



Petrie, J. R. (2020) SGLT2 inhibitors and renal complications in type 1 diabetes. *Lancet Diabetes and Endocrinology*, 8(10), pp. 803-805. (doi: [10.1016/S2213-8587\(20\)30311-9](https://doi.org/10.1016/S2213-8587(20)30311-9)).

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

This is the peer reviewed version of the following article:  
Petrie, J. R. (2020) SGLT2 inhibitors and renal complications in type 1 diabetes. *Lancet Diabetes and Endocrinology*, 8(10), pp. 803-805, which has been published in final form at [10.1016/S2213-8587\(20\)30311-9](https://doi.org/10.1016/S2213-8587(20)30311-9). This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#).

<http://eprints.gla.ac.uk/180711/>

Deposited on: 26 November 2020

# Functional regression models for South African economic indicators: A growth curve perspective

---

## Abstract

In this paper we compare the performance of three functional regression models from a growth curve perspective to predict the relationship between two economic variables. Specifically our focus is on the functional concurrent model, functional historical model and the functional autoregressive model (FAR). The dependent and the independent variables are cumulated over the annual time window for the growth curve analysis. These models of interest are used to predict exports (real) for the South African economy in terms of electricity demand (or electricity consumption). The data analysed consists of 31 years of exports (in ZAR million) at annual quarterly frequency and the electricity demand (in GwH) are monthly totals. Exploratory analysis involved plotting phase-plane plots for the two series. For the prediction exercise, the baseline concurrent model was evaluated against the historical model and the FAR model. The performance of the models are compared using the root-mean-square measure, and the relative performance in terms of the ratio of the RMSEs. Our results show that from the growth curve perspective, for the prediction of exports in terms of electricity for the SA economy, the FAR model of lag(1) outperforms both the concurrent model and the historical model by far. While the historical model performed better than the concurrent model that only relies on drivers within the same time window.

### *Keywords:*

Functional data analysis, growth curves, functional regression, electricity demand, exports, forecasting.

---

## 1. Introduction

In this paper, three novel competing functional growth curve regression models are explored, namely, the functional concurrent model, the functional historical model and the functional autoregressive model to investigate the relationship between electricity demand and the real exports of goods and services index (simply referred to as exports, hereafter). Both electricity demand and exports are vital indicators of economic growth, and it is useful to ascertain how the changes in one variable, in our case electricity demand, affects the changes in the other variable, in our case exports. We use novel features in Functional Data Analysis (FDA)

methods to compare the within year seasonal changes and fluctuations in the two series, and to forecast the dependent variable. In FDA, data is modelled as functions rather than as data points. The researchers in this paper do not assume causality or prove causality between the two variables subjected to investigation here, but are interested in fully evaluating the relationship between exports and electricity demand for the period under study. It is our view that the standard regression methods do not focus on the differences between inter-annual seasonal variations. Also, it has been argued that seasonal variations are an integral component of the continuous time dynamics represented in certain data sets, and that seasonality should not be treated as a “nuisance component” that must be eliminated, as is the common assumption in the analysis of economic time series in which the calculation of seasonally adjusted data is an important aspect (Ramsay and Ramsey, 2002). It is our hypothesis too, that these inter-annual seasonal variations are important and should not be ignored.

South Africa’s electricity demand and exports have been on an increasing trend since the early 1990’s (see Figure 1). This is to be expected as South Africa experienced huge political and economic changes during 1990-1994, with the South African economy being opened to the rest of the world post 1994. The economy is maintained and driven by energy, which in South Africa continues to be primarily electricity. So, electricity demand is vital to economic growth in South Africa. Indeed, it has been shown by Ferguson et al. (2000) in a study of over one hundred countries constituting over 99% of the global economy that electricity consumption is correlated with wealth creation. Ferguson et al. (2000) do not prove causality, but do “show that, in the technological world of the late 20th Century, economic development occurred ‘hand in hand’ with electricity consumption and, in particular, with an increase in the proportion of energy used in the form of electricity”. It is our view that the same can be said of exports as they bring in much needed foreign revenue and directly contribute to job creation and economic growth in South Africa.

### *1.1. Literature review of the problem*

In analysing the relationship between any two economic variables, time series analysis methods are very popular and the theoretical literature is well established. To our knowledge, literature investigating relationships between electricity demand and economic growth, and between exports and economic growth are limited. Exports are a major contributor to the Gross Domestic Product (GDP), an acceptable measure of economic growth (according to World Bank data, exports accounted for just less than 30% of South Africa’s GDP in 2016). Also, most literature is aimed at more than just examining the relationship between the variables, but at also investigating causality between the two variables. Narayan and Smyth (2009) note that there is sizable literature on Granger causality between energy consumption and GDP, and between exports and GDP. In their study, Narayan and Smyth (2009) used Granger causality tests to ascertain the relationship between electricity consumption, exports and GDP for a panel of six

Middle Eastern countries. They found that there are positive effects of electricity demand and exports on economic output and reported that a 1% increase in electricity consumption increases GDP by 0.04%, while a 1% increase in exports increases GDP by 0.17%. They suggested that policymakers must increase electricity production infrastructure to improve economic growth, and encourage exports to promote economic growth. Narayan and Smyth (2008) found similar results when they investigated the relationship between energy consumption and real GDP in a panel of the G7 group of countries. Sami and Makun (2011) investigated the relationship between electricity consumption, GDP and exports in Brazil. They found that exports have a positive impact on real income, and that electricity consumption has a directly positive effect on economic growth. Sami (2011) found that there is evidence that exports, electricity consumption and real income per capita are cointegrated in the Japanese economy. Lean and Smyth (2010) found that there is unidirectional Granger causality running from economic growth to electricity consumption in Malaysia. However, there does not seem to be much literature investigating specifically the relationship between exports and electricity demand, and as such, given that these two are intrinsically important indicators of economic activity and growth, especially for South Africa, our focus here is on these two variables.

Standard multiple regression techniques have been used (Agalega and Antwi (2013); Babalola et al. (2015); Divya and Devi (2014); and Syed and Shaikh (2013)) to analyse the relationship between economic growth and other covariates. For example, Divya and Devi (2014) did a study to ascertain the relationship and impact of certain indicators on the GDP of the Indian economy, with a focus on identifying factors which may provide early signals regarding the growth of the economy. The study made use of the Coefficient of Correlation and Analysis of Variance to study the relationship between the variables, and of Multiple Linear Regression with GDP as the dependent variable to study the impact of the covariate variables on GDP. The chosen independent variables were inflation, exchange rate, foreign exchange reserves, FII's (Foreign Institutional Investors), Sensex (a free-float market-weighted index of 30 well established and financially sound companies in the Bombay Stock Exchange), balance of payments and fiscal deficit. They concluded that while inflation is highly correlated with GDP, it does not significantly influence the GDP. The study revealed that the exchange rate, Sensex and Balance of Payments are better predictors of GDP.

In the South African case, Odhiambo (2009) used Granger causality tests to ascertain the relationship between electricity consumption and economic growth. He found that there is a (bidirectional) relationship between the variables. Wolde-Rufael (2006) did a study on the relationship between electricity consumption and economic growth for 17 African countries. He found that there was a long-run relationship between the variables, although no evidence of causality. Wolde-Rufael (2006) concluded that their results showed that past values of economic growth have predictive ability in determining future values of electricity

consumption for some countries. The opposite is true for some other countries, past values of electricity consumption have predictive ability in determining future values of economic growth. Inglesi-Lotz and Blignaut (2011) used decomposition methods to identify factors that led to the increase in South African electricity consumption for the period 1993-2006. They found that the production effect, that is the change in production, contributed more to changes in electricity consumption than other effects.

### *1.2. Literature review of Functional Data Analysis*

The term “functional data” refers to data where each observation is a curve. The individual datum in FDA is a whole function defined on a common interval, usually time. Although the original data is measured at discrete intervals, measurements could, in theory, be taken at any time during the period of analysis, the discrete points are then considered finite instances of an underlying data generating process. It is therefore natural to analyse the data as if it were defined over a continuum. A major underlying assumption in FDA methods is that the function is differentiable to some order. Levitin et al. (2007) defines FDA as a set of quantitative methods which can answer a number of questions that traditional statistics cannot, such as FDA allows the researcher to ask questions about when in a time series differences might exist. Functional data has in the recent past become important because of improved technological capability in the collection and analyses of high-dimensional data by assuming that the data are continuous. Although classical statistical methods can be used in the analysis of such data, they do not take advantage of the additional information that could be implied by the smooth underlying function and the insights gained from exploring the derivatives of the curves.

FDA is highly flexible in that the timing of the intervals for data observations do not have to be equally spaced (Muller, 2011). Also, FDA methods are not necessarily based on the assumption that the values observed at different times for a single object are independent. Ullah and Finch (2013) notes that since FDA treats the whole curve as a single entity, there is no concern about correlations between repeated measurements, while Wang et al. (2016) notes that any measurement errors which occur during the data observation or collection stage can be viewed as random fluctuations around a smooth trajectory. FDA can accommodate measurement errors easily because for each subject, one observes repeated measurements. Regarding the issue of seasonality in the data, Ramsay and Ramsey (2002) argues that seasonal variations are an integral component of the continuous time dynamics represented in the data, and that seasonality is not a “nuisance component” that must be eliminated, as is a common assumption in the analysis of economic time series in which the calculation of seasonally adjusted data is an important aspect. Further, the assumption that the components of times series are additive, and that the data must be seasonally adjusted, is critical in time series analysis and to the deseasonalising procedures used (Ramsay and Ramsey, 2002). The class of models used to dealing with time series with seasonal components has been previously criticised by Harvey and Jaeger (1993) and Harvey and Scott (1994)

as being misleading and uninformative. It should also be noted that many time series methods are primarily based on the idea of stationarity and some data violate this assumption. Muller (2011) noted that with FDA, stationarity of the underlying process is not needed. As such, FDA is in some way, a set of techniques that overcomes the limitations of traditional time series approach and extends classical statistical methods to the functional framework, without having to first extract statistical indexes out of the data, a process which can lead to information loss (Muller, 2011). FDA methods are also essentially nonparametric, and thus allow for more attractive flexible modelling (Muller, 2011). Applying FDA to a data series involves two main steps: First, data preparation – which consists of transforming the data series into functional form (e.g B-splines); Second, data analysis – many techniques from classical and multivariate statistics have been extended to the functional form (e.g. Functional Principal Component Analysis and Functional Linear Regression).

In this paper, we focus on analysing the derivatives of the data series as well as apply three functional growth curve regression models to explore the relationship between two important economic indicators of the South African economy, namely exports (real) and electricity consumption. Our main aim is to get insights into how electricity consumption may help predict exports. As mentioned before, electricity is the primary source of energy in South Africa for both industrial and domestic use. Literature on FDA analysis of economic data is quite sparse with only two of 84 papers identified by Ullah and Finch (2013) as using FDA methods for analysis being economic papers. One being the well cited paper of Ramsay and Ramsey (2002) who analysed the US monthly non-durable goods index and in which the seasonal variations embedded within the data were analysed as intergral component of the model. The second of the 84 papers was by Kneip and Utikal (2001) in which they modelled the income distribution of households in Great Britain for the years 1968-1978. No analysis of economic data for the South African economy has been undertaken to the best of our knowledge using functional data methods, and as such we see our contribution as adding to the body of this literature, in this instance focusing on the two time series of interest, though the methods dealt with in the paper can be adapted and applied to other time series as well. The remainder of the paper introduces the data to be analysed in Section (2), the three competing growth curve functional regression models are detailed in Section (3), in Section (4) we present the results from the interrogation of the data using the proposed methods and we conclude our exercise in Section (5).

## 2. Data

The investigation in this paper focuses on South African exports (real) and aims to interpret it in a functional framework in terms of electricity demand. Exports (in ZAR million) is available in quarterly frequencies. Exports of goods and services comprise all transactions between residents of a country and the rest of the world involving a change of ownership from residents to non-residents of general merchandise (IMF,

2013). Although the exports data was obtained for the period from Q1-1960 to Q4-2015 from the South African Reserve Bank’s website<sup>1</sup>, for this investigation we limit to the period Q1-1985 to Q4-2015 to match the period of the electricity data.

Electricity demand (in GwH) is available in monthly frequency, and our data ranges from M1-1985 to M12-2015. The electricity data was obtained from the Statistics South Africa report “Electricity generated and available for distribution (201508)”<sup>2</sup>. The specific series analysed was the series “ELEKTR10 - Electricity generated and available for distribution”. An assumption is made that the data represents the electricity consumption in the country, as producers would adjust productions so as to minimise wastage as a result of over production. This assumption is based on the fact that Eskom is a monopoly supplier, and supplies 96% of South Africa’s electricity consumption (Baker, 2015). The reader may note that the two series are in two different frequencies. The raw time series plot of the two data series is provided in Figure 1. The graphs are for the time period 1985 to 2015 as this is the common time frame between the two data sets, and will be the period that will be the focus of this paper. It can be seen that both series had an increasing trend for the period 1985 to about 2008, albeit with seasonality embedded within, especially in the electricity demand time series. A sharp decrease for exports is observed around 2008, possibly capturing the global financial crisis of that time. According to Verick and Islam (2010), South Africa was one of the countries worst hit by the 2007/2008 global economic crisis, experiencing an economic recession as a result. This crisis began with a housing bubble in the USA, and spiralled into one of the worst recessions in history (Verick and Islam, 2010).

The data is measured and stored in discrete form and thus needs to be smoothed into a function. Smoothing is the first step in FDA, and its main purpose is to convert raw discrete data points into a smooth function (so called data preparation). This is done to emphasise patterns in the data, by minimising short-term fluctuations due to observational errors (Ullah and Finch, 2013). For this paper, the data was smoothed using B-splines. In the study by Ullah and Finch (2013), a B-spline smoothing technique was proposed as the most popular method, as it is a rather simple method to use. The choice of a smoothing technique depends on the underlying behaviour of the data, but the *smoother* should have features that match those of the data. For example, Fourier basis for periodic data and spline basis for non-periodic data is the general recommendation (Ullah and Finch, 2013).

### 2.1. What is the B-spline?

A B-spline is a piecewise polynomial function constructed by dividing the interval of observation into subintervals with boundaries at points called break points or knots. Spline functions, by definition, approximate the shape of a

---

<sup>1</sup>It is calculated according to stipulations from the sixth edition of the International Monetary Fund’s “Balance of Payments Manual” (IMF, 2013).

<sup>2</sup>[http://www.statssa.gov.za/?page\\_id=1854&PPN=P4141&SCH=7189](http://www.statssa.gov.za/?page_id=1854&PPN=P4141&SCH=7189)

curvilinear stochastic function without the necessity of pre-specifying the mathematical form of the function. That is, it is unnecessary to restrict the estimate to a straight line, a polynomial of pre-specified degree, an exponential, or any other particular form (Suits et al., 1978). Spline functions are defined as piecewise polynomials of degree  $q$ . The pieces join at the endpoints (knots), and must fulfill continuity conditions for the function itself, with a spline function of degree  $q$  being a continuous function with  $q - 1$  continuous derivatives (Wold, 1974). In Suits et al. (1978), the number of basis functions is equal to the sum of the order of the functions and the number of knots. A functional datum would then be a linear combination of the basis functions.

Eilers and Marx (1996) defines the few steps required to create a B-spline function in the following manner. If we let the domain from  $x_{min}$  to  $x_{max}$  be divided into  $k'$  equal intervals by  $k' + 1$  knots. Each interval must be covered by  $q + 1$  B-splines of degree  $q$ . The total number of knots for construction of the B-splines will be  $k' + 2q + 1$ . And the number of B-splines in the model is  $k = k' + q$ . Now, if we let  $B_j(x; q)$  denote the value at  $x$  of the  $j$ th B-spline of degree  $q$  for a given grid of knots, then the fitted curve  $\hat{y}$  to data  $(x_i, y_i)$  is the linear combination

$$\hat{y}(x) = \sum_{j=1}^k \hat{c}_j B_j(x; q), \quad (1)$$

where  $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_k$  (Eilers and Marx, 1996).

Our objective is to produce curves that are not too locally variable. Our two data sets, electricity demand and exports, fit this description well. Hence the decision to use the simple B-splines as a smoother for both of them, with the knots placed as and when required using trial and error.

### 3. Methodology

To model dependency for the two data series, traditional regression models have their limitations because firstly, the time points in the two series have to be in the same frequency, unlike in this study where electricity is recorded in monthly frequencies and the exports is recorded in quarterly frequencies; and secondly, both data series are expected to be a function of South African economic growth, meaning they are highly correlated. In this section, the notion and usefulness of the growth model is discussed within the functional analysis framework. Also a discussion on phase-plane plots, a useful method in the exploratory analysis of functional data, follows. Finally, the functional regression growth curve models used in the analysis of our data are discussed in detail, namely the concurrent model, the historical model and the autoregressive model.

#### 3.1. Growth curves

Growth curve models typically refer to statistical methods that allow for the estimation of inter-individual variability in intra-individual patterns of change over



time (Curran et al., 2010). Here the interest is on the inter-annual variability embedded within economic data. As such, the annual curves in our data are considered from a growth perspective which essentially means that the quarterly (exports) and monthly (electricity) values are cumulated for each year. That is for the quarterly series, for each year, the first time point corresponds to Q1 value, the second time point corresponds to Q1+Q2 value, the third time point corresponds to Q1+Q2+Q3, and finally the fourth time point corresponds to the value Q1+Q2+Q3+Q4. Similarly for the monthly series. This gives us annual functions which are, at the very least, monotonically increasing. Thus, an analysis of the changes in the growth patterns of the series can be undertaken. In addition, the analysis of the velocity (first derivative) and acceleration (second derivative) of the growth curves will provide further insights into the analysis.

### 3.2. Phase-plane plots as an Exploratory Analysis Tool

As has been mentioned earlier, the smoothness of a curve implies that derivatives can be computed from the data series. This immediately provides the opportunity to investigate the rates of change in the series. The first derivative indicates rate of change (velocity) while the second derivative indicates curvature of the function (acceleration). A phase-plane plot is a graph of the second derivative (curvature or acceleration) of the fitted function against its first derivative (slope or velocity) (Ramsay and Ramsey, 2002). The phase-plane plot is used to explore the within-period (here year) patterns in a given time series. Considerable information is gained by plotting acceleration against its rate of change and is interpreted as the element of energy exchange within the economic system under consideration for each series.

### 3.3. Functional Regression Methods

As an extension of (multiple) linear regression, it is possible to explain the variability of a functional variable in terms of the variability of covariates which may or may not be functions themselves as follows (Ramsay et al., 2009):

- where the dependent variable is a scalar and the independent variable(s) are functions.
- where the dependent variable is a function and the independent variables are scalars.
- where all variables are functions, which is discussed below.

#### 3.3.1. Functional Regression: Concurrent Model

The general functional linear model is defined by:

$$y_i(t) = \beta_0(t) + \sum_{j=1}^q \beta_j(t) x_{ij}(t) + \epsilon_i(t) \quad (2)$$

where  $x_{ij}(t)$  are  $q$  functional observations which are the covariate variables and  $y_i(t)$  is the  $i$ th response curve. Even if  $x_{ij}$  is a scalar observation, it is simply treated as a function that is continuous over  $t$ . It is to be noted that  $\beta_0(t)$  and the  $\beta_i(t)$ , the regression constant and coefficients respectively, are smooth functions that can be interpreted the same way as multiple regression, over time  $t$ . As aforementioned, in general, neither the independent variable nor the dependent variables are required to be functions.

To compute the regression coefficients, we use the method of Ramsay and Ramsey (2005) and Ramsay et al. (2009). Let the  $N$  by  $q$  vector function matrix  $\mathbf{Z}$  contain the  $x_{ij}$  functions, and let the vector  $\beta$  of length  $q$  contain each of the regression functions. The concurrent functional linear model, in matrix notation, is thus defined as

$$\mathbf{y}(t) = \mathbf{Z}(t)\beta(t) + \epsilon(t) \quad (3)$$

where  $\mathbf{y}$  is a function vector of length  $N$  containing the response functions. A basis function for each of the regression functions  $\beta_i$ ,  $i = 1, 2, \dots, q$  must be estimated. The regression functions  $\beta_i$  have the expansion

$$\beta_i(t) = \sum_k^{K_i} b_{ki}\theta_{ki}(t) = \theta_i(t)'\mathbf{b}_i(t), \quad (4)$$

in terms of  $K_i$  basis functions  $\theta_{ki}$ . Now,  $K_\beta = \sum_i^q k_i$ , thus a vector  $\mathbf{b}$  of length  $K_\beta$  can be constructed by stacking the vectors vertically,  $\mathbf{b} = (\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_q)$ .

Now the  $q$  by  $K_\beta$  matrix function  $\Theta(t)$  is given by:

$$\Theta(t) = \begin{bmatrix} \theta_1(t)' & 0 & \dots & 0 \\ 0 & \theta_2(t)' & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \theta_q(t)' \end{bmatrix}$$

It must be noted then that  $\beta(t) = \Theta(t)\mathbf{b}$ . Thus the model in equation (3), can be expressed as:

$$\mathbf{y}(t) = \mathbf{Z}(t)\Theta(t)\mathbf{b} + \epsilon(t) \quad (5)$$

and the vector of residuals can be expressed as  $\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{Z}(t)\Theta(t)\mathbf{b}$ . It should be apparent from the above that to get the coefficient functions we simply need to estimate  $\mathbf{b}$ .

Now,

$$\hat{\mathbf{b}} = \frac{[\int \Theta'(t)\mathbf{Z}'(t)\mathbf{y}(t)dt]}{[\int \Theta'(t)\mathbf{Z}'(t)\mathbf{Z}(t)\Theta(t)dt + \mathbf{R}(\lambda)]}, \quad (6)$$

where  $\mathbf{R}(\lambda)$  is a symmetric block diagonal matrix included to take care of the roughness (penalties) of the respective coefficients. A solution to the above can be found using numerical integration. Substituting  $\hat{\mathbf{b}}$  into equation (4) gives us the estimated coefficients. For a detailed explanation of the full model and estimation, the reader is referred to Ramsay and Ramsey (2005) and Ramsay et al. (2009).

### 3.3.2. Functional Regression: Historical Model

The concurrent model has a disadvantage because it assumes that  $y(t)$  and  $x(t)$  follow the traditional linear regression model. However, a situation exists where covariates effects may change with time  $t$ , that is when the value of  $y$  at time  $t$  depends not only at the current value of  $x(t)$ , but also depends on the historical values of  $\{x(s) : s \leq t\}$  (Wang et al., 2016). The historical model aims to capture this scenario. The historical model, defined by Malfait and Ramsay (2003), Ramsay and Ramsey (2005) and Kosiorowski (2014) is described as follows:

$$y_i(t) = \beta_0(t) + \int_0^t x_i(s) \beta(s, t) ds + \epsilon(t), \quad (7)$$

where  $\beta(s, t)$  is a bivariate regression coefficient function which defines the dependence of  $y_i(t)$  on covariate  $x(s)$ . Thus for the historical model, at any given time  $t$ , the value of  $y(t)$  depends on the concurrent as well as the past values of  $x$ . It may be pointed out that the concurrent model is a special case of the historical model.

Estimating the historical model is somewhat analogous to the concurrent model. We consider the simpler case when there is only one covariate. We consider the method of Ramsay and Ramsey (2002), Malfait and Ramsay (2003) and Ramsay and Ramsey (2005). The regression function  $\beta_i$  has the basis expansion:

$$\beta(s, t) = \sum_k^{K_\beta} b_k \theta_k(s, t) = \theta'(s, t) \mathbf{b}. \quad (8)$$

A simplification of model (7) is effected by integrating out  $s$  this way

$$x_{ik}^*(t) = \int_{\Omega_t} x_i(s, t) \theta_k'(s, t) ds, \quad (9)$$

and thus equation (7) can be rewritten as

$$y(t) = \mathbf{X}^*(t) \mathbf{b} + \epsilon(t) \quad (10)$$

The other steps in the estimation are analogous to those of the concurrent model, save to say,  $\hat{\mathbf{b}}$  can be solved as

$$\hat{\mathbf{b}} = \frac{\int \mathbf{X}(t)^* \mathbf{y}(t)}{[\int \mathbf{X}(t)^* \mathbf{X}(t)^* + \mathbf{R}(\lambda_s) + \mathbf{S}(\lambda_t)]}. \quad (11)$$

It should be noted that the functions  $\mathbf{R}(\lambda_s)$  and  $\mathbf{S}(\lambda_s)$  are functions which are introduced to take care of the roughness (penalties) of the coefficients. The interested reader is once again referred to Malfait and Ramsay (2003), Ramsay and Ramsey (2005) and Kosiorowski (2014) for the complete derivation of the estimators.

### 3.3.3. Functional Regression: Autoregressive Model

The functional autoregressive model (FAR) is a special case of the concurrent model in so far as the independent variable is a lag of the dependent variable. We define a FAR model of lag 1 (FAR(1)) as:

$$y_i(t) = \beta_0(t) + \sum_{j=1}^q \beta_j(t) y_{(i-1)j}(t) + \epsilon_i(t). \quad (12)$$

The autoregressive model is a special case of the concurrent model. And thus the estimation of the coefficient function follows the same method as that used in the concurrent model.

The concurrent model in our paper is the bench-mark model against which the functional historical model and the functional autoregressive model are evaluated.

### 3.4. Measures of Goodness of Fit

Malfait and Ramsay (2003) suggests a time related measure of fit which is the root mean squared error as a function of time denoted by  $RMSE(t)$ , defined as

$$RSME(t) = \sqrt{SSE_1(t)/N} \quad (13)$$

where  $SSE(t) = \sum_{i=1}^N \{y_i(t) - \hat{y}_i(t)\}^2$ . To compare the performance of our three models we will use the RMSE(t) ratio of competing model to the functional concurrent regression model. If the RMSE ratio is  $> 1$  then the competing model performs worse than the functional concurrent model, and if the RMSE ratio is  $< 1$  then the competing model performs better than the functional concurrent model.

Further, we use the one step-ahead approach to compare the performance of the three models. We start by using the period 1985-2010 to estimate, and forecast the export curve for 2011, 2012, ..., 2015 individually. Thereafter, we include 2011 in the estimation period, and forecast export curve for the remaining years 2012, 2013, ..., 2015 individually, and so forth.

All analysis was carried out using R Core Team (2013) software, in particular the two R packages by *fda* and *fda.usc* (Ramsay et al., 2017; Febrero-Bande and Oviedo de la Fuente, 2012), and their dependencies, within the R-Studio environment.

## 4. Results and Discussions

### 4.1. Phase-plane plots

The phase-plane plots in Figures 2 and 3 depict the acceleration against velocity of the functional curves, and aims to capture the energy in the series visually, especially capturing the variation in the seasonal component of the series.

This introduces an element of energy exchange in the economic system. With potential energy (in vertical axis) representing the volume of goods (electricity) that the system can produce. Kinetic energy represents the goods (electricity) that are

produced, when they are now being used. To read the plot, one must find the lower-case “j” in the plot, and move around the diagram clockwise, noting that the letters indicate the months as you go. For electricity, for most of the years the phase-plane plots have a cyclical shape. The electricity demand plots have three cycles in a year (see in Figure 1 the 2001 plot for a typical electricity demand phase-plane plot), showing positive kinetic energy (electricity use) in the middle months of the year and eventually decreasing to negative values by the end of the year. The two small cycles begin in March and July, and end in June and September, respectively. The plot finally ends with a larger cycle and negative velocity in December. The phase-plane plot confirms that most electricity is consumed around the middle of the calendar year, during the winter months. Both derivatives are zero around May signifying the start of the winter months.

The phase-plane plots of the 1990’s show that the two small cycles were insignificant during those years, but by the early 2000’s, as the South African economy increased, the cycles seemed more significant. The phase-plane plots of 2007 and 2008 once again show the decreasing consumption of electricity, capturing the economic crisis of that era affected consumption. The 2008 plot also seems to highlight the response to the loadshedding which took place during the winter months of 2008.

The exports phase-plane plots are similar for most years, showing strong velocity, with two cycles in each year, a large one from July to December and a smaller one from February to July. The phase-plane plots are mostly similar, the increasing trends emphasised by observing the increasing cyclical pattern of the plots from 1994 to 2014 using the same axis.

#### *4.2. Functional Growth Curve Regression Model Results*

Table 1 show the results of RMSE for the base model (i.e. the concurrent model). While Table 2 shows the relative RMSE for the historical and AR(1) models (that is, relative RMSE to the RMSE of the base model). The first column shows that the data for the years 1985 to 2010 was used to build the model. Then a 1-step ahead forecast was computed (i.e. 2011 forecast), then a 2-step ahead forecast (i.e. 2012 forecast), etc. Surprisingly, the historical model is the worst model for prediction. Also, the results indicate that the 1-step ahead forecast is better if one chooses to use the concurrent and historical models.

In addition, the results show that both competing models outperform the baseline functional concurrent regression model conclusively. The results also reveal that the functional autoregressive model is the best model for the 1-step ahead prediction of exports annual curve for all the years from 2011 to 2015 with electricity consumption as the driver. Also, the difference between the 1-step ahead forecast and the other forecasts is negligible, that is, the differences between the different forecast horizons is small.

## 5. Concluding Remarks

In this paper our intent was to explore the strengths of functional data analysis and associated growth curve regression models to analyse two economic indicators of the South African economy - exports (real) and electricity consumption and explore the derivatives of the series and their phase-plane plots to enable visual insights into the energy contained with the series. Specifically, our focus was on analysing the performance of the baseline function concurrent growth curve regression model with two competing models namely the functional historical model and the functional autoregressive model from a growth curve perspective (cumulative over the time window). The phase-plots revealed that the direction of the rate of change over the period did not alter that much during the period under consideration (1985-2015), though around the period of the global economic crisis the phase-plane plots, especially for electricity demand, did display a pronounced change for the years 2007 to 2008. There was a pronounced increase in the size of the cycle in the export data for the period under study. For the period analysed, the autoregressive FAR(1) model outperformed both the concurrent and historical models. In addition, the historical model performed much better than the baseline model when comparing the ratio of predicted root mean square errors. Most importantly, our exercise has contributed to the body of literature in economic analysis and forecasting insofar as the annual curves (functions) of the series retain the inter-annual growth (and variations) embedded in the data. This is a novel approach to modelling of South African economic indicators and provides practitioners with additional tools for quantitative analysis. In future, it would be of interest to see how the addition of more economic indicator covariates would improve the accuracy of the models in predicting South African exports.

## 6. References

### References

- Agalega, E. and Antwi, S. (2013), 'The impact of macroeconomic variables on gross domestic product: Empirical evidence from Ghana', *International Business Research* **6**, 108 – 116.
- Babalola, O. O., Danladi, J. D., Akomolafe, K. J. and Ajiboye, O. P. (2015), 'Inflation, interest rates and economic growth in Nigeria', *European Journal of Business and Management* **7**, 91–102.
- Baker, L. (2015), 'Renewable energy in South Africa's minerals-energy complex: a 'low carbon' transition?', *Review of African Political Economy* **42**(144), 245–261.  
**URL:** <https://doi.org/10.1080/03056244.2014.953471>
- Curran, P. J., Obeidat, K. and Losardo, D. (2010), 'Twelve frequently asked questions about growth curve modeling', *Journal of Cognition and Development* **11**(2), 121–136. PMID: 21743795.

- Divya, K. H. and Devi, V. R. (2014), ‘A study on predictors of GDP: Early signals’, *Procedia Economics and Finance* **11**, 375 – 382.
- Eilers, P. H. C. and Marx, B. D. (1996), ‘Flexible smoothing with  $B$ -splines and penalties’, *Statistical Science* **11**(2), 89–102.  
**URL:** <http://www.jstor.org/stable/2246049>
- Febrero-Bande, M. and Oviedo de la Fuente, M. (2012), ‘Statistical computing in functional data analysis: The R package *fda.usc*’, *Journal of Statistical Software* **51**(4), 1–28.  
**URL:** <http://www.jstatsoft.org/v51/i04/>
- Ferguson, R., Wilkinson, W. and Hill, R. (2000), ‘Electricity use and economic development’, *Energy Policy* **28**, 923 – 934.
- Harvey, A. and Jaeger, A. (1993), ‘Detrending, stylized facts and the business cycle’, *Journal of Applied Econometrics* **8**, 231–247.
- Harvey, A. and Scott, A. (1994), ‘Seasonality in dynamic regression models’, *The Economic Journal* **104**, 1324–1345.
- IMF (2013), *Balance of Payments Manual*, 6th edn, International Monetary Fund.
- Inglesii-Lotz, R. and Blignaut, J. N. (2011), ‘South Africa’s electricity consumption: A sectoral decomposition analysis’, *Applied Energy* **88**, 4779 – 4784.
- Kneip, A. and Utikal, K. J. (2001), ‘Inference for density families using functional principal component analysis’, *Journal of the American Statistical Association* **96**, 519–542.
- Kosiorowski, D. (2014), ‘Functional regression in short-term prediction of economic time series’, *Statistics in Transition new series* **15**.
- Lean, H. H. and Smyth, R. (2010), ‘Multivariate granger causality between electricity generation, exports, prices and GDP in Malaysia’, *Energy* **35**, 3640 – 3648.
- Levitin, D. J., Nuzzo, R. L., Vines, B. W. and Ramsay, J. O. (2007), ‘Introduction to functional data analysis.’, *Canadian Psychology* **48**, 135 – 155.
- Malfait, N. and Ramsay, J. O. (2003), ‘The historical functional linear model’, *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **31**, 115–128.
- Muller, H. (2011), ‘Functional data analysis’, *StatsProb: The Encyclopedia Sponsored by Statistics and Probability Societies*. .
- Narayan, P. K. and Smyth, R. (2008), ‘Energy consumption and real GDP in G7 countries: New evidence from panel cointegration with structural breaks’, *Energy Economics* **30**, 2331 – 2341.

- Narayan, P. K. and Smyth, R. (2009), ‘Multivariate granger causality between electricity consumption, exports and GDP: Evidence from a panel of Middle Eastern countries’, *Energy Policy* **37**, 229 – 236.
- Odhiambo, N. M. (2009), ‘Electricity consumption and economic growth in South Africa: A trivariate causality test’, *Energy Economics* **31**, 635 – 640.
- R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.  
**URL:** <http://www.R-project.org/>
- Ramsay, J. O., Hooker, G. and Graves, S. (2009), *Functional Data Analysis with R and MATLAB*, Springer.
- Ramsay, J. O. and Ramsey, J. B. (2002), ‘Functional data analysis of the dynamics of the monthly index of nondurable goods production’, *Journal of Econometrics* **107**, 327 – 344.
- Ramsay, J. O. and Ramsey, J. B. (2005), *Functional Data Analysis*, 2nd edition edn, Springer.
- Ramsay, J. O., Wickham, H., Graves, S. and Hooker, G. (2017), *fda: Functional Data Analysis*. R package version 2.4.7.  
**URL:** <https://CRAN.R-project.org/package=fda>
- Sami, J. (2011), ‘Multivariate cointegration and causality between exports, electricity consumption and real income per capita: recent evidence from Japan’, *International Journal of Energy Economics and Policy* **1**, 59.
- Sami, J. and Makun, K. K. (2011), ‘The relationship between exports, electricity consumption and economic growth: Empirical evidence from Brazil’, *World Review of Business Research* **1**, 28–37.
- Suits, D. B., Mason, A. and Chan, L. (1978), ‘Spline functions fitted by standard regression methods’, *The Review of Economics and Statistics* **60**, 132–139.  
Contributor:.
- Syed, A. A. S. G. and Shaikh, F. M. (2013), ‘Effect of macroeconomic variables on gross domestic product (GDP) in pakistan’, *Procedia Economics and Finance* **5**, 703–711.
- Ullah, S. and Finch, C. F. (2013), ‘Applications of functional data analysis: A systematic review.’, *BMC Medical Research Methodology* **13**, 1 – 12.
- Verick, S. and Islam, I. (2010), ‘The great recession of 2008-2009: Causes, consequences and policy responses’, *Institute for the Study of Labor* .  
**URL:** <ftp://ftp.iza.org/dps/dp4934.pdf>



- Wang, J.-L., Chiou, J.-M. and Muller, H.-G. (2016), 'Functional data analysis', *Annual Review of Statistics and its Application* **3**, 257–295.
- Wold, S. (1974), 'Spline functions in data analysis', *Technometrics* **16**(1), 1–11.  
**URL:** <http://www.jstor.org/stable/1267485>
- Wolde-Rufael, Y. (2006), 'Electricity consumption and economic growth: a time series experience for 17 African countries', *Energy Policy* **34**, 1106 – 1114.

## Tables

Table 1: RMSE from the Concurrent Model

	Period used for model building				
	1985-2010	1985-2011	1985-2012	1985-2013	1985-2014
2011	21588.89	—	—	—	—
2012	28644.9	24835.95	—	—	—
2013	39706.53	34891.18	29986.42	—	—
2014	48210.31	42710.83	37057.03	30644.68	—
2015	52703.52	46877.98	40889.29	34011.83	27928.55

Table 2: Relative RMSE of the historical and FAR(1) models to the concurrent model

		Period used for model building				
		1985-2010	1985-2011	1985-2012	1985-2013	1985-2014
2011	Historical	0.94	—	—	—	—
	FAR(1)	0.63	—	—	—	—
2012	Historical	1.05	1.05	—	—	—
	FAR(1)	0.45	0.52	—	—	—
2013	Historical	1.01	0.96	0.98	—	—
	FAR(1)	0.32	0.37	0.42	—	—
2014	Historical	1	0.92	0.85	0.8	—
	FAR(1)	0.27	0.3	0.35	0.42	—
2015	Historical	0.94	0.96	0.92	0.92	0.9
	FAR(1)	0.25	0.28	0.32	0.38	0.47