The Asymmetric Behaviour of Tax Revenue over the Business Cycle

Wian Boonzaaier

ESSA Conference, September 2015

Abstract

Tax revenue forecast errors on average constituted more than 80 percent of the forecast error of the budget balance-to-GDP ratio between 2000/01 and 2010/11. Some portion of the tax revenue forecast error could be attributed to the use of constant or long-term tax elasticities over the course of the business cycle, which leads to systematic over- and under-estimation during downswing- and upswing phases respectively. Even when distinguishing between short- and long-run tax elasticities via an error-correction framework, the implicit assumption is made that tax revenue responds to changes in economic activity in a symmetric fashion. This paper therefore aims to explicitly test for possible tax revenue asymmetries relative to the business cycle via a smooth transition autoregressive moving average framework which allows for the estimation of two separate regimes, i.e. a low growth (or contractionary) regime and a high growth (or expansionary) regime. Using this framework, the movement in tax revenue at all times will be governed by a weighted average of two different linear models, where the weighting of the two models depend on the recent history of the tax revenue series as well as a measure for the output gap. Preliminary results show that tax revenue collections do react differently depending on whether it is in a high growth phase or a low growth phase. This has implications for policymakers in terms of the revenue forecasting process and the way the cyclically adjusted budget balance is calculated.

1 Introduction

Wolswijk (2007: 6-7) notes that the sensitivity of tax revenue collections to changes in the macroeconomic environment (i.e. the buoyancy or elasticity) is an essential input into the revenue forecasting process. This elasticity is often assumed to be constant over time[1], although in practice this may lead to systematic over- and underestimation of tax revenue. This is especially true for cyclically-sensitive tax products. In addition, Jooste and Naraidoo (2011: 114-115) argues that an expansionary business cycle supports tax revenue via automatic stabilisers and a general improvement in people's willingness to pay tax. Conversely, when the economy is in a downswing phase, tax revenue collections tend to worsen. Evidence of significant revenue forecast errors for the case of South Africa and its consequences are highlighted in the paper by Calitz, Siebrits, and Stuart (2013), who observes that revenue forecast errors were largely responsible for Budget projections deviating from the actual budget balance-to-GDP ratio outcome. More specifically, revenue forecast errors on average contributed 82.1 percent to the overall forecasting error of the budget balance-to-GDP ratio over the fiscal period 2000/01 to 2010/11. A similar trend was observed during more recent financial years. This necessitates the important research question of whether this systematic bias in revenue forecasts is partly the result of asymmetric movements in tax revenue relative to the phases of the business cycle that are not captured through a traditional linear regression framework. This question has relevance for policy makers given that failure to account for transitory movements in revenue and expenditure components may reduce the effectiveness of fiscal policy interventions.

Interest in non-linear modelling methods have been increasing in recent years given the need for models that can analyse the regime-switching elements present in time series. According to Tsay (2002: 137-138), many non-linear models have been put forward in the literature, from the bilinear models of Granger and Andersen (1978), the threshold autoregressive model of Tong (1978) and later Tong and Lim (1980), as well as the Markov switching model of Hamilton (1989). While in the Markov switching type models the regimes are determined exogenously via an unobservable state variable, the regimes in the threshold model are determined

[1]When using a linear regression model to forecast tax revenue, the coefficient obtained from the regressor (which is assumed to represent the tax base) is the average elasticity over the period. This coefficient is typically used to forecast tax revenue.

endogenously. Given that threshold models are useful for explaining the asymmetric nature of declining and rising patterns in a time series, it will be useful for purposes of testing the behaviour in tax revenue growth. As noted by Tsuay (2002: 130), the TAR model makes use of threshold space to obtain a better linear approximation of the conditional mean equation as opposed to the traditional piecewise linear model.

This paper is an attempt to build upon the work done on the asymmetric behaviour of tax revenue over the business cycle by Jooste and Naraidoo (2011), and employs a non-linear methodology that has not been explicitly applied to tax elasticity estimation and forecasting in a South African context. More specifically, this paper will attempt to develop a smooth transition autoregressive moving average model with external regressors for the three main tax products: Personal income tax (PIT), corporate income tax (CIT), and value-added tax (VAT). The plan of this paper is as follows. Section 2 surveys the available South African literature on tax elasticities and the various methods used to calculate them. Section 3 provides an overview of threshold autoregressive models and its extensions, in addition to discussing the various LM-type tests for the presence of STAR non-linearity. Section 4 provides a description of the dataset, followed by a data-based modelling cycle process of the various smooth transition ARMAX models. Section 5 concludes.

2 Literature review

According to Leal, Tujula, and Vidal (2007: 9-10), the majority of institutions responsible for producing fiscal projections in their respective countries make use of some form or combination of the following techniques: simple linear regression, time series analysis, structural macro-econometric models, or subjective judgement based on experience. An attempt is also made to balance theoretical grounding with specific policy requirements when deciding on which techniques to use. However, it is not clear from available studies which methods are superior when preparing fiscal projections. Most of the available papers make use of accuracy measures to select a superior forecasting method, while others focus on the forecast properties of a specific technique or method. What follows is a survey of the different techniques that have been used in the South African literature to estimate tax elasticities and ultimately forecasting tax revenue collections using these elasticitities over the course of a full business cycle. Note that the focus of this paper will only be on model development and analysis.

South African studies have focused predominantly on linear elasticities. The role and impact of tax revenue as an automatic fiscal stabiliser in the South African economy since the 1970s were investigated by Swanepoel and Schoeman (2002). As part of their analysis, they calculated the sensitivity of broad tax categories with respect to gross domestic product. Linear regression was used with no distinction between short- and long-term elasticity estimates, while no attempt was made to remove the effect of discretionary tax policy changes. The majority of coefficients were larger than one, and signifies a tax structure that is highly responsive to movements in economic activity. In addition, the cyclical component of tax revenue and its standard deviation is calculated. The standard deviation, which provides a rough approximation of the sensitivity of tax revenue to the business cycle, shows that cyclical tax revenue is less volatile than the business cycle as a whole. Notwithstanding the lower volatility, the authors find a high correlation between the output gap and cyclical tax revenue.

Du Plessis and Boshoff (2007) analysed the cyclicality of fiscal outcomes since the beginning of 1990 in South Africa in an attempt to evaluate the role of fiscal policy during this period of economic stability. The authors focused on the estimation of long-run tax elasticities with respect to output, either directly or indirectly. For direct estimation, an unrestricted vector autoregression with a lag order of four quarters and the tax product and output as endogenous variables is used to calculate the long-run elasticity. The majority of the tax elasticities are calculated using the direct method. In select instances the indirect method is used, where the tax elasticity with regards to output is calculated as the product of the tax elasticity with regards to the applicable tax base and the tax base elasticity with regards to output.

According to Wolswijk (2007), a distinction can be made between short- and long-run elasticities for PIT, CIT, and VAT. The long-run tax elasticity is obtained from the cointegrated relation between the tax products and its tax base, while the short-run tax elasticity is derived from the error correction model. A further distinction can be made between positive and negative values of the error correction term to allow the short-term tax response to be asymmetric. Moreover, asymmetric responses arising from whether the tax base is above or below equilibrium are also accounted for. The author removes the effect of discretionary policy changes, thereby deriving the 'true' elasticity as opposed to the tax buoyancy factor. The author finds evidence for the case of the Netherlands that short-term elasticities deviate substantially from long-term elasticities, especially when tax revenue is below equilibrium. Assuming that short-term elasticities are not different from long-term elasticities may lead to 'budget surprises' as well as inaccurate cyclically-adjusted fiscal indicators, which has consequences in terms of policy. Although the error-correction terms have some correlation with cyclical indicators such as the output gap, it does not fully capture the effect of the business cycle on tax elasticities.

The paper by Jooste (2009) estimates tax elasticities for PIT, CIT and VAT as part of a broader framework to calculate the South African structural budget balance. The author proceeds by first estimating potential output and the associated output gap. This is followed by econometric estimation of two types of elasticities for each respective tax product. The first elasticity relates changes in tax revenue collected to changes in the tax base, and are obtained by regressing each tax product in nominal terms on its respective inflation-adjusted tax base. The second elasticity captures the cyclicality of the respective tax bases, and is obtained by regressing the cyclical components of each tax base on the output or income gap. The overall elasticity is obtained by multiplying the two calculated elasticities. The author finds firstly that adjusting for temporary changes in tax revenue may improve policy makers' ability to construct a sustainable fiscal framework. Second, given the disparity in elasticity estimates using different methodologies, it would be prudent to supplement these estimates with professional judgement and in-depth research.

Jooste and Naraidoo (2011) represented a significant departure from the South African tax elasticity literature at the time by employing an autoregressive distributed lag (ARDL) model augmented with a smooth transition regression model to estimate tax elasticities over the period 1994 to 2009. Within the STARDL framework, the authors directly estimate tax elasticities for PIT, CIT and VAT by combining estimates of tax proceeds relative to changes in the tax base with estimates of the sensitivity of the tax base to the business cycle. One of the authors' main findings is that tax collections move asymmetrically over the cycle, which has implications for how policy makers project tax revenue. Given this conclusion, it follows that tax revenue collections tend to be underestimated during upswing phases of the economy, and overestimated during downswing phases of the economy when using linear elasticities.

3 Methodology

This section will provide a brief overview of the threshold autoregressive model and its variants, followed by the introduction of the logistic smooth transition autoregressive model. Moving average terms are then introduced to the model, along with the ability to incorporate external regressors.

3.1 Threshold autoregressive models

According to Enders (2010: 439), regime-switching models make the behaviour of some $\{y_t\}$ process conditional on the state of the system. A relevant example is VAT revenue, which are highly buoyant during the upswing phase of the economy but becomes relatively unresponsive to movements in consumption during a downswing phase as individuals buy relatively more zero-rated goods or lower-valued goods (Wolswijk 2007: 20). As a result, the dynamic adjustment equation of VAT collections will therefore depend on whether the economy is in an expansionary or contractionary regime.

Based on Enders (2010: 439-442), consider the following two-regime threshold autoregressive (TAR) model:

$$y_t = \begin{cases} \alpha_{10} + \alpha_{11} y_{t-1} + ... + \alpha_{1p} y_{t-p} + \epsilon_{1t} & \text{if } y_{t-d} > \tau \\ \alpha_{20} + \alpha_{21} y_{t-1} + ... + \alpha_{2p} y_{t-p} + \epsilon_{2t} & \text{if } y_{t-d} \leq \tau \end{cases}$$

where y_t is some variable of interest (e.g. tax revenue collections), y_{t-d} is the threshold variable, d is the delay parameter, and τ is some threshold. It can be stated that the TAR model is essentially a piecewise linear AR model in threshold space (Tsay 2002: 131). As illustration, when y_{t-d} is larger than the threshold τ, the $\{y_t\}$ sequence will be modelled using the first autoregressive process. Similarly, when y_{t-d} is equal to or less than τ, the $\{y_t\}$ sequence will be governed by the second autoregressive process. The purpose of the delay parameter d is to account for the fact that the regime switch may take more than one period to occur. It is standard procedure to estimate a TAR model for each value of the delay parameter, and then choosing the value where the residual sum of squares is minimised or alternatively the parameter value which yields the smallest AIC or SBC value. Note that y_t is linear in each regime, but is non-linear overall given the possibility of regime switching. The threshold variable y_{t-d} can be generalised to z_{t-d} so that it is measurable to a function of elements of F_{t-1}. The generalised version of the model is known as an open-loop TAR model.

3.2 Logistic smooth-transition autoregressive models

In some instances, it may not be optimal to model some variable of interest using a threshold rule that assumes immediate adjustment. In addition, as noted by Van Dijk and Franses (1997: 3), the value of the threshold is unknown[2] in most instances and the estimation of such threshold is problematic. A suggested alternative, according to Enders (2010: 457-462), is the smooth-transition autoregressive (STAR) model, which allows the autoregressive parameters to change gradually and alleviates the problem of estimating the threshold directly. Consider the following two-regime logistic[3] STAR (LSTAR) model of order p using (adapted) notation contained in Van Dijk and Franses (1997) and Ghalanos (2014):

$$y_t = \phi_1' y_t^{(p)} F(z_{t-d}; \gamma, \alpha, c) + \phi_2' y_t^{(p)} (1 - F(z_{t-d}; \gamma, \alpha, c)) + \epsilon_t$$

[Eq.1]

$$F(z_{t-d}; \gamma, \alpha, c) = (1 + \exp(-\gamma(\alpha' z_{t-d} - c)))^{-1}$$

[Eq.2]

where y_t is some variable of interest, $y_t^{(p)} = (1, \tilde{y}_t^{(p)})'$ where $\tilde{y}_t^{(p)} = (y_{t-1}, ..., y_{t-p})'$ is a vector of lagged values of the variable of interest, $z_{t-d} = (z_{1t-d}, ..., z_{jt-d})'$ is a vector of k observed variables known as the transition variable which is assumed to explain the state transition with delay parameter d, $\phi_i = (\phi_{i0}, \phi_{i1}, ..., \phi_{ip})'$, $i = 1, 2$, ϵ_t is white noise with zero mean and standard deviation σ, and $F(.)$ represents the logistically transformed transition function bounded by zero and one. Also, $\alpha = (\alpha_1, ..., \alpha_p)'$ is a vector of parameters where it is usually assumed that only the d-th row takes on a value of one and the rest is zero (for identification purposes). The parameters γ and c are scalars representing the smoothness parameter and threshold respectively.

The above notation makes it clear that the $\{y_t\}$ sequence can be expressed as a weighted linear combination of the two autoregressive processes at any point in time, with the weights assigned to each autoregressive process being dependent on the transition function. According to Van Dijk and Franses (1997: 3-4), when $\alpha' z_{t-d}$ is small, the transition function $F(.)$ will tend to zero, which implies that the largest weight will be allocated to the second equation. Similarly, when $\alpha' z_{t-d}$ is small, the transition function will tend to one, which will result in the first model equation receiving the largest weight. A higher value of γ will result in a faster change in weights as $\alpha' z_{t-d}$ increases. In addition, as γ tends to zero, the model becomes linear and the model weights become constant and equal to 50 percent. As γ tends to positive infinity, the LSTAR model will reduce to a two-regime self-exciting TAR model.

[2] Chan (1993) suggested an iterative procedure where, after some percentage of the highest and the lowest observations have been removed, the threshold is set equal to the remaining observations in a consecutive manner, where the threshold is contained in the regression with the smallest residual sum of squares.

[3] There are two general forms of the STAR model: (1) The logistic form, also known as the LSTAR model and (2) the exponential form, also referred to as the ESTAR model. The focus of this paper (and most studies) is on the logistic form mainly for analytical reasons.

3.3 Further extensions to the standard LSTAR model

Based on Teräsvirta (1998) and Ghalanos (2014), the LSTAR model as depicted in [Eq.1] can be extended to include external regressors (also known as a smooth transition regression model) as well as moving average terms in the two regimes. Combining the aforementioned extensions lead to the model known as the smooth transition ARMAX model (or STARMAX), and can be written as follows:

$$y_t = (\phi_1' y_t^{(p)} + \xi_1' x_t + \psi_1' e_t^{(q)}) F(z_{t-d}; \gamma, \alpha, c, \beta) + (\phi_2' y_t^{(p)} + \xi_2' x_t + \psi_2' e_t^{(q)})(1 - F(z_{t-d}; \gamma, \alpha, c, \beta)) + \epsilon_t$$

[Eq.3]

where $e_t^{(q)} = (e_{t-1}, .., e_{t-q})'$, $\psi_i' = (\psi_{i1}, ..., \psi_{iq})'$, $x_t = (x_1, ..., x_l)'$, $\xi_1' = (\xi_{i1}, ..., \xi_{il})'$. The moving average terms and associated coefficients in the two regimes are represented by $e_t^{(q)}$ and ψ_i' respectively. The l external regressors are represented by x_t with coefficients ξ_1'.

3.4 Testing for the presence of STAR non-linearity

Based on the discussion by Van Dijk et al (2000: 11-14), the presence of non-linearity is tested by equating the autoregressive parameters in the two regimes to create the null hypothesis, $H_0 : \phi_1 = \phi_2$, while the alternative hypothesis is $H_0 : \phi_1 \neq \phi_2$ for at least one $j \in 0, ..., p$. However, a standard Lagrange multiplier test for the presence of STAR behaviour cannot be performed given that testing the null hypothesis of non-linearity on a STAR model would be equal to setting the smoothing parameter γ equal to zero. This problem arises due to the presence of unidentified parameters under the null hypothesis. These parameters are not restricted by the null hypothesis, and do not add any information in the case that the null hypothesis holds true. As a result, it is not possible to use conventional statistical theory to obtain the asymptotic distributions of the test statistics.

The framework created by Luukonen et al (1988) outlines a specification procedure for STAR models. Due to some model parameters being unidentified when γ is set to zero, their test replaces the transition function with an appropriate Taylor series approximation around $\gamma = 0$. The identification problem is resolved, and a Lagrange multiplier statistic with a standard asymptotic χ^2 distribution can be used to test for model linearity.

Consider the simple logistic STAR model as depicted in [Eq.1], which can be re-arranged to give:

$$y_t = \phi_2' y_t^{(p)} + (\phi_1 - \phi_2)' y_t^{(p)} F(z_{t-d}; \gamma, \alpha, c) + \epsilon_t$$

[Eq.4]

where the ϵ_t process is identically and independently distributed with zero mean and variance equal to σ^2. Now, the linearity test can be derived by approximating the function $F(z_{t-d}; \gamma, \alpha, c) = (1 + \exp(-\gamma(\alpha' z_{t-d} - c)))^{-1}$ using a first-order Taylor approximation around $\gamma = 0$. The resulting auxiliary regression is:

$$y_t = \beta_0' y_t^{(p)} + \beta_1' y_t^{(p)} z_{t-d} + e_t$$

[Eq.5]

where $\beta_i = (\beta_{i,0}, \beta_{i,1}, ..., \beta_{i,p})'$, $i = 0, 1$, and $e_t = \epsilon_t + (\phi_1 - \phi_2)' y_t^{(p)} R_1(z_{t-d}; \gamma, \alpha, c)$ where $R_1(z_{t-d}; \gamma, \alpha, c)$ is the remainder term of the Taylor expansion. The remainder term does not affect the properties of the errors under the null hypothesis. Testing the null hypothesis of $\phi_1 = \phi_2$ on [Eq.1] will be equal to testing the null hypothesis that $\beta_1 = 0$ in [Eq.5]. The test statistic has an asymptotic χ^2 distribution with $p + 1$ degrees of freedom. Since the test statistic is not used to test the original null hypothesis, it is referred to as an LM-type statistic.

In the case that the transition variable z_{t-d} only consists of some lagged value of y_t, the term $\beta_{1,0}z_{t-d}$ should be removed from [Eq. 5] to avoid perfect multi-collinearity. A solution to this problem involves approximating $F(z_{t-d};\gamma,\alpha,c) = (1+\exp(-\gamma(\alpha' z_{t-d} - c)))^{-1}$ using a third-order Taylor approximation. The resulting auxiliary regression can then be expressed as follows:

$$y_t = \beta_0' y_t^{(p)} + \beta_1' y_t^{(p)} z_{t-d} + \beta_2' y_t^{(p)} z_{t-d}^2 + \beta_3' y_t^{(p)} z_{t-d}^3 + e_t$$

[Eq.6]

where $e_t = \epsilon_t + (\phi_1 - \phi_2)' y_t^{(p)} R_3(z_{t-d};\gamma,\alpha,c)$. The original null hypothesis of $\gamma = 0$ can now be approximated using $\beta_1 = \beta_2 = \beta_3 = 0$ and tested using a standard LM-type test. Note that, in order to avoid perfect multi-collinearity in the case where the transition variable is some lagged value of y_t, the terms $\beta_{i,0}s_t^i$, $i = 1,2,3$, should be excluded.

The above specification does not consider the presence of external regressors. Based on the discussion by Ghalanos (2014: 12-13), [Eq.6] can be modified as follows:

$$y_t = \beta_0' \mathbf{X_t} + \beta_1' \mathbf{X_t} z_{t-d} + \beta_2' \mathbf{X_t} z_{t-d}^2 + \beta_3' \mathbf{X_t} z_{t-d}^3 + e_t$$

[Eq.7]

where $\mathbf{X_t} = (\tilde{y}_t^{(p)}, \tilde{x}_t)'$ consists of the vector of lagged values of the variable of interest as well as external regressors.

4 Modelling multiple regimes in tax revenue collections

4.1 Overview of the data

The focus of this paper will be on constructing two-regime smooth transition ARMAX models for the three main tax products: (1) PIT, (2) CIT, and (3) VAT[4]. The data used for the analysis cover quarterly South African cash flow tax revenue (as opposed to tax liability) and their respective tax bases over the period 1994Q1 to 2014Q3. The output gap[5] is also included to capture the cyclical effect of the business cycle on tax revenue collections. A breakdown of the tax products and its respective tax bases are shown in Table 1.

-Insert Table 1 here-

Tax revenue data was obtained from the South African Revenue Service while national accounts data were sourced from the South African Reserve Bank; quarterly observations for compensation of employees and the gross operating surplus were obtained from Statistics South Africa. All variables are nominal given the lack of available price deflators, in addition to not being adjusted for seasonality. It may seem unusual not to seasonally adjust the data series. The argument for not seasonally adjusting the data series for this paper is two-fold. First, administrative seasonality present in tax revenue collections pose challenges in some instances where certain rules have been implemented by the South African Revenue Service which either distorts the seasonal pattern or changes it completely; seasonal adjusting may add noise to the data instead of removing it. Second, based on the paper by Bell and Hillmer (1984), it is often better to identify and estimate seasonal coefficients and ARMA coefficients jointly as opposed to separately. A logarithmic transformation was applied to all data series as is customary when working with volatile and trending data. Seasonal differencing was applied to obtain year-on-year growth. No consideration was given to the impact of legislative (or policy) changes on tax revenue growth, given that tax policy expenditure exercises only take place on an annual basis and are not revised when new information becomes available.

Figure 1 shows year-on-year growth for PIT, CIT and VAT respectively for the period 1994Q1 to 2014Q3. The coincident business cycle indicator is represented by the grey line (second y-axis not shown). Note that

[4] VAT is a composite of domestic VAT, import VAT and VAT refunds.
[5] The output gap is estimated using a simple Hodrick-Prescott filter.

the previous two upward phases of the business cycle occurred from June 1993 to November 1996 and from September 1999 to November 2007 respectively. The last two downward phases took place from December 1996 to August 1999 and from December 2007 and August 2009 respectively. According to SARS (2014: 30), PIT is a tax levied on the taxable income (including taxable capital gains) of individuals and trusts. PIT remains the largest source of tax revenue and constitutes roughly a third of all tax collected. PIT revenue on average grew by approximately 10.6 percent over the measured period with a maximum rate of 22.5 percent achieved in the third quarter of 2007 prior to the global financial crisis; a minimum of -2.5 percent was achieved during the first quarter of 2001. Quarterly growth during the early 2000s were pushed down mainly as a result of the significant tax relief that was granted to individuals, keeping all other factors constant. What is interesting is how closely growth in PIT revenue tracked movements in the business cycle, given that PIT revenue tends to be fairly resilient to business cycle movements.

-Insert Figure 1 here-

CIT is a tax levied on the taxable income of companies and close corporations. CIT has been the third largest tax revenue contributor over the past ten years, having briefly surpassed VAT revenue during 2008/09 prior to the start of the recession in South Africa. Over the measured period, the rate of normal tax on taxable income declined gradually from 40 percent in the first quarter of 1994 to 28 percent during the second quarter of 2008. Some sectors are subject to sector-specific tax dispensations and deductions, examples of which include the gold mining formula as well as accelerated depreciation of capital assets for qualifying sectors. Furthermore, companies classified as small business corporations with a turnover of less than R20 million (for tax years from 2013 onward) are subject to a graduated income tax schedule, while micro businesses with annual turnover of less than R1 million are subject to a graduated income tax schedule with a maximum marginal rate of 6 percent (SARS 2014: 129-130).

Company revenue is clearly more volatile as compared to individual income tax revenue and partly reflects the cyclicality of company earnings. The average quarterly growth in CIT over the measured period was 13.2 percent, with a maximum rate of roughly 99 percent just prior to the 9/11 terrorist attacks and the recession in the US. A minimum of -32 percent was achieved during the first quarter of 2002. An administrative change known as the '80 percent rule' was introduced for all years of assessment beginning on or after 1 March 2008, and requires provisional taxpapers with taxable income of more than R1 million to cover at least 80 percent of their liability in the first two provisional payments. This has reduced the third provisional payment to less than 5 percent of total provisional payments. Despite CIT payments being made on a lagged basis relative to liability, it has generally moved in tandem with the coincident business cycle indicator.

VAT in South Africa is destination-based, which implies that it is payable on the supply of goods and services within South Africa, including imported goods. A standard rate of 14 percent has been levied since the 7th of April 1993 and most products and services are subject to this rate apart from select exempt and zero-rated goods. Companies registered as VAT vendors can be in a net refund position if: (1) Mainly zero-rated goods are produced as output and input tax can be claimed, (2) large capital investments are made which will result in large input tax claims, and (3) selling goods or inventories below cost value. In the 2013/14 tax year there were roughly 662,000 registered VAT vendors, with the majority (63.5 per cent) being active vendors as opposed to being dormant (SARS, 2014: 182-183).

In terms of growth characteristics since 1994, percentage growth in VAT has remained between zero and 20 for roughly 80 percent of the time. The contraction in growth during the third quarter of 2009 as a result of the recession was fairly muted given that the negative performance in domestic VAT and import VAT was partly offset by the significant slowdown in VAT refunded.

4.2 Identification, specification and estimation

A 'specific-to-general' approach as recommended by Granger (1993) for non-linear models will be used to build smooth transition ARMAX models for the three main tax products. Further to this, based on Teräsvirta (1994), a data-based modelling cycle approach will be taken which consists broadly of the following steps: (1) Specify a linear autoregressive model of order p for each tax product, (2) test for the presence of STAR non-linearity, (3) estimate the STAR model, (4) run model diagnostics, (5) make changes if required, and (6) use the model for forecasting and analysis.

As suggested by Van Dijk and Franses (1997: 13), a simple second order autoregressive model can be used as the starting point for constructing two-regime smooth transition ARMAX models. The results from this exercise are summarised in Table 2, 3 and 4.

-Insert Tables 2,3 and 4 here-

It can be observed that a second-order linear autoregressive process represents a good approximation of movements in PIT revenue. All included variables are statistically significant, based on the Ljung-Box test for autocorrelation there is no evidence of significant autocorrelation and the assumption of normality is not rejected based on the Jarque-Bera test for normality. This outcome therefore calls into question whether a two-regime equation would add value in the case of PIT. On the other hand, CIT and VAT revenue are not well approximated using a second-order autoregressive process, as evidenced by the significant autocorrelation. Normality is rejected in both instances due to significant positive excess kurtosis. Further investigation will involve testing explicitly for STAR non-linearity.

To determine whether the null hypothesis of model linearity is valid, a third-order Taylor approximation of the transition function will be used. For this test, a simple logistic STAR model as shown in [Eq.1] is estimated for each tax product, where a combination of lagged tax revenue growth and the estimated output gap serve as the transition variable; the applicable tax base enters exogenously into the equation. The outcome of the non-linear tests are shown in Table 5. Based on the p-values of both the F-statistic and the χ^2 statistic, the null hypothesis of model linearity can be rejected for PIT, CIT and VAT at the 1 percent significance level.

-Insert Table 5 here-

Following rejection of linearity, the specification and estimation of the STARMAX models for each tax product will be performed. Information criteria will be used to guide the specification decision. Based on Ghalanos (2014: 7), STARMAX models are estimated via maximum likelihood, with no inequality restrictions on the regime intercepts or other parameter bound restrictions apart from requiring a positive variance and smoothness parameter γ. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm will be used to solve the unconstrained non-linear optimisation problem.

The estimated model parameters for each tax product are shown in Table 6, 7, and 8. Note that $s1.alpha1$ is not estimated and is fixed to one. In terms of PIT, the fitted STARMAX model appears to provides a good fit to the data. The threshold value c equates to 0.5 percent and is statistically significant. The tax base elasticity of PIT in the low growth regime is 0.4 percent, compared to the 1.5 percent in the second regime. The γ value indicates that the transition from the low growth regime to the high growth regime occurs fairly quick. The appendix contains more information on the transition function and a graphical representation of how well the model fits the actual data.

-Insert Table 6 here-

The fitted model for CIT provides an optimal threshold value of approximately 6 percent. The tax base elasticity during the low growth regime amounts to 0.7 percent, and moves to 2.5 per cent during the high growth regime. This outcome implies that CIT collections become relatively stagnant relative to its approximated tax base during periods of low growth, and become very buoyant during periods of expansion. The larger disparity between the high and low growth tax base elasticities can be most likely attributed to the higher volatility of CIT collections.

-Insert Table 7 here-

A similar picture is obtained when analysing the optimal parameters of VAT. A threshold value of approximately 11 is obtained, while the low growth and high growth tax base elasticities are equal to 0.5 and 1.7 respectively. It can also be observed that, in terms of model volatility, CIT remains the most volatile of the three products analysed, followed by VAT and then PIT.

-Insert Table 8 here-

Making comparisons to previous South African studies are problematic given that there exists only one published paper on asymmetric non-linear elasticities. In addition, the estimates would not be directly comparable given methodological differences. However, based on Table 9 it does appear that the estimated tax base elasticities from the two-regime STARMAX models are generally lower than what was obtained using the autoregressive distributed lag model that was augmented with a smooth transition regression model.

-Insert Table 9 here-

5 Conclusion and policy recommendations

This paper was an attempt to build on the work done by Jooste and Naraidoo (2011) in terms of the asymmetric behaviour of tax revenue in South Africa. Given that tax revenue forecast errors tend to be positive during periods of low growth and negative during upswing phases, it may prove useful to test whether accounting explicitly for business cycle cyclicality improves the forecasting accuracy of tax projections.

Smooth transition ARMAX models were fitted for the main tax products in South Africa following the rejection of linearity. Preliminary results show that PIT, CIT and VAT do respond differently to its respective tax bases depending on the value of the transition variable, which consists of lagged tax revenue growth and the output gap. Generally, the tax base elasticity becomes fairly inelastic during periods of low growth and very buoyant during boom times. Compared to a previous South African study on non-linear elasticities, the coefficients appear to be slightly lower in both regimes. This result has implications in terms of the current tax revenue forecasting process, given that tax revenue cyclicality is not explicitly considered when generating forward-looking estimates. It is therefore recommended that policy makers should base their tax revenue forecasts on a weighted linear combination of the tax base elasticities in the respective regimes, where the weights will be governed by a measure of the output gap and past values of tax revenue.

South African research on the behaviour of tax revenue over the course of the business cycle is currently in its initial stages, and there are many ways of improving upon this paper. Future research will focus on the identification of the most appropriate transition variable and the choice between the logistic form of the STARMAX model and the exponential form. The option to include multiple regimes (i.e. more than two) will also be investigated. In addition, the existence of cointegrating relationships will be tested. Additional model dynamics that will be investigated include: specifying a dynamic variance model using a GARCH implementation; and the option to include an autoregressive term in the state dynamics equation based on work done by Kauppi and Saikkonen (2008) and Nyberg (2010).

6 References

Calitz, E., Siebrits, K. and Stuart, I. (2013), "The accuracy of fiscal projections in South Africa," *Paper presented at the 2013 ESSA Conference.*

Chan, K.S. (1993), "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *The Annals of Statistics*, 21, 520-33.

Chan, K.S. and Tong, H. (1986), "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 178-190.

Du Plessis, S. and Boshoff, W. (2007), "A fiscal rule to produce countercyclical fiscal policy in South Africa," *Stellenbosch Economic Working Paper Series* 13/07.

Enders, W. (2010), *Applied econometric time series.* 3rd edition. New Jersey: Hoboken.

Granger, C.W.J. (1993), "Strategies for modelling non-linear time series relationships," *The Economic Record*, 69, 233-238.

Granger, C.W.J., and Andersen, A.P. (1978), *An Introduction to Bilinear Time Series Models*, Vandenhoek and Ruprecht: Gottingen.

Hamilton, J.D. (1994), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Jooste, C. (2009), "Estimating tax elasticities: The case of corporate income tax, personal income tax and value-added tax," [Online] Available from: www.africametrics.org/documents/conference09/papers/Jooste.pdf

Jooste, C. and Makrelov, K.H. (2012), "Forecasting tax revenues using a non-linear econometric approach," *National Treasury Research Note* (Unpublished).

Jooste, C. and Naraidoo, R. (2011), "Non-linear tax elasticities and their implications for the structural budget balance," *The Journal of Applied Business Research*, 27, 113-126.

Kauppi, H. and Saikkonen, P. (2008), "Predicting US recessions with dynamic binary response models," *The Review of Economics and Statistics*, 90, 777-791.

Leal, T., Pérez, J.J., Tujula, M. and Vidal, J. (2007), "Fiscal forecasting: Lessons from the literature and challenges," *ECB Working Paper Series* 2007/843.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988), "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491-499.

Nyberg, H. (2010), "Dynamic probit models and financial variables in recession forecasting," *Journal of Forecasting*, 29, 215-230.

Swanepoel, A.S. and Schoeman, N.J. (2002), "Tax revenue as an automatic fiscal stabiliser: A South African perspective," *South African Journal of Economics*, 5, 566-588.

Teräsvirta, T. (1994), "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, 89, 208-218.

Teräsvirta, T. (1998), "Modelling economic relationships with smooth transition regressions," in *Handbook of Applied Economic Statistics*, ed. A. Ullah, & D.E.A. Giles: New York.

Tong, H. (1978), "On a threshold model," in *Pattern Recognition and Signal Processing*, ed. C.H. Chen, Sijhoff & Noordhoff: Amsterdam.

Tong, H. and Lim, K.S. (1980), "Threshold autoregression, limit cycles and cyclical data," *Journal of the Royal Statistical Society*, 42, 245-292.

Tsay, R.S. (2002), *Analysis of financial time series*, 1st edition, New York: John Wiley & Sons

Van Dijk, D. and Franses, P.H. (1999), "Modelling multiple regimes in the business cycle," *Macroeconomic Dynamics*, 3, 311-340.

Van Dijk, D. and Franses, P.H. (2000), *Non-linear time series models in empirical finance*, 6th edition, New York: Cambridge University Press.

Van Dijk, D., Teräsvirta, T. and Franses, P.H. (2000), "Smooth transition autoregressive models – A survey of recent developments," *Econometric Reviews*, 21, 1-47.

Wolswijk, G. (2007), "Short and long run elasticities: The case of the Netherlands," *European Central Bank Working Paper* 2007/763.

Figure 1a: PIT(y/y%)

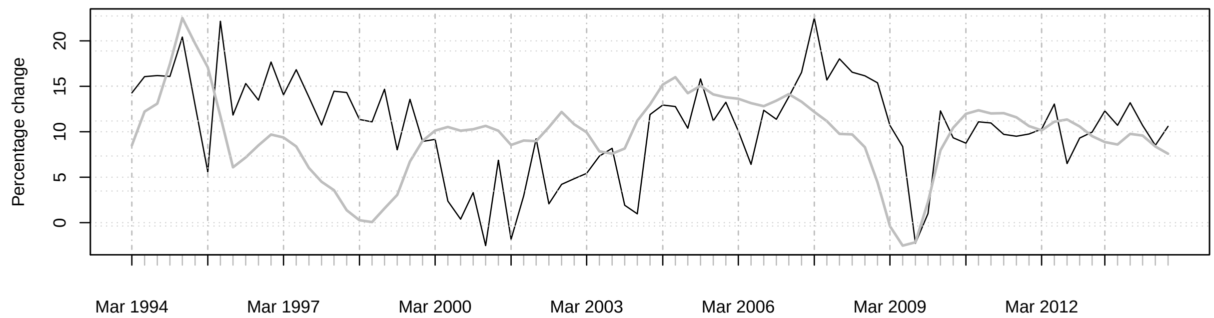

Figure 1b: CIT

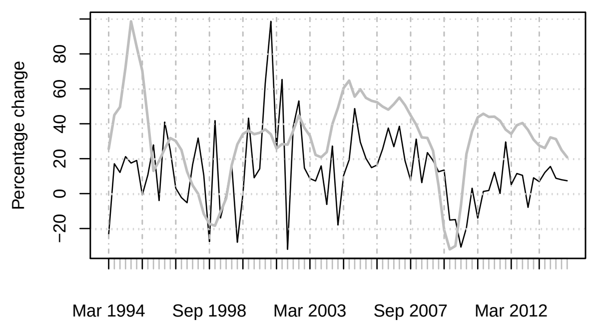

Figure 1c: VAT

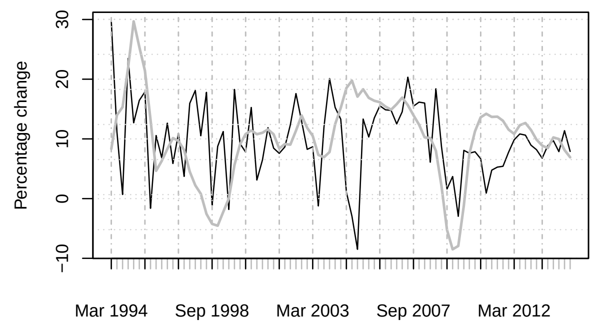

Table 1: Tax products and their respective tax bases

Tax.type	Tax.base	Source
PIT	Compensation of employees	StatsSA
CIT	Gross operating surplus	SARB
VAT	Final household consumption	SARB

Table 2: Second-order linear autoregression: PIT

	Estimate	Std.Error	t.value	p.value
phi.0	0.10	0.02	6.62	0.00
phi.1	0.42	0.11	3.86	0.00
phi.2	0.28	0.11	2.61	0.01

Table 3: Second-order linear autoregression: CIT

	Estimate	Std.Error	t.value	p.value
phi.0	0.14	0.04	3.76	0.00
phi.1	0.16	0.11	1.42	0.16
phi.2	0.19	0.11	1.74	0.09

Table 4: Second-order linear autoregression: VAT

	Estimate	Std.Error	t.value	p.value
const	7.24	1.50	4.82	0.00
phi.1	0.16	0.11	1.47	0.15
phi.2	0.07	0.10	0.72	0.48

Table 5: LM-type test for STAR non-linearity

	F.stat	p.value	Chisq.stat	p.value
PIT	8.52	0.00	34.05	0.00
CIT	3.18	0.01	17.03	0.01
VAT	4.69	0.00	12.72	0.01

Table 6: PIT optimal parameters

	Coefficient	Std.error
s1.phi0	-1.80	2.00
s1.phi1	0.78	0.11
s1.xi1	0.41	0.24
s1.psi1	-0.39	0.17
s2.phi0		
s2.phi1	-1.71	0.33
s2.xi1	1.45	0.11
s2.psi1		
s1.gamma	89.14	6.37
s1.c	0.46	0.01
s1.alpha1	1.00	
s1.beta	-0.25	0.00
sigma	3.65	0.34

Table 7: CIT optimal parameters

	Coefficient	Std.error
s1.phi0	0.52	5.31
s1.phi1	0.90	0.28
s1.xi1	0.66	0.52
s1.psi1	-1.19	0.31
s2.phi0	-31.55	12.75
s2.phi1	-0.04	0.30
s2.xi1	2.47	1.27
s2.psi1	-2.00	0.42
s1.gamma	20.77	2.96
s1.c	5.85	0.07
s1.alpha1	1.00	
s1.beta	-0.44	0.00
sigma	17.44	2.02

Table 8: VAT optimal parameters

	Coefficient	Std.error
s1.phi0	-6.12	3.01
s1.phi1	0.82	0.09
s1.xi1	0.50	0.30
s1.psi1	-1.66	0.25
s2.phi0	7.76	2.49
s2.phi1	-1.21	0.01
s2.xi1	1.67	0.22
s2.psi1	2.00	0.13
s1.gamma	0.19	0.04
s1.c	10.65	2.39
s1.alpha1	1.00	
s1.beta	-0.83	0.00
sigma	4.51	0.43

Table 9: Non-linear tax elasticities

	Jooste,Naraidoo(2011)	Boonzaaier(2015)
s1.pit	0.88	0.41
s1.cit	0.86	0.66
s1.vat	0.81	0.50
s2.pit	1.87	1.45
s2.cit	2.77	2.47
s2.vat	2.18	1.67