Province of KwaZulu-Natal

Provincial Treasury

IMES Unit

KWA-ZULU-NATAL GROSS DOMESTIC PRODUCT TEMPORAL DISAGGREGATION WITH A FUEL CONSUMPTION INDICATOR[1]

Clive Coetzee

General Manager: IMES Unit

Economist

clive.coetzee@kzntreasury.gov.za

033 897 4538

Working Paper 2b: 10 November 2014

[1] The views expressed in this working paper are the views of the author and might not necessarily reflect the views of the Provincial Treasury.

Introduction

Not having a time series at the desired frequency is a common problem for researchers and analysts. While there is no way to fully make up for the missing data, there are useful workarounds: with the help of one or more high frequency indicator series, the low frequency series may be disaggregated into a high frequency series (Sax and Steiner, Temporal Disaggregation of Time Series).

The aim of temporal disaggregation, according to Sax and Steiner, is to find an unknown high frequency series y, whose sums, averages, first or last values are consistent with a known low frequency series y_l (The subscript l denotes low frequency variables). In order to estimate y, one or more other high frequency indicator variables can be used. We collect these high frequency series in a matrix X.

The theoretical literature has outlined two alternative approaches to the temporal disaggregation of a single time series. The first one is the purely mathematical or time series model such as the smoothing technique of Boot, Feibes and Lisman (1967) and the ARIMA model of Wei and Stram (1990). The second approach includes methods that make use of related indicators observed at the desired high frequency, such as the static regression-based method of Chow and Lin (1971) and its variants by Fernández (1981) and Litterman (1983). Recent developments in the literature extend the regression methods to include a dynamic component such as Salazar et al. (1997), Santos Silva and Cardoso (2001), Di Fonzo (2003) and Proietti (2006).

Notation

According to Salazar et al. (1994), the following notation convention should be adopted. Single observations of low-frequency (LF) data are denoted by a single subscript, i.e. y_t, and are observed in T consistently spaced periods, which is a key assumption for the various temporal disaggregation methods

The aim is to derive an estimate of the underlying high-frequency (HF) series, whose unknown values are denoted by a double-subscript, so that $y_{t,u}$ denotes the HF value of

Y in sub-period u of period t = 1,...T, which is assumed to have periodicity s. For example, s=3 if we require monthly estimates of a quarterly observed series, s=4 if we want quarterly estimates for yearly data and s=12 if monthly data are required for an annually observed series.

The $(T \times 1)$ vector of LF data is denoted by

$$y_l = (y_1,..., y_t,...,y_T)'$$

while the $(n \times 1)$ vector of HF data is denoted by y_h. Importantly we must have $n \geq sT$. If $n = sT$, then $y_h = (y_{1,1},...,y_{T,s})'$ and we face a problem of distribution or interpolation. If $n > sT$, also an extrapolation issue has to be considered, with the difference $n - sT$ being the number of HF sub periods not subject to temporal aggregation constraints.

Similarly, any (TxK) matrix of LF data is denoted by X_l and its $(n \times K)$ HF counterpart is written as X_h. The columns of the LF matrix X_l are denoted by $X_{l'k}$ and those in the HF matrix X_h by $X_{h,k}$, where k = 1,...K, denotes the relevant variable of the matrix X_l and X_h, respectively. Accordingly, x_t and $x_{t,u}$ are $(K \times 1)$ vectors containing the LF and HF observations on K related series in period t and (t,u), respectively.

Techniques for Temporal Disaggregation

The Office for National Statistics (2010) states that there are many different methods for temporally disaggregating a time series (see Chen 2007 for a good survey). The choice of method will critically depend on the basic information available as well as preferences. But the fundamental objective is to construct a new time series that is consistent with the low frequency data whilst preserving the short–term movements in the higher frequency indicator series (if available).

Their article considers a number, although not an exhaustive, selection of techniques for deriving higher frequency data. If no higher frequency indicator is available then a smoothing method will be required such as:

- Cubic spines
- Boot, Feibes and Lisman (BFL) smoothing method

However, when a higher frequency indicator is available not only can the smoothing methods be applied but also a range of statistical methods. In particular three variants of the Chow–Lin regression method are frequently used:

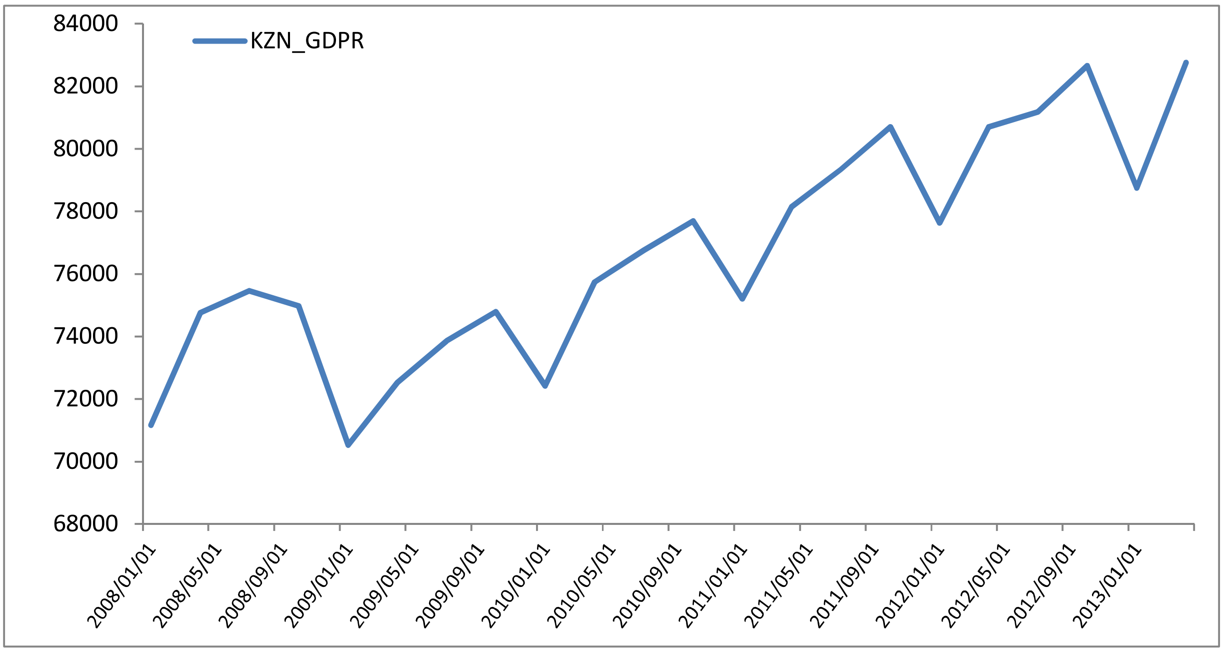
- Fernandez random walk model ($\mu = \mu_{t-1} + \varepsilon_t$)
- Random walk Markov model (Litterman, Min SSR and Max Log) ($\mu = \mu_{t-1} + \varepsilon_t$ where $\varepsilon_t = \alpha\varepsilon_{t-1} + e_t$)
- AR(1) model ($\mu = p\mu_{t-1} + \varepsilon_t$)

KZN Application

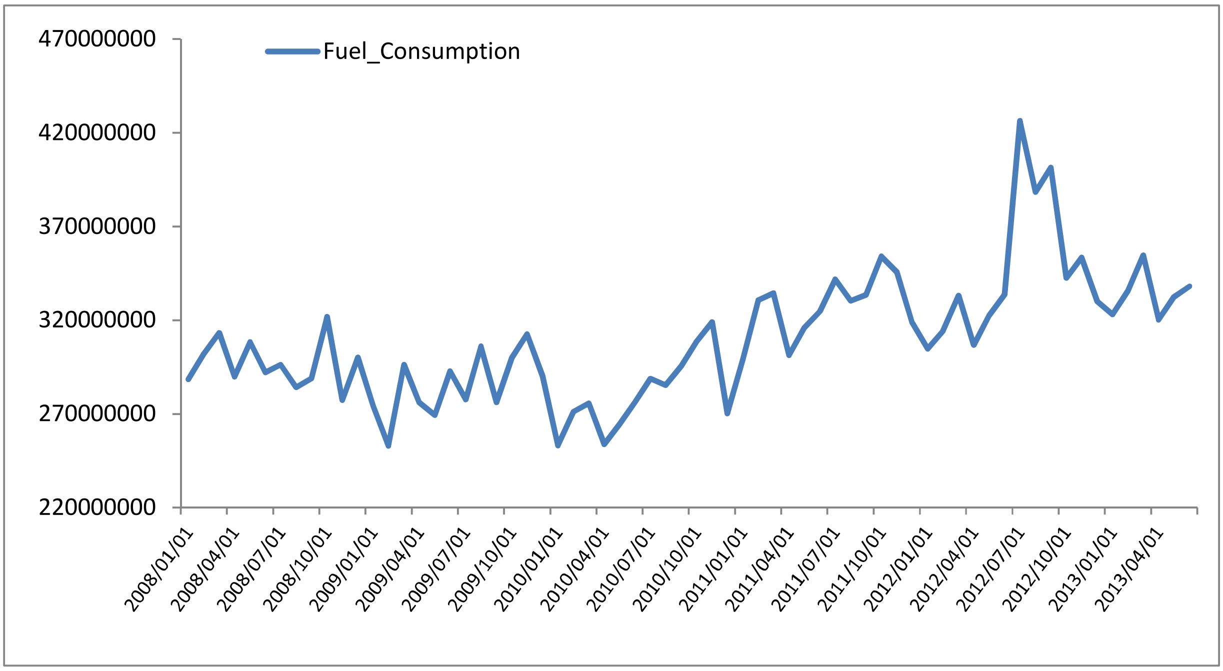
This study is an application of model-based approaches to the GDP quarterly data (high frequency) of the province, using a related indicator namely fuel (diesel and petrol) consumption (low frequency) for the period 2008 to 2013. The ECOTRIM software for temporal disaggregation is used to perform the analysis. ECOTRIM is a computer based programme developed by Eurostat specifically for temporal disaggregation of time series.

KZN quarterly real GDP for the period is displayed the below graph. A very clear seasonal pattern in the time series is noticeable. KZN monthly fuel consumption (litres) is displayed in graph 2. Coetzee (2012) argues that it is therefore (results of an error correction model) possible to conclude that there indeed exists a short and long run statistically significant relationship between national GDP and national fuel consumption. Fuel consumption can therefore be used as a proxy for GDP

Graph 1: Quarterly Real GDP (R'm, 2008 to 2013)

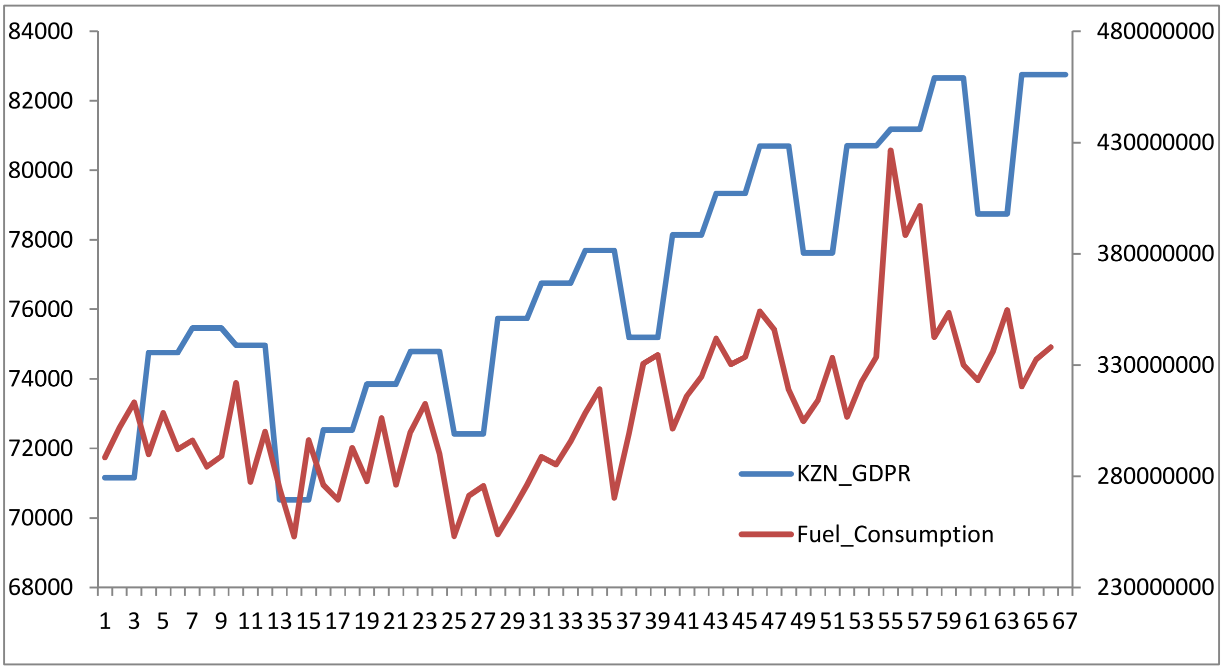


Graph 2: Monthly Fuel Consumption (Litres, 2008 to 2013)



Graph 3 displays the quarterly GDP and monthly fuel consumption variables.

Graph 3: Quarterly GDP and Monthly Fuel Consumption



The descriptive statistics for the two variables are displayed below:

Table 1: Average Quarterly and Monthly Descriptive Statistics (GDP, R'm and Fuel, L'm)

	KZN_GDP	FUEL_CONSUMPTION
Mean	76 716	310
Median	76 249	306
Maximum	82 753	426
Minimum	70 528	253
Std. Dev.	35 65	33.59
Skewness	0.07	0.90
Kurtosis	2.07	4.50
Jarque-Bera	0.80	15.02
Probability	0.67	0.00
Observations	22	66

KZN Temporal Disaggregation

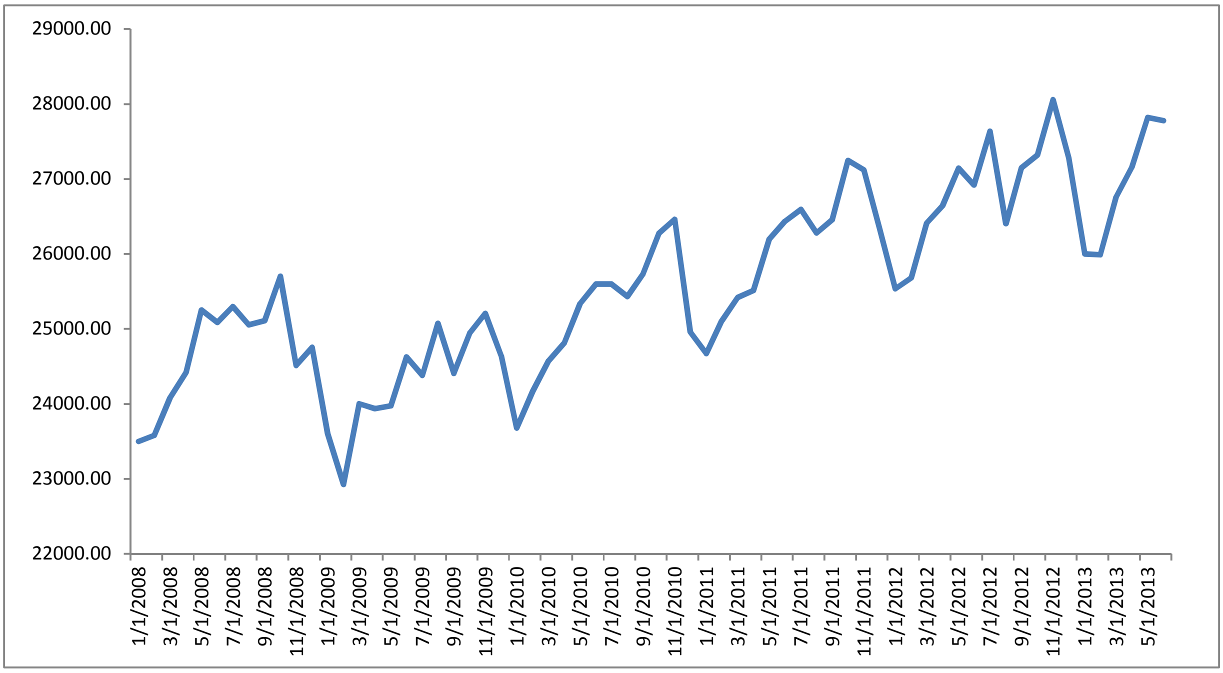
Graph 4 displays the monthly time series of KZN GDP, derived from the quarterly series using the monthly fuel consumption time series, and estimated using the flow series type of aggregation and the AR(1) Max Log model. The value of the parameter is 0.47. The results of the regression equation indicate that the monthly fuel consumption time series is indeed statistically significant.

Variable	Estimate	Std Error	t-Stat
CONSTANT	18486	2036.98	9.08
Fuel Consumption	0	0	3.5

Additional results include the following

- R-Squared = 0.38
- Probability of F = 0.17
- Durban Watson = 2.08
- Residual Sum of Square = 25 221 587
- Total Sum of Squares = 40 683 883
- Jarque-Bera normality = 0.24

Graph 4: AR (1) MAX LOG model



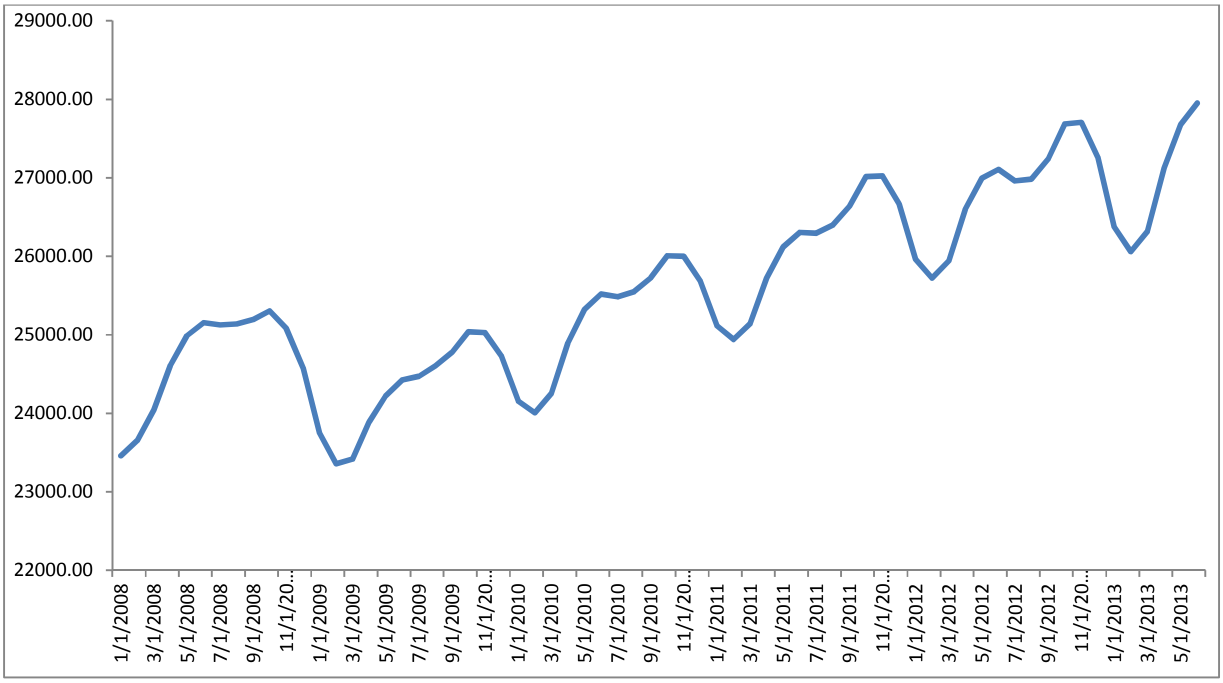
Graph 5 displays the monthly time series of KZN GDP, derived from the quarterly series using the monthly fuel consumption time series, and estimated using the flow series type of aggregation and the Fernandez model. The value of the parameter is 1. The results of the regression equation indicate that the monthly fuel consumption time series is not statistically significant.

Variable	Estimate	Std Error	t-Stat
CONSTANT	23354	2197.25	10.63
Fuel Consumption	0	0	0.05

Additional results include the following

- R-Squared = 0
- Probability of F = 0.01
- Durban Watson = 2.79
- Residual Sum of Square = 8 613 971
- Total Sum of Squares = 8 615 233
- Jarque-Bera normality = 0.86

Graph 5: Fernandez model



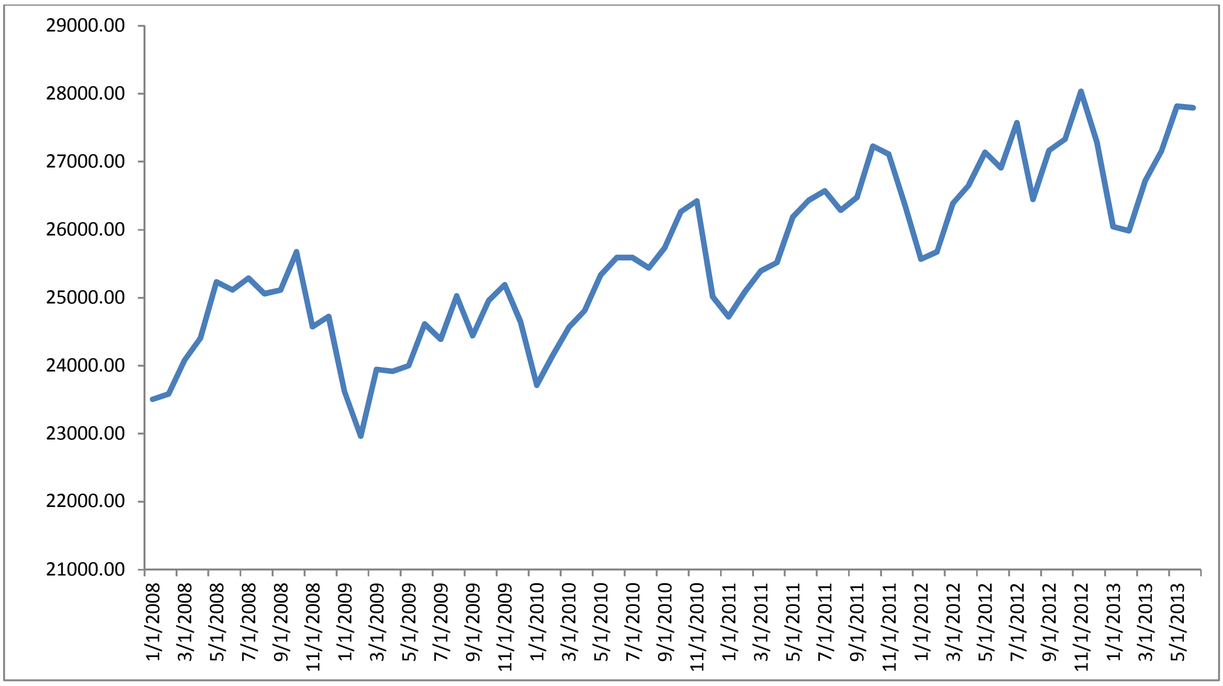
Graph 6 displays the monthly time series of KZN GDP, derived from the quarterly series using the monthly fuel consumption time series, and estimated using the flow series type of aggregation and the AR(1) Min SSR model. The value of the parameter is 0.55. The results of the regression equation indicate that the monthly fuel consumption time series is statistically significant.

Variable	Estimate	Std Error	t-Stat
CONSTANT	19162	2128.7	9
Fuel Consumption	0	0	3.03

Additional results include the following

- R-Squared = 0.32
- Probability of F s = 0.17
- Durban Watson = 2.14
- Residual Sum of Square = 24 879 372
- Total Sum of Squares = 36 329 280
- Jarque-Bera normality = 0.28

Graph 6: AR(1) Min SSR model



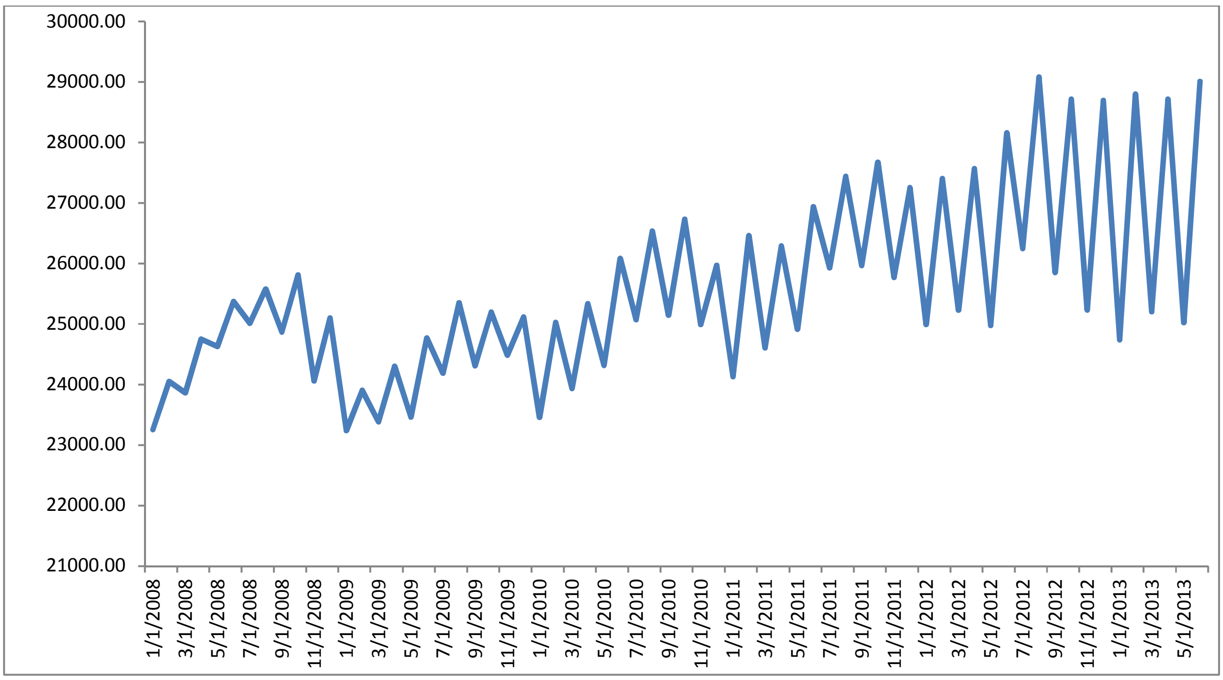
Graph 7 displays the monthly time series of KZN GDP, derived from the quarterly series using the monthly fuel consumption time series, and estimated using the flow series type of aggregation and the Litterman Max Log model. The value of the parameter is -0.99. The results of the regression equation indicate that the monthly fuel consumption time series is statistically significant.

Variable	Estimate	Std Error	t-Stat
CONSTANT	19567	1586.89	12.33
Fuel Consumption	0	0	2.8

Additional results include the following

- R-Squared = 0.28
- Probability of F = 0.17
- Durban Watson = 1.79
- Residual Sum of Square = 8 724 316
- Total Sum of Squares = 12 135 794
- Jarque-Bera normality = 0.37

Graph 7: Litterman Max Log model



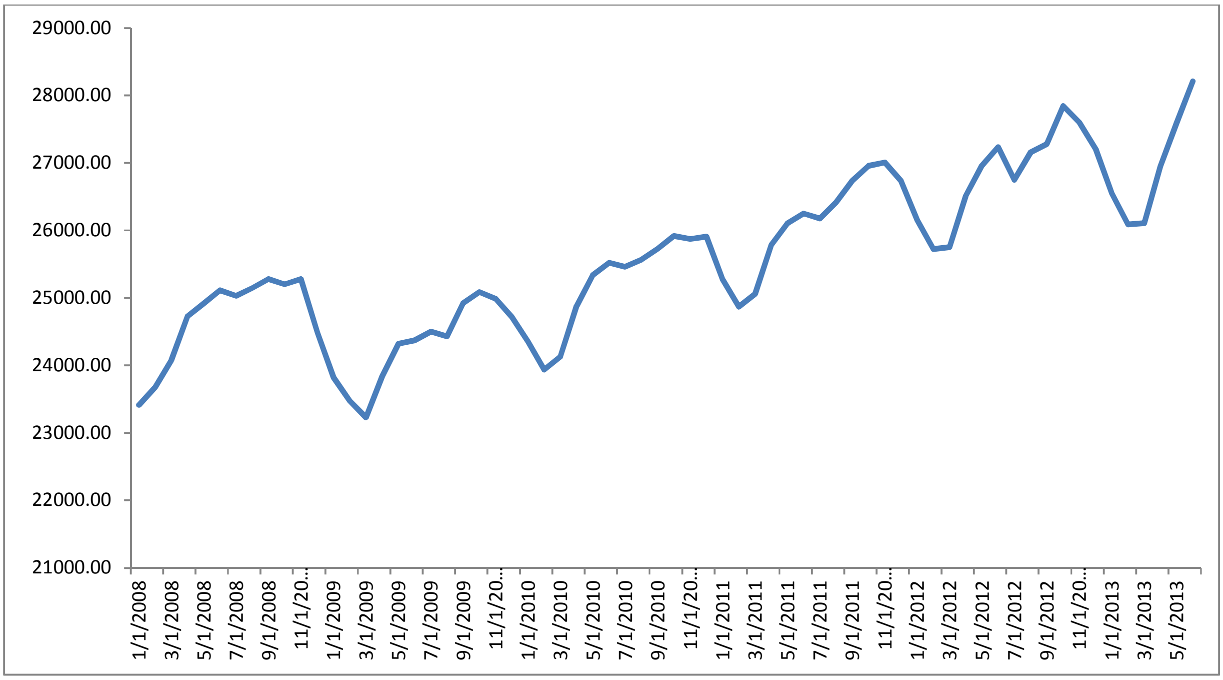
Graph 8 displays the monthly time series of KZN GDP, derived from the quarterly series using the monthly fuel consumption time series, and estimated using the flow series type of aggregation and the Litterman Min SSR model. The value of the parameter is 0.76. The results of the regression equation indicate that the monthly fuel consumption time series is statistically significant.

Variable	Estimate	Std Error	t-Stat
CONSTANT	25190	2126.24	11.85
Fuel Consumption	0	0	-1

Additional results include the following

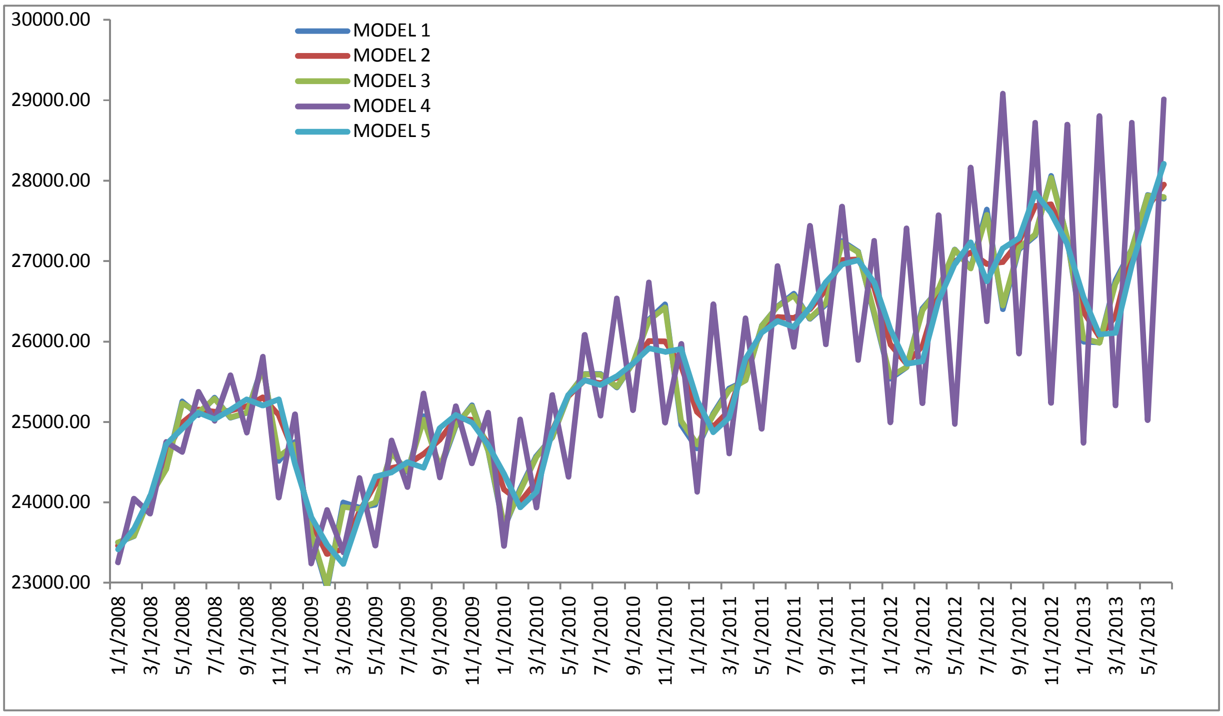
- R-Squared = 0.05
- Probability of F = 0.14
- Durban Watson = 3.08
- Residual Sum of Square = 4 282 650
- Total Sum of Squares = 4 496 825
- Jarque-Bera normality = 1.54

Graph 8: Litterman Min SSR model



The results of the above five models are displayed in the below graph.

Graph 9: KZN Temporal Disaggregation



It is evident that the trend behaviour of the five models is very similar. However the fluctuations around the trend are very different between the five models. The results of five models are displayed the table below:

Table2: Results of Temporal Disaggregation

	AR (1) MAX LOG	FERNANDEZ	AR(1) MIN SSR	LITTERMAN MAXLOG	LITTERMAN MIN SSR
Value of the parameter	0.47	1	0.55	-0.99	0.76
T stat	3.5	0.05	3.03	2.8	-1
R-Squared	0.38	8 615 233	0.32	0.28	0.05
Probability of F	0.17	0.01	0.17	0.17	0.14

Durban Watson	2.08	2.79	2.14	1.79	3.08
Residual Sum of Square	25 221 587	8 613 971	24 879 372	8 724 316	4 282 650
Total Sum of Squares	40 683 883	8 615 233	36 329 380	12 135 794	4 496 825
Jarque-Bera normality	0.24	0.86	0.28	0.37	1.54

Conclusions

This article makes use of a monthly time series (fuel consumption) to disaggregate the quarterly provincial GDP time series. The article further analysis five flow model based techniques to disaggregate the provincial quarterly GDP to monthly frequency. The efficacy of temporal disaggregation techniques depends on the validity of the regression model and finding a suitable related series. This helps in compiling timely data and enhances the quality of the data.