specific sense of being some type of random walk as described by equations (5), (7) or (8) above. This means that each *variable* must have a stochastic trend and therefore can be made stationary by first differencing. Such variables are integrated of order 1, I(1), as described in the above section on Box-Jenkins methods.

b. The coefficients β_0 and β_1 can be identified such that $e_{x,t}$ is stationary in the regression (equation 57):

57. $\widehat{x_t} = \beta_0 + \beta_1 y_t + e_{x,t}$

Relative to equation (1), equation (56) and equation (57) make the strong assumption that no discretionary changes have been made in the tax rate or tax base. Since discretionary changes are normally made, two options must be considered. One is to adjust the above forecast model to control for the effects of discretionary changes in the tax rate. Retaining the seasonal effects, let D_d be a categorical variable for the period (say years) when the relevant discretionary change in the tax rate or the tax base is in effect. This dummy variable is represented by the value of "0" for the years before the policy and a value of "1" for the years of after the policy comes into effect and until it expires. For example, a tax amnesty to encourage payment of taxes should be included as a categorical variable for the period of the amnesty, say 2 years. The dummy variable would be given a value of 1 for each of the two years. Then, the necessary forecast model should be:

58. $\widehat{x_t} = \beta_0 + \beta_1 y + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \cdots + \beta_d D_d + \beta_{dy} D_d * y$

The income elasticity of the tax policy is then $\beta_1 + \beta_{dy}$. This option can work if sufficiently long time series data are available, so that the coefficients can be reliably estimated.

For an example, we use the series in Figure 1. We suspect that crime has something to do with international tourism arrivals. We are also interested in the elasticities, so both crime and international arrivals are transformed into their logarithms. Then, we use Stata to check. Run the following code first:

```
>xi:reg liva i.moy*lcrime
```

The code would include all the dummy months and all their interactions with **lcrime**. If none of the interactions has a non-zero coefficient, then it is better to run:

```
>xi:reg liva lcrime i.moy
```

The result is reported in Table 3. The elasticity of IVA to crime is -0.25. That is, every 1% increase in crime leads to a 0.25% decrease in IVA.

Table 3-3: Illustration of an explanatory regression model

Source	SS	df	MS
Model	4.28573908	12	.357144923
Residual	3.85592213	71	.054308762
Total	8.14166121	83	.098092304

Number of obs	84
F(12, 71)	6.58
Prob > F	0.0000
R-squared	0.5264
Adj R-squared	0.4464
Root MSE	.23304

liva	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lcrime	-.2494314	.1333408	-1.87	0.066	-.5153054	.0164426
_Imoy_2	.1031455	.1253712	0.82	0.413	-.1468376	.3531287
_Imoy_3	.195576	.1246141	1.57	0.121	-.0528976	.4440495
_Imoy_4	.0238481	.1283472	0.19	0.853	-.2320689	.2797652
_Imoy_5	-.1029395	.1278864	-0.80	0.424	-.3579377	.1520586
_Imoy_6	-.3537371	.1262304	-2.80	0.007	-.6054334	-.1020409
_Imoy_7	-.4502495	.1262515	-3.57	0.001	-.7019878	-.1985112
_Imoy_8	-.0644664	.1299163	-0.50	0.621	-.3235122	.1945793
_Imoy_9	-.0877231	.1259187	-0.70	0.488	-.3387978	.1633517
_Imoy_10	-.3789734	.1310024	-2.89	0.005	-.6401848	-.1177619
_Imoy_11	-.321834	.1330555	-2.42	0.018	-.5871393	-.0565288
_Imoy_12	-.371671	.1272601	-2.92	0.005	-.6254205	-.1179216
_cons	4.544897	.5269474	8.62	0.000	3.494193	5.5956

Finally, regarding the aggregate tax forecasts, at least one exogenous macro variable, such as the logarithm of population size (p_{op}), should be added directly to all the aggregate tax forecast equations above. For example, equation (58) should be modified to:

59. $\widehat{x_{t+1}} = \beta_0 + \beta_1 y_{t+1} + \beta_2 p_{op,t+1} + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \cdots + \beta_d D_d + \beta_{dy} D_d * y$

Here too, we are working with a special case of equation (1). Thus, in building a forecast of aggregate revenues with any of these GDP-based forecast equations, it would be necessary to forecast the GDP or the GDP growth rate at time $t+1$. In the case of equation (58) it would also be necessary to forecast the population at time $t+1$. The population can be forecasted using any of the methods presented in Section 4 above.

When long enough time series data are not available, one approach is to forecast these variables independently, without using a macroeconometric model featuring multiple dimensions. Independent forecasts work well (Auerbach, 1999; Koirala, 2014).[18] For independent forecasts of the GDP and GDP growth rate, uninfluenced by the path of other variables in the economy, the required approach is as follows:

60. $y_{t+1} = \ln \widehat{Y_{t+1}}$
61. $\widehat{Y_{t+1}} = Y_t + \frac{(\bar{\bar{g}}_y Y_t)}{100}$
62. $\bar{\bar{g}}_Y = \frac{1}{n}\sum_{t=1}^{n} \frac{(Y_t - Y_{t-1})}{Y_{t-1}} * 100$

The process is recursive. To get the forecast of y_{t+1}:

1. First, forecast the GDP growth rate as the average over the selected number of years, and convert to a percentage.
2. Then, multiply the forecasted percentage growth rate and the GDP for year t, and reconvert to a proportion.
3. Next, add the result to the GDP of year t.
4. Then, take the natural logarithm of the forecast and apply the result to get the forecast of aggregate revenues.

Note that $\bar{\bar{g}}_y$ is just an overall average covering the entire data period arbitrarily chosen, say with $n = 5$, so the time series does not have to be long. However, the average growth rate can also be independently forecasted using any of the methods specified in **Section 4**. Use of equation (62) assumes that discretionary policy has zero impact in equation (1). In these cases, the influence of discretionary policy can be added by rewriting equation (62) as:

63. $\bar{\bar{g}}_Y = \frac{1}{n}\sum_{t=1}^{n}[\frac{(Y_t - DP_{t-1})}{Y_{t-1}} - 1] * 100$

[18] Auerbach, A J. (1999). *"On the Performance and Use of Government Revenue Forecasts."* University of California, Berkeley and NBER, USA. Koirala, T.P. (2014). *Government Revenue Forecasting in Nepal.* NRB Economic Review, Nepal Rastra Bank.

Instead of an independent forecast, the forecast of Y_{t+1} can be done with a macroeconometric model. The best approach is to update and use the model specially developed for preparation of the budget to reflect the role of effective consumption capacity. The next best alternative is to use a model such as that recommended in Annex 2. The output of Equation (9) of this model is already in the form of y_t, so a one-step-ahead forecast with the estimated equation will yield y_{t+1}. As illustrated by the model, macroeconometric forecasts are structured similar to the regression equations set out above, but can include estimates of influential and consequential change of other key variables shaping the budget: the debt to GDP ratio, the unemployment rate, the CPI, per capita income, and the effective consumption capacity per dollar of imports. More complex events and relationships can be considered and dynamic feedback can be incorporated using the output from one equation to inform adjustments in another equation. The types of revenue for which macroeconometric forecasts are most useful include those for which economy-wide elasticities can be computed: (i) corporate tax, (ii) personal income tax, (iii) real estate taxes, (iv) value-added or sales taxes, and (v) user fees and charges.

It is important to emphasize here that qualitative judgement is an integral part of forecasting with macroeconometric models. Macroeconomic forecasts are important in framing forward-looking medium-term macroeconomic policy and the budget. The model used must capture the salient features of the economy in a tractable framework that takes account of data availability. It must also take account of key external influences on the economy, particularly continual structural change and technological advances abroad and general inherent randomness. Judgement necessarily plays an important role in the preparation of forecasts with such variables and hence with models incorporating them. The use of a macroeconomic model for forecasting is illustrated here, using Stata code. The language of macroeconometric forecasting is as follows:

1. A macroeconometric forecasting model is a system of equations whose interactions jointly determine the outcomes of one or more *endogenous* variables. The term endogenous variables contrasts with exogenous variables, the values of which are not determined by the interactions of the system's equations.

2. A macroeconometric forecasting model has
   i. One or more stochastic equations that describe the behaviour of the endogenous variables.
   ii. Zero or more non-stochastic equations (identities) that usually describe the behaviour of endogenous variables that are based on accounting identities or summation conditions.
   iii. Zero or more exogenous variables that must be declared.
   iv. Equations that are identified, in the sense that there is enough information to estimate the structural equations that have been specified.
   v. Stationary residuals

3. A macroeconometric forecast model can produce forecasts under alternative scenarios.

4. 'Solve the model' and 'forecast with the model' mean the same thing.

We have concentrated on revenues so far, so we choose to illustrate forecasting of government spending with a macroeconometric model. The government expenditure forecast is needed to set the overall budget ceiling, taking into account the development path of the economy. In such a model, the share of government spending in GDP must be treated as endogenous. Substantively, it should depend on the success of the economy in solving its fundamental problem of growing the effective consumption capacity per dollar of foreign exchange, on how the sectors are restructuring, and on how the technology of the economy is evolving. The success of the economy in solving its development challenge depends partly on how the technology of the economy is adjusting relative to that of its main trading partner, here assumed to be the US economy. Inflation is not merely a by-product in this process. It influences the rate of development of the technology of the economy. Here is a dynamic statistical forecasting model that can be used with the GDP forecast equations (60-62) to forecast aggregate government expenditure:

64. $lgovgdp = \alpha_0 + \alpha_1 lpk + \alpha_2 lothmin + \alpha_3 lpk_{t-1} + \alpha_4 lexppop_{t-2} + \alpha_5 lapcon_{t-1} + \alpha_6 lgovgdp_{t-1} + \alpha_7 lklratio_{t-1} + e_1$
65. $lapcon = \beta_0 + \beta_1 lresus + \beta_2 leximp + \beta_3 ldevgap2 + \beta_4 lexchr + \beta_5 ltbusa + e_2$
66. $lcpi = \gamma_0 + \gamma_1 lm2 + \gamma_2 lgdpdef + \gamma_3 lsavr_{t-1} + \gamma_4 lapcon + \gamma_5 lexchr + \gamma_6 ltbusa + e_3$
67. $lklratio = \theta_0 + \theta_1 t + \theta_2 tech2 + \theta_3 tech3 + \theta_4 lcpi + \theta_5 ltbusa + \theta_6 ltbill + \theta_6 ltbusa_{t-1} + \theta_8 lgov_{t-1} + \theta_9 lsavr + \theta_{10} lsavr_{t-1} + e_4$

The model has four interacting equations, The variables are all in logarithms to allow response elasticities to be measured:

$lgovgdp$ – the logarithm of the ratio of government spending to GDP
lpk – the logarithm of the investment deflator, which is a proxy for the price of capital
$lothmin$ - the logarithm of the ratio of other economic activity to the total of mining and manufacturing activity
$lexppop$ - the logarithm of exports per capita
$lapcon$ - the logarithm of effective consumption per dollar of imports
$lklratio$ - the logarithm of the capital per capita, which is a proxy for the capita-labour ratio
$lresus$ - the logarithm of the ratio of official reserves to government spending
$ldevgap2$- the logarithm of the ratio of the domestic capital-labour ratio to the capital-labour ratio of the USA
$leximp$- the logarithm of the ratio of exports to imports
$lexchr$ - the logarithm of the exchange rate
$ltbusa$ - the logarithm of the ratio of US Treasury bill rate to local treasury bill rate
$lm2$- the logarithm of the money supply
$lgdpdef$- the logarithm of the GDP deflator

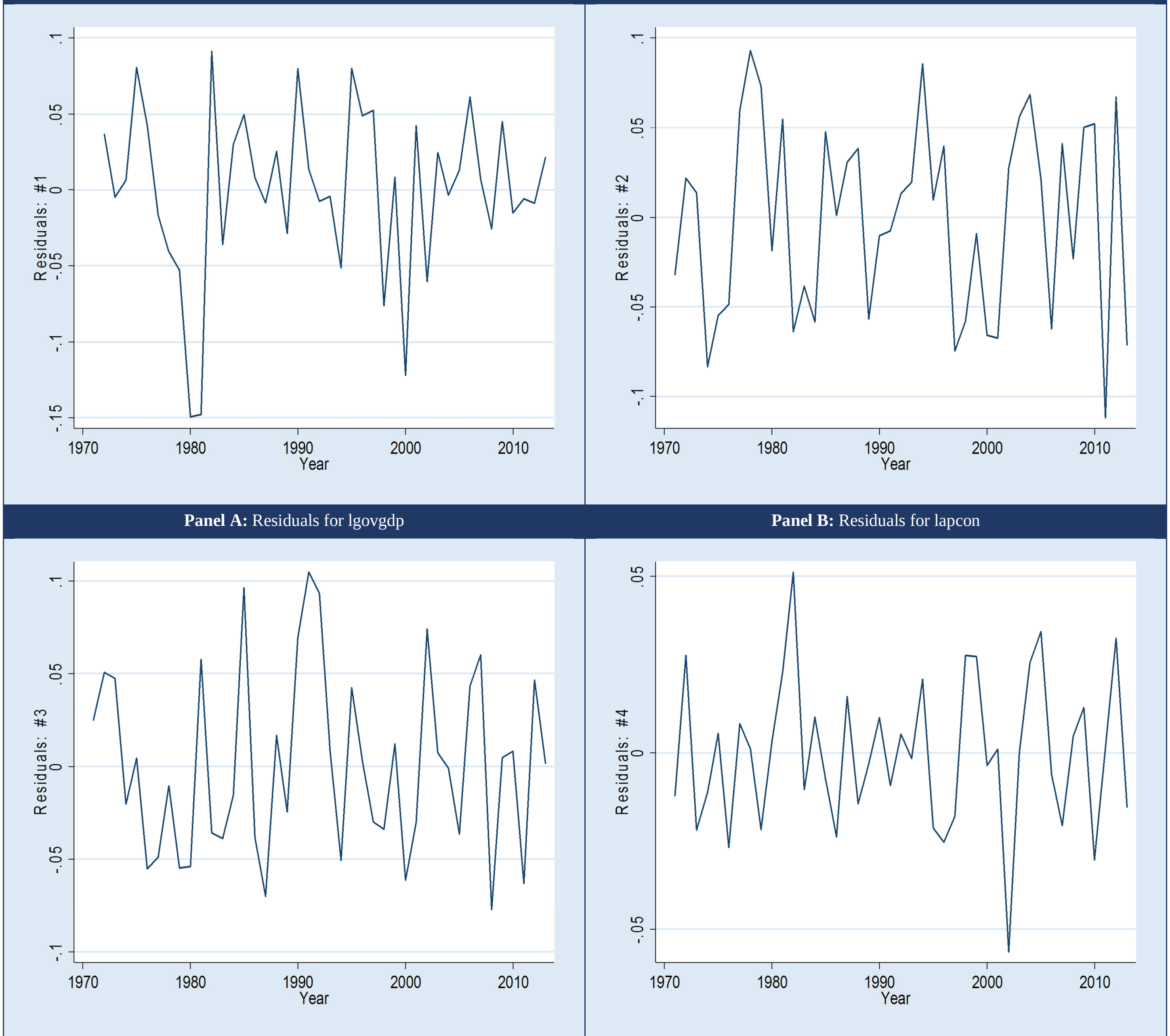
$lsavr$ – the logarithm of the **savings rate**
$ltbill$ – the logarithm of the local treasury bill rate
t – linear time-trend used
$tech2$ - the quadratic of the time trend
$tech3$ – the interaction of the linear time trend and the logarithm of the time trend

This expenditure forecasting model can be adjusted and estimated to forecast government spending in each of the partner countries, using the **data now publicly available.** Data are available from 1970 to 2013. Further, if other official data are available, the method can be further developed. The forecast equation (64) can be estimated independently with dynamic OLS regression. However, an approach that yields the most efficient estimates of the parameters is to estimate all together using 3SLS. Here is some simple Stata code for this purpose, written without macros:

```
>reg3 (lgovgdp l2.lexppop l.lapcon lpk l.lpk lothmin l.lgovgdp l.lklratio)
///
(lapcon lothmin lresus leximp ldevgap2 lexchr ltbusa l.lgovgdp l.lsavr) ///
(lcpi lm2 lgdpdef lapcon lexchr ltbusa l.lsavr) ///
(lklratio t tech2 tech3 lcpi ltbusa ltbill l.lgov lsavr l.lsavr l.ltbusa),
///
endog(lgovgdp lapcon lcpi lklratio) ///
exog(lothmin lsavr t tech2 tech3 lresus ldevgap2 ltbusa ///
lm2 lgdpdef lexchr lgov leximp lexppop lpk ltbill)
>predict reseq1,equation(#1) residuals
>tsline reseq1
>predict reseq2,equation(#2) residuals
>tsline reseq2
>predict reseq3,equation(#3) residuals
>tsline reseq3
>predict reseq4,equation(#4) residuals
>tsline reseq4
>estimates store devmodeqs
>forecast create devmod,replace
>forecast estimates devmodeqs
>forecast solve,begin(2010)
>tsline lgovgdp f_lgovgdp
>list f_lgovgdp
```

The **reg3** lines of code call the Stata routines for 3SLS. The **predict** lines of code allow the analyst to generate the residuals for each equation. The **tsline** codes allow visual examination of the residuals while specifying and estimating the model. In practice, the analyst would need to adjust to ensure that the final forecast model has stationary residuals, with the aim being to bring them as close to as possible to being white noise. Figure 12 displays the residuals generated with data for the case of Trinidad and Tobago at the point when the specification and estimation process was brought to a halt. They are stationary even if not white noise and can be improved with more research.

Figure 3-12: Residuals for Expenditure Forecasting Model



Panel A: Residuals for lgovgdp

Panel B: Residuals for lapcon

Panel C: Residuals for lcpi

Panel D: Residuals for lklratio

The forecast of **lgovgdp** comes from four additional steps.

a. The code **'estimates store devmodeqs'** stores the results of the estimation process under the name **devmodeqs** **and** makes them available to the next line and step**,** which name where the set of forecast equations will be held with the code **'forecast create devmod, replace'.**

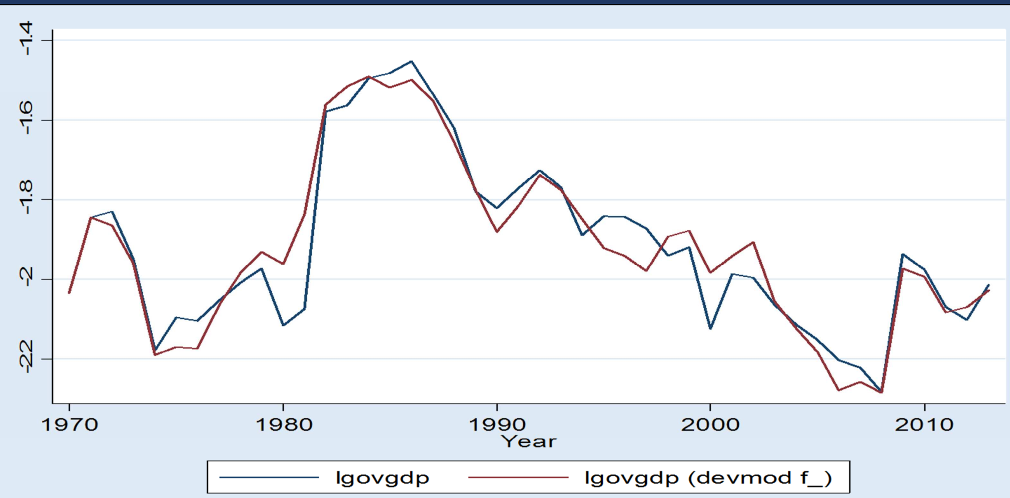
b. The **'replace'** component allow the model to be run repeatedly without having to change the name of the model pool.

c. The code **'forecast estimates devmodeqs'** loads the estimated equations into **devmod.**

d. The code **'forecast solve, begin(2010)'** tells the computer to solve the system of equations numerically, which is the same as generating the actual forecasts. In this case, the instruction is to report forecasts starting from 2010. Any start data within the time series can be used, after the number of lags and leads are considered. The above model uses at most 2 lags, so forecast can start in 1972 and go all the way to 2013.

The graph of the revenue forecasts can be produced with the code **'tsline lgovgdp f_lgovgdp',** and the forecasts can be listed with **'list f_lgovgdp'.** Figure 13 is the graph of the forecast from 1972 to 2013. It is clear from inspection that the model gives a fairly good 'in-sample' forecast of **lgovgdp**. To forecast to 2015 or 2016, all the other variables will have to be updated to the chosen budget year, using one of the independent forecasting methods presented above.

Figure 3-13: In-sample forecast of lgovgdp



Since the forecast equation generates $lgovgdp_{t+1} = \ln(\frac{G}{Y})_{t+1}$, it is necessary to recover the forecasted level of government spending with $\widehat{Y_{t+1}}$ from equation (61) and the following equation:

68. $\widehat{G_{t+1}} = \widehat{Y_{t+1}} * \exp lgovgdp_{t+1}$

As indicated above, this aggregate estimate provides an estimate of the budget ceiling that can be put through the judgement forecasting process and be subjected to further political influence when the budget is being made.

Disaggregated Budget Forecasts

The most widely used method of preparing disaggregated budget forecasts is an application of the elasticity approach specified to cater directly for cases when long time series are not available. The general method is used to forecast categories that change due to changes in underlying macroeconomic, fiscal, structural and socio-demographic variables. The general formula for the method is to use equation (1) and obtain the forecast of the rate of growth of $g_{x,t}$ using independent elasticity measures. The general formula is:

69. $g_{x,t} = \varepsilon_{X,Y} g_{y,t}$

where, $\varepsilon_{X,Y}$ is an independent estimate of elasticity. Application of the formula is subject to the conditions specified above in relation to equation (56): (i) non-stationarity, and (ii) cointegration between X and Y.

Regarding the independent estimates of $\varepsilon_{X,Y}$, an elasticity of 1 implies that a 1% increase in Y_t results in a 1% increase in X_t. This is most widely evident with taxes on habitual consumer expenditures, such as sugar, salt, alcohol and cigarettes. However, if data are not available for an independent estimate, the PER Team should use an elasticity of 1. The PER Team should use an elasticity other than 1 only when it is backed by economic rationale and tends to reduce bias and increase precision in the estimates. Taxes with a **progressive schedule,** such as many types of income tax**,** will yield disproportionately high revenues compared with the tax base. In those instances, the elasticity will be higher than 1. Taxes with a **regressive schedule** will yield disproportionately low revenues compared with the tax base. In those instances, the elasticity will be lower than 1. The outcome is similar for taxes which are based on a linear schedule but are subject to a depreciation tax allowance or an earnings cap. Elasticities other than 1 may also reflect sustained **structural changes** within the underlying variable. For example, an ongoing movement from the consumption of goods and services taxed at the standard VAT rate to goods and services taxed at a reduced VAT rate would cause the elasticity of VAT revenues from consumption to fall below 1.

Disaggregated Expenditure Forecasts

The disaggregated approach to expenditure forecasting is a ‘bottom-up’ mapping of expenditures to underlying indicators. The core list of expenditure categories to be forecasted has been provided earlier in the Annex. Table 4 lists the expenditure categories, suggested forecasting methods including equation (69), and suggested matching indicators to be used when generating the forecasts. In general, expenditure forecasts are derived from underlying economic and social variables.

Within a broad COFOG classification, all the listed budget categories should be forecasted, even if on an *ad hoc* basis. The underlying variables used in forecasting a budget category is an analytical decision based on the PER Team's understanding of the socio-economic factors that are highly correlated with it. The choice of underlying variables should be limited to variables for which forecasts exist or can be obtained through one or more of the forecasting methods described above. Thus, the use of equation (69) may be combined with other forecasting methods. The disaggregated forecasts are used together with the budget ceilings and strategic allocations when evaluating the allocative efficiency of the budget.

Table 3-4: Methods and Indicators used to Forecast Government Expenditure

Budget Category	***Forecasting Method***	***Main Underlying Indicators Suggested***
Intermediate Consumption	x_{t-1}; trend; elasticity	Budget subcategories, health sector reform
Compensation of Employees		
Wages and salaries	Trend; elasticity	Negotiated wages, CPI, employment growth, structural or socio-demographic changes of employment; Financial rules and regulations
Employers' social contributions	Elasticity	Growth of wages and salaries
Employers' imputed social contributions	Trend	
Tax **Incentives/Allowances** on Production and Imports	Elasticity	Growth of wages and salaries
Subsidies	x_{t-1}; trend	
Tax **Incentives/Allowances** on Property Income to Encourage Entrepreneurship and Investment	Elasticity	Level of market interest rates, debt maturity profile, growth of government debt
Tax **Incentives/Allowances** on Current Income, Wealth, etc.	x_{t-1}	Budget subcategories
Social Benefits other than Transfers in Kind		
Pensions	Trend; elasticity	Official growth pension payments, structural and sociodemographic changes and number of pensioners
Unemployment benefits	Elasticity	Number of unemployed
Long-term care benefits	Trend	Recipients of long-term care benefits (by levels of care)
Family benefits	Trend; elasticity	Budget subcategory, demographic changes
Social benefits n.e.c.	x_{t-1}	
Social Transfers in-Kind	Trend; elasticity	Budget subcategory, CPI, number of recipients of long-term care benefits, pupils, old-age or invalidity
Other Current Transfers		
To Households	Trend	Budget subcategories
To CARICOM, etc.; To OECS	Subjective/qualitative	CARICOM Budget
Current transfers n.e.c.	Trend	Budget subcategories
Capital Transfers		
Bank support package	Subjective/qualitative	Level of market interest rates, annual reports, budget plan
Capital transfers n.e.c.	x_{t-1}; trend	
Gross Capital Formation by Government		
Other general services	Trend	Budget subcategory
Health	Subjective/qualitative	Health care reform
Research and development	Trend	Budget subcategory
Road, air, and related infrastructure for transportation	Trend	Budget subcategory; Public works plan
National Security	x_{t-1}	
Education	Subjective/qualitative	Education development plan
Depreciation/Maintenance	Subjective/qualitative	Financial rules and regulations; Existing Public Sector Reform programs
Contingencies	x_{t-1}	

To develop the forecasts, the following specific steps must be followed:

1. Identify the expenditure category to be forecasted.
2. Identify the underlying indicator or predictor.
3. Deflate the nominal expenditures and nominal indicators where applicable, in order to eliminate price effects. The GDP deflator or an alternative price index would be needed for this purpose.
4. Identify, quantify and deflate all discretionary changes in expenditure policy.
5. Adjust the real values of the expenditure categories to remove the effects of changes in discretionary policy.
6. Identify the applicable independent estimate of elasticity.
7. Forecast the growth of the underlying indicator variable.
8. Using the forecasted indicator growth rate and the elasticity indicator to forecast the future expenditure.

Disaggregated Revenue Forecasts

Instead of using dummy variables to monitor the effects of discretionary policy changes in equation (58) or equation (59), an important step is to adjust the taxes directly to reflect the changes. This is best considered in the context of the application of equation (56) to each distinct tax base. If data are available, it is better to proceed by forecasting each component of the tax collection using forecasts of its associated tax base generated by the macroeconomic growth model. There should be a close relationship between the tax base and the taxes collected. Also, forecasting of the distinct components keeps in view the structure of the tax system and the tax base, which is also of interest in sector PERs. Sector PERs must monitor the revenue sources that are generated by the activities/programs and agencies funded by sector allocations. This is a bottom-up mapping approach to revenue forecasting.

To apply this method, the following specific steps must be followed:

1. Identify the tax categories that should be forecasted.
2. Identify the matching tax base of each tax.
3. Deflate the nominal taxes and tax base to eliminate price effects. The GDP deflator or an alternative price index would be needed for this purpose.
4. Identify, quantify and deflate all discretionary changes in tax policy– the tax rate and the tax base.
5. Adjust the real values of the tax categories to remove the effects of changes in discretionary tax policy.
6. Regress the natural logarithm of the adjusted real taxes on the natural logarithm of the relevant tax base.
7. Forecast the tax base with a multi-sector macroeconomic model, such as indicated in Annex 2.
8. Using the forecasted tax base, forecast the future real tax flow using the β_0 and β_1 for that tax base.

The rule of thumb normally used is that a tax should be forecasted if it accounts for more than 5% of total taxes (Jenkins et al, 2000: 53).[19] Taxes should also be forecasted if they are based on consumption habits, such as alcohol and cigarettes, or are difficult to escape because they are collected at a point of transaction, such as purchase of an airline ticket. In the partner countries, the set of taxes to be considered are at least the following:

1. Income, profits taxes – PAYE
    i. Individuals
    ii. Corporations
    iii. Other
2. Domestic taxes on goods and services
    i. Sales taxes – VAT
    ii. Motor vehicles
3. Property taxes
    i. Inheritance and gift
    ii. Land and other property taxes
    iii. Other recurrent property taxes
4. Other taxes
    i. Stamp duties
    ii. User fees and other taxes not elsewhere classified
5. Draw-downs from sales of government-owned business enterprises
6. Receipts of social security contributions
    i. From employers
    ii. From employees
    iii. From self-employed
    iv. Other social security contributions

Table 5 lists the matching tax-bases of some of these taxes:

Table 3-5: Important Taxes and Tax Bases in Partner Countries

Tax Category (X_j)	Tax Base (Y_j)
PAYE	Personal Income (wages and salaries; emoluments; dividends; profits)
Alcohol and cigarette taxes	Alcohol and cigarette sales
Value-added Taxes	Consumer spending (subject to taxation)
Travel taxes	Airline sales
Corporate taxes	Business income (operating surplus), including household unincorporated enterprises
Real estate tax	Real estate sales
User fees	Construction and building permits service fees for education and health
Trade tariffs	Exports; imports

[19] Jenkins, G.P., Kuo, C., and Shukla, G.K. (2000). *Tax Analysis and Revenue Forecasting: Issues and Techniques.* Harvard Institute for International Development, Harvard University.

Equations (9) and (10) of the macroeconomic development model in Annex 2 imply that corporate tax forecasting model should distinguish between the characteristics of different industrial sectors, with particular regard to their capital intensive nature – their capital-labour ratio. This is also related to the volatility of depreciation-related tax deductions accompanying differing investment adjustments to economic conditions. It is also related to the problems of measuring income when doing national income accounting for some sectors such as the finance sector.

In the bottom-up approach, the aggregate revenue forecasting models are largely accounting frameworks designed to account for the conceptual differences between the taxable base, which reflects tax law and changes in the economic base. These differences are most pronounced for the income tax heads of revenue, in particular corporate income tax. Econometric techniques are not usually required to uncover the quantitative relationships between a head of revenue and its taxable base when the elasticity of revenue to its taxable base is one. Intuitively, this means that taxation revenue increases by one per cent for each increase of one per cent in the taxable base.

The main exception is Treasury's model for income tax withholding, which incorporates an elasticity which has been econometrically estimated, to capture the progressivity of the individuals' income tax system. The revenue forecasting models generate forecasts on an income year or accrual basis. To generate forecasts on a cash basis these forecasts must be adjusted for the revenue's payment arrangements. To produce a forecast for each head of revenue the cash tax revenue forecasts are then adjusted for the estimated impact of government policy decisions, court decisions and compliance activity.

The payment arrangements introduce a lag between the timing of the economic activity and the receipt of the associated revenue — for example, 60 per cent of corporate income tax is typically received in the year that the profit is generated, with the remaining 40 per cent received in the following year. Further adjustments are made to take into account any information on recent taxation collections. It follows from this approach to forecasting that the PER should issue an opinion on the need for continued work to understand each taxable base and the tax payments system. The PER should consider the difference between the taxable base and the economic bases, and to identify and resolve instances where taxation receipts and economic data are providing conflicting signals about the state of the nominal economy.

3.2.1.17 A Note on Microsimulation of Revenues

Microsimulation uses a wide range of models mainly to impute missing data about tax liability in the light of particular changes of public policy. The models estimate tax liability using data supplied by tax filers and non-filers.

Although it is a data- and computationally-intensive method, microsimulation is growing in use in revenue forecasting because it has a key readily available data source – the tax returns and other data of individuals and businesses. Good available references for the method are (Jenkins et al, 2000)[20] and Mitton, Sutherland and Weeks (2000).[21] The summary presented here follows these references closely. The simulation methods are 'micro' because they rely mainly on data about individuals, households or businesses who make decisions and undertake activity that are affected by the discretionary policies of government, and who report their resulting tax obligations to the authorities. Reliance on this data is a strength of the method, but is also a weakness in the partner countries because micro data availability is an issue.

The method tends to rely on survey data, which has the additional strength of being representative of the population taxpayers and non-filers. Sample sizes range from 1% to 5% of the employed population, adjusted for resource availability to field the survey. Multi-stage stratified random sampling is practical and cost-effective,[22] and is consistent with the methods used by the National Statistical Organisations of the partner countries. The underlying strata for tax simulation would normally be as follows:

1. Source of income
    a. salaried employment
    b. investment,
    c. type of industry, for example
        i. mining
        ii. farming,
        iii. fishing,
    d. type of profession
        i. professional
        ii. business
2. Location of residence
    i. Foreign vs local
    ii. Urban, rural, district
3. Tax status
    i. Taxability
    ii. Income size

[20] Jenkins, G.P., Kuo, C., and Shukla, G.K. (2000). *Tax Analysis and Revenue Forecasting: Issues and Techniques.* Harvard Institute for International Development, Harvard University.
[21] Mitton, L., Sutherland, H. and Week, M. (2000). *Microsimulation Modelling for Policy Analysis.* Cambridge: Cambridge University Press.
[22] The list of possible methods are: (i) simple random sampling; (ii) systematic sampling; (iii) stratified sampling; (iv) cluster sampling; and (v) multi-stage sampling.

However, reliance on survey data also makes the method vulnerable to bias due to the severe under-reporting challenges confronting all income-related survey in these countries.

The basic strategy of the forecasting method is to use modern statistical software to develop the statistical distributions of the taxes reported by individuals and then estimate the effects of a given policy on the distribution. Distributions can be developed and compared for various subpopulations: income groups, sex, age, family size, and the like. The capacity to process large datasets speedily is the core strength of the computational methodology. The great advantage of large datasets is the tendency for all distributions to tend to the normal distribution. This is an advantage because such distributions can be completely described by their **mean** and **variance**, and they allow a wide range of simulation techniques to be used, each with a specifiable level of confidence. The focus of the simulations is on estimation of the differences in the distributions between distributions before and after the policies (treatments) are applied, and on the associated implications for collectible tax revenues. The method also simulates the differences of groups who are affected by the policy and those who are not, but who are otherwise identical. Once micro data are available from a sufficiently large random sample of taxpayers, several steps are following, which are broadly described here:

Data aging: This is a method of imputing new characteristics to the existing micro units, by reweighting and by indexation to particular money amounts. Some methods use no aging, preferring to use updated information on the same units, as is done in a longitudinal panel. Surviving units age by one year in each year of observations. Deterministic and stochastic methods are used to do the change of status, especially in terms of:

1. Income
2. Employment
3. Housing, including cohabitation
4. Parenthood

Addition of Behavioural Response: Most often panel data are used to consider how behaviours change with age – behaviours such as dependence on household, fertility, skill, and the like, that have an effect on income and wealth.

Methods are either static or dynamic. Static methods use the data for a single survey and can suffer from considerable bias. Dynamic methods rely on longitudinal panels with a sufficiently large number of survivors from period to period to give information about the ‘transition probabilities’.

Summarily, with the individual tax returns in hand and with suitable survey data, the microsimulation models will make relevant calculations based on the applicable laws and regulations regarding tax liability, including incentives and allowances. Note the following:

1. Since the data comes from a multi-stage stratified sample, each tax return has an associated weight. The weight defines the number of other filers represented in the stratum from which it comes.
2. Calculations of the model will adjusted the observations by these weights in order to compute sums and estimate total revenue yielded by any type of tax and all taxes.

Taking into account the aging methods and the behavioural assumptions, the data sets in use will contain:

a. Historical database of individual tax returns.
b. Weighted historical data of individual tax returns.
c. Aging or growth factors. These factors grow the historical database to represent the current position and to a chosen future year.
d. Behavioural factors. These factors further adjust the aged data to better represent the position in the chosen future year.
e. Historical database of individual tax returns incorporating respective growth rates and behavioural factors. Thus, for example, the wages in the database are increased by a wage-growth factor up to the current or future year.
f. Weighted historical database of individual tax returns incorporating respective growth rates and other behavioural factors. Here the weights or raising factors are applied to raise the aged historical database to arrive at the situation of the total population.

With this data in hand, the simulation models can be run, including cross-section and longitudinal versions of the regression models that relate tax revenues to their respective tax bases.

4 Summary

Apart from microsimulation, the set of methods described in this Annex comprise a basic set that can be used with the data available in the partner countries. The methods range from expert judgment to macroeconomic forecasting based on a model that treats the economy as a system. The latter is needed to support the search for consistency in sector allocations rather than produce a budget that is merely an amalgam of ministry requests. It is emphasized that judgement is an important part of the forecasting effort. This judgement should be developed through expert opinion and through dialogue with the business community and other informed stakeholders.