



**Strathmore**  
UNIVERSITY

**A Short-Term Now-Casting Model for Kenyan's Quarterly GDP: The Dynamic Factor  
Model Approach**

**Wangila Enock Wafula, 069978**

**Submitted in partial fulfillment of the requirements for the  
Degree of  
Bachelor of Business Science Actuarial Science at Strathmore University**

**[School of Finance and Applied Economics (SFAE)  
Strathmore University  
Nairobi, Kenya**


**[November, 2015]**

## DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, this Research Project contains no material previously published or written by another person except where due reference is made in the Research Proposal itself.

© No part of this Research Proposal may be reproduced without the permission of the author and Strathmore University

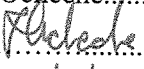
Wangila Enock Wafula..... [Name of Candidate]

..... [Signature]

13<sup>th</sup> November 2015..... [Date]

This Research Proposal has been submitted for examination with my approval as the Supervisor.

John Ocheche..... [Name of Supervisor]

..... [Signature]

13/11/2015..... [Date]

School of Finance and Applied Economics

Strathmore University

## Table of Contents

Abstract.....	iv
1. Introduction.....	1
1.1 Background.....	1
1.2 Problem Statement.....	2
1.3 Research Objective.....	3
1.4 Research Question.....	3
1.5 Statement of Significance.....	3
2. Literature Review.....	4
2.1 Model Selection.....	4
2.2 Dynamic factor Model.....	5
2.3 Factor Estimation.....	6
2.4 Determining the Number of Factors.....	8
2.5 Research / Knowledge Gap.....	9
2.6 Conceptual Framework.....	9
3. Methodology.....	10
3.1 Introduction.....	10
3.2 Research Design.....	10
3.3 Population and Sampling Design.....	10
3.4 Nature and Source of Data Set.....	10
3.5 Data Analysis.....	10
4. Results and Analysis.....	14
5. Discussions.....	15
5.1 In sample evaluation.....	15
5.2 Out of Sample Evaluation.....	18
6. Conclusion.....	19
Appendix.....	20
References.....	21

## Abstract

This research study, *A Short-Term Now-Casting Model for Kenyan's Quarterly GDP: The Dynamic Factor Model Approach*, was carried out with the need to: establish a set of common factors explaining variations in macroeconomic variables, determine the optimal number of factors for nowcasting and to evaluate the performance of the dynamic factor model in nowcasting Kenya's quarterly GDP. Principal component analysis was used to extract the common factors and an information criteria, among other methods, used to determine the optimal number of common factors. The common factors were then regressed against the in sample quarterly GDP figures to obtain the full specification of the dynamic factor model equation. The extraction of the common factors allowed for an out of sample nowcasting exercise, from which, the study concluded that the dynamic factor model performs adequately well relative to historical GDP figures. Consequently, the model would be well useful in making timely estimates of end of quarter GDP. The scope of this research can be extend to analyze the implication of data revisions on the nowcasts.

# 1. Introduction

## 1.1 Background

Tracking real time economic performance and turning points of economic activity has been at the core of recent macroeconomic models. The availability of data on historical economic performance of the economy and present economic conditions guide the decision making process of monetary and fiscal policy makers. Across countries, GDP reports are published with a considerable time lag of two – three months after the reference quarter (Angelini et al., 2011). The delay in publication acts as an impediment to timely understanding of the economic situation and constraints the ability of policy makers to make timely policy interventions (Golinelli & Parigi, 2007). In an attempt to bridge the information gap, short-term models have been developed that explain the current performance of the economy by exploit readily available information.

The use of factor models is a common approach of providing real-time information by use of indicators. In advanced and emerging economies, the indicators of short-term performance have provided a basis for macroeconomic decision making (Dixit & Opoku-Afari, 2012). Bums and Mitchell (1947) pioneered the works on factor models by developing indicators for cyclical revival. Advancements in the field resulted in development of dynamic factor models that attribute variations in macroeconomic time series to a small set of variables, regarded as common factors. Related literature on dynamic factor models spans widely: Stock and Watson, 2002; Forni et al., 2003; Schumacher and Breitung, 2006; Schumacher, 2007; Angelini et al., 2008; Frale et al., 2008; Barhoumi et al., 2010.

Factor models, unlike other models, are superior as they allow for: inclusion of many variables without having a limited number of parameters varying independently, minimization of measurement error and local shocks and accommodate relaxation of assumptions unlike structural models that have overly tight assumptions (Breitung and Eickmeier, 2005). Their application in macroeconomics has been as a consequence of their ability to make use of monthly or higher frequency data as proxies for actual economic performance. This allows for the use of monthly indicators from multiple sectors of the economy to provide estimates for quarterly GDP. The estimates are of value to policy makers especially in cases where, due to prevailing economic conditions, there is need for a prompt policy measure. The models can also provide useful macroeconomic projections alongside other existing models.

This study seeks to develop a short-term now-casting model for Kenya's quarterly GDP using high frequency data, monthly indicators, as proxies for Kenya's economic performance. The study explores a research gap proposed by (Dixit & Opoku-Afari, 2012) in developing a leading indicator model for Kenya by extending the proposed model to the use of principal component analysis as a method of selecting the common factors.

## 1.2 Problem Statement

GDP reports provide an aggregate measure of the pulse of an economy. They form the basis of policy interventions for: correcting declining performance in ailing sectors and stimulating growth in sectors that have potential for higher contribution to economic performance (Barhoumi et al., 2010). Due to the practice of statistical bodies publishing GDP reports with a considerable time lag, there exists an information gap on tracking the turning points of economic activity in real-time. The information gap has been reduced by utilizing high frequency data that is readily available on a monthly basis or higher frequency (Barhoumi et al., 2010).

Unlike developed nations, a limited number for countries in Sub-Saharan Africa rely on leading indicators for macroeconomic decision making (Dixit & Opoku-Afari, 2012). Consequently, the information gap persists and real time economic performance cannot be tracked. Monetary and fiscal policy makers are then unable to make timely policy interventions in cases of adverse changes in the business cycle. Dixit & Opoku-Afari (2012) point out that the absence of short term forecasts of economic performance bars accurate measurement of policy measures instituted by central banks.

Kenya is a developing country and has the same challenge of delay in publication of quarterly GDP reports from the end of the reference quarter. However, there exists a set of high frequency data from multiple economic sectors that can be used to provide an estimate of the country's economic performance. This paper proposes a short term model that utilizes monthly available data to now-cast Kenya's economic pulse. The proposed dynamic factor model would provide a solution to the need to track economic activity in real-time. This paper adds to the body of literature on factor models by applying the model to the Kenyan case, not done before, and consequently making a conclusion as to its performance in the nowcasting exercise.

### 1.3 Research Objectives

1. To determine the full set of common factors explaining the variation of Kenya's economic performance
2. To determine the optimal number of common components to include in the model for an accurate estimate of Kenya's quarterly GDP
3. To determine the performance of the dynamic factor model in nowcasting Kenya's quarterly GDP

### 1.4 Research Questions

1. What are the set of common factors that explain the variation in Kenya's economic performance?
2. What is the optimal number of common factors that provide the most accurate estimate of Kenya's quarterly GDP?
3. How well does the dynamic factor model nowcast Kenya's quarterly GDP?

### 1.5 Significance of the Study

This study would be useful to policy makers at Kenya's Central bank. The dynamic factor model would provide as a timely basis for prompt intervention measures on the overall economy in the short-run guided by a timely measure of economic. The model would also supplement existing forecasting models.

## 2. Literature Review

### 2.1 Model Selection

Macroeconomic policy decisions are based on availability of current information and information on past economic performance. Timely availability of past economic information is however, a challenge in policy making. To address the problem of delay in publication of GDP information, various models have been proposed. Five models, only a subset of the developed models, are representative of the commonly used models by policy makers: bridge equations, dynamic factor models, autoregressive models, bivariate vector autoregressive models and Bayesian vector autoregressive models.

The use of VAR models in macroeconomic forecasting follows from the works by Sims (1980). According to Sims (1980), the restrictions on the structural macro-econometric models were unnecessary and instead needed to be relaxed to allow for no restrictions. The effect of policy interventions could be measured by identifying structural shocks and generating the impulse response function. Finally, a linear transformation of the VAR of the data could be transformed to provide a moving average representation in regard to the structural shock. However, the model presented practical issues evident from the disappointing inflation forecasts at the Federal Bank of Minnesota (Sims, 1999).

Forni and Lipi (2000) qualify the VAR models as an appropriate model for forecasting small datasets. The small dataset does not allow for the over-parameterization of the model compared to the case for a large dataset. Gupta and Kabundi (2011) concur arguing that since in VAR models all variables are used as parameters, the models result in a curse of dimensionality where the model becomes complex and takes much time to complete as the parameters converge slowly to the focal point. The over parameterized model results in inefficient estimates and possibly large forecasting errors (Gupta and Kabundi, 2011). Further, the use of the models results in a scarcity of degrees of freedom for large datasets. The limited number of parameters that vary independently makes the VAR models insufficient in mimicking complex economic relations.

Bayesian VAR models were proposed by Litterman (1980) in a bid to address the over-parameterization of the VAR models. The model involved applying a prior to the data set and

centralizing the equations around a random walk. Using procedures proposed by Litterman (1980), a prior distribution is imposed on the parameters. The approach improves on the VAR model as it overcomes the over-parameterization of the model. Eklund and Kapetanios (2008) states that the Bayesian VAR is limited in application to only small data sets. However, Banburra, Giannone and Reichlin (2007) and Carriero, Kapetanios and Marcellino (2007) apply the approach to a large dataset and present encouraging results. Gupta and Kabundi (2011) suggest the Bayesian VAR as an alternative to the dynamic factor model in macroeconomic forecasting. However, on comparing the performance of the models, the dynamic factor model outperforms the Bayesian VAR for large datasets and short-term to medium term forecasts.

The dynamic factor model is an alternative approach to the forecasting of macroeconomic variables based on common factors. The model has been at the core of literatures on macroeconomic forecasting for policy makers. The support from literature is largely on two grounds: the ability of the model to outperformance other macroeconomic models and ability to exploit timely releases of economic information (Giannone, Reichlin and Small, 2008; Liu, Romeu and Matheson, 2011; and Breitung and Pigorsch, 2013). Due to the model's ability to utilize timely releases of information, the model is a preferred choice for modelling economic variables in real time.

Geweke (1977) postulated the dynamic factor model under the premise that a small number of unobserved common factors as well as individual factors could explain the movement of economic time series. The proposed model was an extension of the classical factor model to incorporate time series. Sargent and Sims (1977) affirmed the proposition by establishing that two dynamic factors could by large explain the variations in the quarterly macroeconomic variables: prices, employment and output, for the United States of America (USA). The findings have been supported by similar conclusion from related studies that include but are not limited to: Stock and Watson (2004); Giannone, Reichlin, and Sala (2004); Doz, Giannone, and Reichlin (2006); Hallin and Liška (2007) and Giannone, Reichlin, and Small (2008).

Literature on the traditional dynamic factor model has developed around the model's set of assumptions, estimation of the common factors and dealing with missing observation. The exact

dynamic factor model is established on the assumption that the measurement error and individual features, idiosyncratic component, are uncorrelated with the common factors (Sargent and Sims, 1997).

### 2.3 Factor Estimation

Stock and Watson (2010) argue that the strict restriction on the idiosyncratic components for the dynamic factor model informed development of models where the assumptions were relaxed. Further, Sargent and Sims (1977) and Geweke (1977) relied on frequency domains to establish the factors that could not be computed directly constraining their use in forecasting. Direct estimates of the common factors could only be through time domain methods of the dynamic factor model: first generation, second generation and third generation (Stock and Watson, 2011).

The first generation models, applicable only to models of few parameters, provide estimates of the common factors based on the maximum likelihood and the kalman filter approach (Stock and Watson, 2011). Engle and Watson (1983) implemented the approach by using the maximum likelihood to estimate the parameters of the model and subsequently using the kalman filter to obtain efficient parameter estimates. The use of the kalman filter approach allowed for handling of data with mixed frequencies and hence variables with missing observations could still be used in the model. Despite the ability of the first generation models to deal with missing observation, the model was limited by assumption of normality of residuals and cumbersome computations for large datasets (Barhoumi, Ferrara and Darné, 2014).

The second generation models extended the dynamic factor model to include application to large datasets. Forni et al. (2000) and Stock and Watson (2002) pioneered the work on these models by relaxing assumptions of the classical model to allow for weak serial and cross-sectional correlation between idiosyncratic components. Using a non-parametric method, principle component analysis, Forni et al. (2000) estimates common factors on the premise that the weighted average of the idiosyncratic factors will converge to zero. The convergence only allows for linear combinations of the common factors and makes computation easier relative to first generation models. Stock and Watson (2002) apply the second generation model and show that efficient forecasts can be obtained by regressing on factors that are predictors.

Boivin and Ng (2003) suggest a weighted principal component analysis approach that improves on the accuracy of the principal component analysis. Stock and Watson (2005) concur with the findings and cite that the forecasts based on principal components and weighted principal components are highly correlated. Further, their work addresses the serial correlation between the error terms through: estimate of latent factors using principal components, regression of the random variable using an autoregressive fit and applying the Boivin and Ng (2003) residual variance to the random variable. However, a static principal component is unable to handle the time variant characteristic of dynamic processes. Addressing the drawback, Forni, Hallin, Lippi, and Reichlin (2004) computed and prove that the dynamic principal components provides consistent estimates of the latent factors.

There exists contradicting evidence on the performance of static and principle components. Boivin and Ng (2005) establish that the static principle component outperforms the dynamic principle components. However, Stock and Watson (2006) and D'Agostino and Giannone (2007), arrive at a different conclusion, the two models perform similarly. Schumacher (2007) forecasts multiple German economic variables using the two models and arrives at mixed results on the performance of the two models. Consequently, due to varying literature on performance of the model, the choice between static and dynamic principle components is subjective to the researcher.

Stock and Watson (2011) refute the direct use of dynamic principle components in forecasting exercises that require estimates of the latent factors for the full sample. The case follows from the estimation method of the dynamic factor model that does not provide estimates of the latent factors at the end of the sample due to the two-sided smoothing. However, the kalman filter and smoother approach can be applied to incomplete datasets to handle the missing values and jointly provide efficient estimates of the latent factors. The usefulness of the kalman filter and smoother inform the case for the third generation models.

The third generation of dynamic factor models is a hybrid of the state space approach and the principle component approach. The combined use of the efficiency of the statistical capacity of kalman filter, state space, and the convenience of the forecasting approach by principle components allows for application of the model in real time (Giannone, Reichlin, and Small, 2008). Kalman filter allows for handling of missing observations and can thus be applied in macro-

economic forecasting, where there exists missing data points due to publication lags. The two-step parametric principle component approach by Doz, Giannone, and Reichlin (2006) and Giannone, Reichlin, and Small (2008) involves estimating the common factors by principal components and then using the estimated common factors to estimate the unknown factors by the kalman filter. Since the kalman filter approach averages both across the time and series, its estimates provide improvements in estimates of the factors.

An alternative approach to estimating the common factors is by use of Bayesian estimates. Stock and Watson (2011) attribute the reason for use of the Bayesian technique to: ease of computation, a higher degree of stability relative to the maximum likelihood approach for many parameters, its applicability in estimating non-normal latent variables and the need to impose prior information through a prior distribution in modelling the latent factors. Bai and Wang (2015) apply the Bayesian estimation approach in studying the minimal identification conditions for dynamic factor models. In comparison with the generation models, most of the reasons for use of the Bayesian estimation method are sufficiently addressed by the third generation of dynamic factor models.

#### Determining the Number of Factors

A key step in computing the factor models lies in determining the number of factors. Methods of estimating the number of factors for factor models largely develop around literature by Bai and Ng (2002). Bai and Ng (2002) proposed the use of an Information Criterion (IC) in determining the number of static factors for large data sets. To establish the number of static factors, the IC is minimized to obtain the maximum number of static factors. Alessi et al. (2010) concurs with the approach by Bai and Ng (2002) and provides an extension of the approach. The proposed procedure is dependent on the variance of the estimated static factors. Kapetanios (2010) further builds up on the IC method by basing the estimates on the random maximization theory.

Multiple approaches have been proposed for estimating the number of factors for dynamic factor models. Bai and Ng (2007) develop an Information Criterion by estimating static factors and applying a VAR model to the factors. Using a set of predefined criteria, Bai and Ng (2007) apply a variance-covariance matrix to obtain optimal number of factors for the dynamic factor model. Stock and Watson (2005) develop a computational approach from Bai and Ng (2002) by applying the criterion to the errors resulting from lagged values on static factors.

Breitung and Pigorsch (2013) argue that the Bai and Ng (2002) approach is unable to distinguish the dynamic factors from their lags. The limitation results in the number of factors being equal to the sum of lags of dynamic factors and the number of dynamic factor. As an alternative, Breitung and Pigorsch (2013) propose a criterion for selection based on the correlation between the past and current values of the common factors. Though the proposed approach is argued to outperform other suggested selection methods, it is limited in application to only cases where the variance of the common factors is medium or low. Breitung and Pigorsch (2013) establish that their approach performs at least similarly to the approach by Stock and Watson (2005) and of Bai and Ng (2007).

### 3.5 Research Knowledge Gap

The body of literature on nowcasting models is focused on two key aspects: identifying a reliable model for nowcasting and providing sufficient proof for the ability of the models to make efficient nowcast. While studies have been done across Europe to validate the dynamic factor model as a precise model for nowcasting, no such studies have been done across Sub-Saharan Africa. This adds to the body of literature by testing the hypothesis that the dynamic factor model provides a precise and timely measurement of quarterly GDP.

### 3.6 Data Source and Selection

The theory behind factor models is that a few set of common factors can be extracted to explain the movement of economic time series Geweke (1977). Factor models allow for the bridging of high frequency data with quarterly or annual GDP. High frequency data on macroeconomic performance is often drawn from indicator variables produced by statistical bureaus. In the scope of this study, the macroeconomic variables relied upon were: monthly inflation rate, exchange rate, coffee and tea production volume and export values, value of buildings, value of horticultural export, tourist arrivals, car registrations, cement production and consumption, imports and exports, electricity generation and consumption, and fuel consumption. The variables are related as they jointly they provide a signal as to the trends in the economy.

### 3. Methodology

#### 3.1 Introduction

This study adopts the methodology proposed by Giannone, Reichlin, and Small (2008) for the dynamic factor model. The choice of the parametric model is based on its ability to exploit high frequency data and handle missing observations via the kalman filter and smoother approach.

#### 3.2 Research Design

This research is exploratory in nature as it seeks to establish a set of common factors that are responsible for almost all the co-movements in macroeconomic time series. In addition, since the study seeks to assess the predictive accuracy of the dynamic factor model in forecasting quarterly GDP, an exploratory research design is ideal.

#### 3.3 Population and Sampling Area

The entire set of macroeconomic indicators served as the population for the study. To obtain the sample from the population, non-probability sampling was used based on the availability of the indicators over the study period. Indicators that were available throughout the sample period were included in the study and formed the basis for the analysis and conclusion. The study relied upon monthly economic indicators available throughout the sampling period January 2004 to June 2015.

#### 3.4 Data Sources and Data Collection

The study will rely on secondary data from Kenya Bureau of Statistics. It will include monthly lead economic indicators from multiple sectors of the economy: financial sector and price indices, building and construction sector, agricultural sector, tourism sector, energy sector and transport sector. For the forecasting exercise to assess the predictive accuracy of the model, the data will be drawn from two other sources: Nairobi Securities Exchange (NSE) and the Central Bank of Kenya (CBK).

#### 3.5 Model Specification

The dynamic factor model proposed by Giannone, Reichlin, and Small (2008) is based on the need to exploit the colinearity in the dataset by summarizing information in a set of few common factors. The colinearity in the series of data allows for a parsimonious representation of a bulk of the

interaction among the series. Due to the limited number of common factors, the model allows for feasible projections based on few parameters.

Giannone, Reichlin, and Small (2008) propose a dynamic factor model in obtaining short-term forecasts of GDP. Prior to the specification of the model, Giannone, Reichlin, and Small (2008) highlight the issue of an unbalanced dataset. By defining the information available at the end of the month  $v$  as to be contained in the information set,  $\Omega_v^n$ , where  $n$  denotes the number of monthly time series datasets, they propose that the GDP can be projected as a function of the information set:

$$proj [GDP_t | \Omega_v^n]$$

Further, the information set is defined as:

$$\Omega_{v_j} = \{Y_{it|v_j}; i = 1, \dots, n; t = 1, \dots, T_{iv_j}\}$$

Where  $v$  represents the month of release of the data, and  $v_j$ , the vintage, the exact day of the month of the  $j$ th data release. The variable  $Y_{it|v_j}$  is composed of the individual time series  $i$  and the time component in months,  $t = 1, \dots, T_{iv_j}$ . The last value in the equation indicates the last period for which the monthly time series data has an observed value.

Giannone, Reichlin, and Small (2008) note that evaluating the model using the entire information set is complex due to the large amount of information making the information set very large. As an alternative, they propose utilizing the co-linear relationship between the variables to summarize the large information set into a few common factors,  $F_t$ . A projection on these set of common factors is able to capture a large part of the covariance in the dataset.

The parametric dynamic factor model proposed by Giannone, Reichlin, and Small (2008) is as given:

$$y_{it|v_j} = u_i + \lambda_i F_t + \xi_{it|v_j}$$

Where  $u_i$  denotes a constant,  $\lambda_i F_t$  the common components and  $\xi_t$  the idiosyncratic component, that are unobserved stochastic processes. The proposed model assumes that the common component is a linear function of unobserved common factors,  $F_t$  while the idiosyncratic

component is driven by shocks that are specific to the model's variables and are inclusive of the data revision errors.

According to the proposed methodology, the first step involves transforming the data into stationary data. The transformed data is then standardized to variables with zero sample mean and unitary variance.

$$x_{it} = y_{it} - \hat{u}_i$$

$$z_{it} = \frac{1}{\hat{\sigma}_i} (y_{it} - \hat{u}_i)$$

Where  $\hat{u}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$  and  $\hat{\sigma}_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_{it} - \hat{u}_i)^2}$

The estimates of the common factors are then obtained through principal component analysis which is the solution to the stated problem:

$$(\tilde{F}_t, \Lambda^{\wedge}) = \arg \min \sum_{t=1}^T \sum_{i=1}^n (z_{it} - \lambda_i F_t)^2$$

The stated computation relies on the sample covariance matrix (S) of observed values. By extracting a matrix of eigenvectors,  $V$ , from a diagonal matrix of the largest eigenvalues of the sample covariance matrix (S),  $D$ , the factors,  $\tilde{F}_t$ , can be estimated as:

$$S = \frac{1}{T} \sum_{t=1}^T z_t z_t'$$

$$\tilde{F}_t = V' z_t \text{ where } V' \text{ is the matrix transpose}$$

The number of factors of the model is to be determined as by techniques proposed by Bai and Ng (2007). For the number of static factors, Bai and Ng (2007) proposes minimizing the information criteria for the number of factors,  $j = 0, \dots, r_{max}$ . The information criterion is defined by:

$$IC(j) = \ln(V(j, F)) + j \left( \frac{N + T}{NT} \right)$$

Where  $V(j, F)$  is such that:

$$V(j, F) = (NT)^{-1} \sum_{t=1}^T (x_t - \Lambda \tilde{F}_t)' (x_t - \Lambda \tilde{F}_t)$$

The proposed model relies on regression of the estimated factors to obtain the factor loadings,  $\Lambda$ , of the common factors and the covariance matrix,  $\psi$ , of the idiosyncratic component. Their estimates are obtained as follows:

$$\Lambda = \sum_{t=1}^T x_t \tilde{F}'_t \left( \sum_{t=1}^T \tilde{F}_t \tilde{F}'_t \right)^{-1}$$

and

$$\hat{\psi} = \text{diag}(S - VDV)$$

For unbalanced datasets, the Kaman smoother is used to compute the factors by weighting each variable's content according to its news, an element driven by the variable's common shocks, to noise (idiosyncratic component) ratio. Estimates under the kalman filter are given as:

$$F^* t = (\bar{\psi} I_r + \hat{\lambda}' \hat{\lambda})^{-1} \hat{\lambda}' x_t$$

The corresponding estimates of the factors are plugged into the general dynamic factor model equation and a quarterly estimate is defined as:

$$\hat{z}_{k|v_j}^q = \left( \frac{y_{z,k|v_j} + y_{z,k-1|v_j} + y_{z,k-2|v_j}}{3} \right)$$

Under the model and parameter assumptions as defined above, an early estimate of quarterly GDP can be computed.

#### 4. Results and Analysis

The data relied upon for the study was obtained from secondary sources: statistical abstracts from the Kenya Bureau of Statistic, leading economic indicators publications and the Central Bank of Kenya (CBK). Data was readily retrieved online from the aforementioned sources and consequently relied upon in analysis.

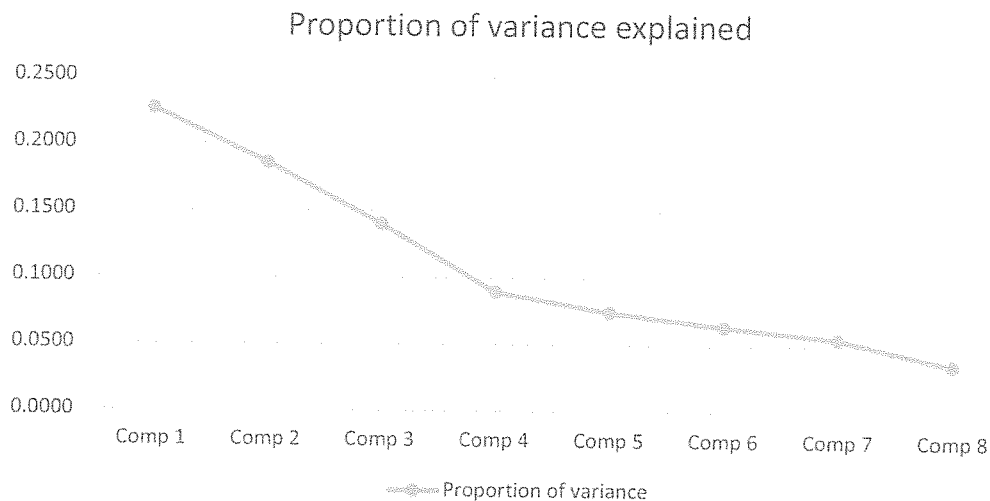
Data analysis relied on the procedure described in the methodology section of this paper. The data was first standardized to provide a uniform comparable scale. Logs and first differences were taken to obtain the growth rate of the monthly indicators over a quarter period. A table to show the data transformations applied on the data is provided in the appendix section. Principal component analysis was used to extract the common factors and factor loadings as is the practice with factor models. The values of the common components were then regressed against quarterly GDP to obtain coefficients for the dynamic factor model equation. A discussion of the same is as presented in the next section.

## 5. Discussions

### 5.1 In sample evaluation

The nowcasting exercise involve both an in sample and an out of sample evaluation. The in sample exercise was performed on a balanced data panel and facilitated the extraction of the common components and coefficients for the dynamic factor model equation. The first stage in implementation was the extraction of the common components from the monthly macroeconomic indicators by principal component analysis. Components were extracted on the assumption of a high degree of correlation between the monthly economic indicators. A total of 20 common components were extracted that represented the unobservable factors that explained much of the variation in the dataset.

To determine the optimal number of factors to retain, the Bai and Ngai (2007) information criterion was applied. The criterion, as was discussed, considered the number of time series variables and forecasting horizon. The criterion yielded the optimal number of common components as 8 that jointly accounted for 87% of the variation in the indicator variables. The number of factors was verified by applying another criteria, retaining only common factors whose eigenvalues are greater one. The results is in line with the underlying theory of factor model, a few unobservable factors can explain a large amount of comovement in economic time series (Angelini et al., 2008). A plot of the eight common factors against their eigenvalues is shown below.



*Figure 1: The minimum factors to retain is at least those at and before the first*

Common components	Eigenvalues	Proportion of variance	Cumulative proportion of explained variance
Comp 1	2.1266	0.2261	0.2261
Comp 2	1.9272	0.1857	0.4118
Comp 3	1.6731	0.1400	0.5518
Comp 4	1.3356	0.0892	0.6410
Comp 5	1.2191	0.0743	0.7153
Comp 6	1.1289	0.0637	0.7790
Comp 7	1.0501	0.0551	0.8341
Comp 8	0.8466	0.0358	0.8700

Table 1: Table showing eigenvalues, proportional and cumulative variance of common components

Factor loadings for the common components were also extracted by principal component analysis. The loadings represent how best each principal component explains the 20 monthly macroeconomic indicator variables. The results indicated that the first common component best explained the variance in coffee productions quantity, the second variations in new vehicle registrations, the third variations in total imports, the fourth variations in tea production quantity, the fifth variation in visitor arrivals, the sixth variations in foreign reserves, the seventh variations in exchange rates and the final component best explained the variations in monthly inflation rate. A summary presentation of the factor loadings on the indicator variables is presented in the graph below.

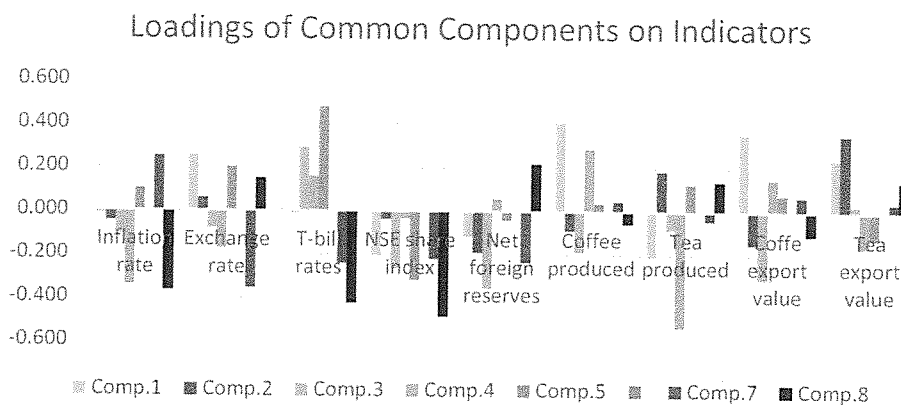


Figure 2: Common factor loadings on the monthly macroeconomic indicators

The extraction of common components allows for regressing of quarterly GDP growth on the eight factors. The ordinary least squares approach was used to extract the coefficients of the regression equation. Coefficients were extracted at a 95% confidence interval. The resulting nowcasting equation was consequently defined by the given equation. The equation provides the full specification of the dynamic factor model.

$$GDP^Q = 0.041 + 0.084f_1 - 0.124f_2 + 0.140f_3 + 0.30f_4 + 0.012f_5 + 0.326f_6 - 0.239f_7 - 0.042f_8 + 0.02422$$

The full specification of the dynamic factor model equation allow for the fitting of quarterly GDP on common components. As shown on the graph below, the fitted values match up to the actual growth rates trends. The trends in economic cycle, slumps in performance and upswings are both timely captured by the model. The fitted values provide comparable estimates of current quarter economic performance relative to the actual rates. As noted by Giannone and Reichlin (2006) the dynamic factor model nowcast outperforms benchmark models by making timely and near accurate GDP estimation with marginal errors.

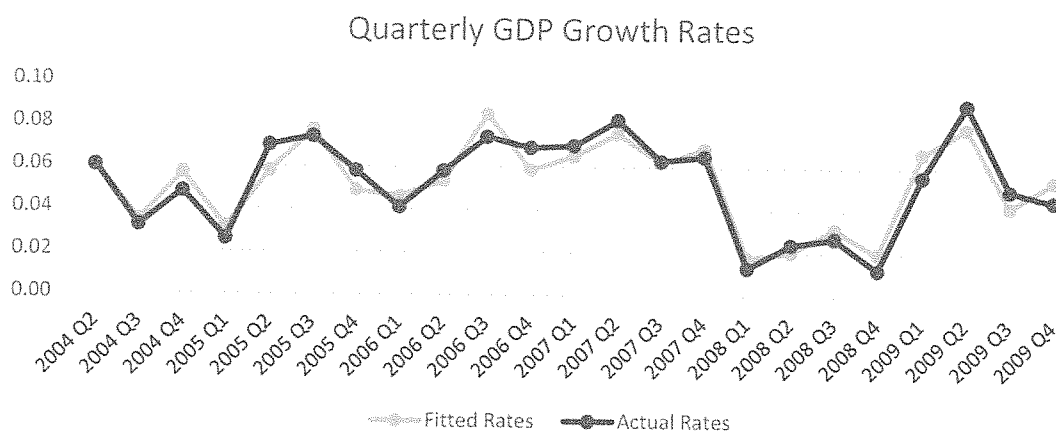


Figure 3: A graph of actual versus fitted quarterly GDP growth rates

### 5.1 Out of Sample Evaluation

The in sample evaluation provided the basis for the out of sample evaluation. The sampling period ranged from the first quarter of 2010 to the second quarter of 2015. Missing values in the data set were computed by applying an algorithm to past historical values for an estimate, the kalman filter technique. As was with the in sample exercise, the eight principal components were extracted by principal component analysis. The loadings of the common factors are as shown below.

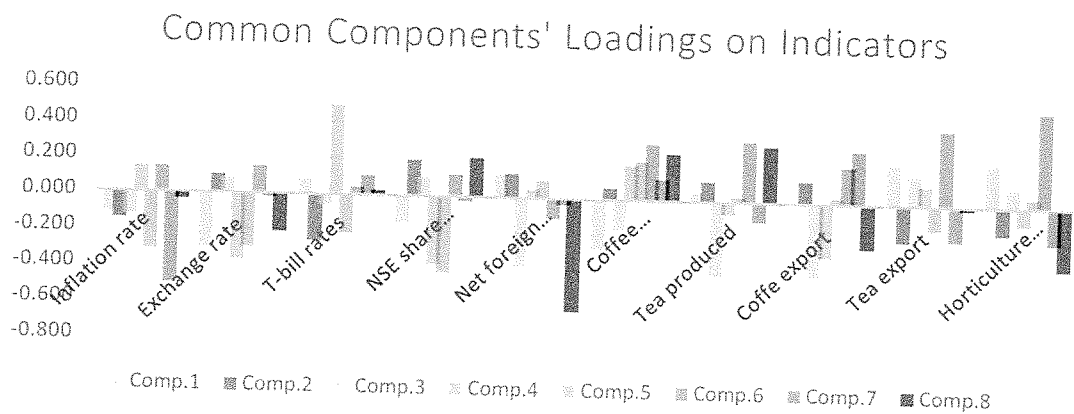


Figure 4: Common component loadings on monthly macroeconomic indicators

The common components are regressed against the dynamic factor model equation to obtain the quarterly economic growth rates. The graph below shows the plot of the fitted values versus the actual values over the nowcasting horizon, 2010 to 2015. The out of sample results are consistent with the in sample result as the nowcasts provides a satisfactory estimate of current quarter GDP. The average percentage error in the nowcast is 0.23%.

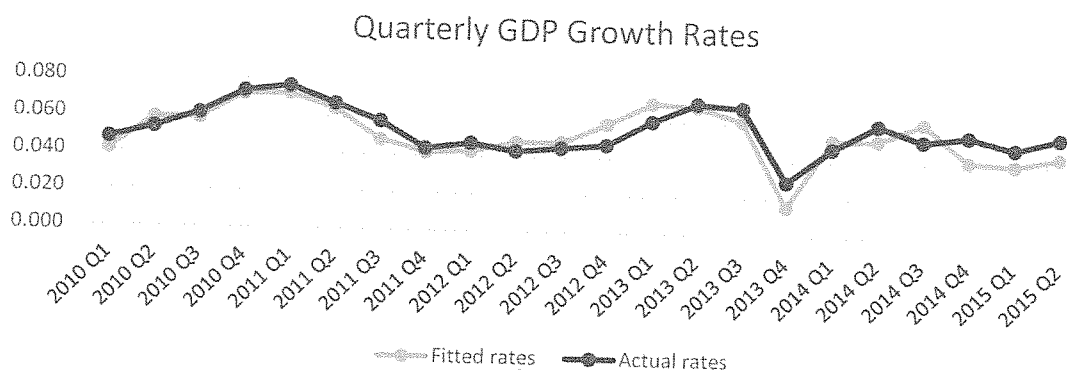


Figure 5: A graph of actual versus fitted quarterly GDP

## 6. Conclusion

The dynamic factor model provides a means to make timely estimates of quarterly GDP. The model allows for extraction of a few unobservable factor that explain the variation in indicator variables. This study established that principal component analysis and the information criteria are accurate methods in reducing the number of unobservable factors to an optimal number. Based on the eight extracted components, the dynamic factor model provides a precise estimate of current quarter GDP. In addition, based on the findings, the model can be applied to the Kenyan case and used to come up with early realistic growth rates for GDP.

The scope of this research centered on establishing the usefulness of the dynamic factor in providing timely estimates for Kenya's quarterly GDP. Areas for further research in the application of the model would be to examine the effects of data revisions and news to the factor model.

## Appendix

Monthly Macroeconomic Indicator	Logs	First Difference
Kenyan month on Inflation rate		X
Exchange rate to US Dollar	x	x
Interest rates (T-Bills)		x
NSE share index	x	x
Net foreign exchange reserves	x	x
Coffee sales quantity	x	x
Tea sales quantity	x	x
Coffee export value	x	x
Tea export value	x	x
Horticulture export value	x	x
Export value	x	x
Total imports	x	x
Total exports	x	x
Electricity generation	x	x
Electricity consumption	x	x
Consumption of petroleum fuels	x	x
Cement consumption	x	x
Cement production	x	x
Value of buildings approved	x	x
Visitor arrivals at JKIA	x	x
New vehicle registration	x	x

*Table 2: X indicates the type of transformation applied to the dataset. Where absent, it means the specific data transformation was not applied.*

## References

- Alessi, L., Barigozzi, M. & Capasso, M. (2010). *Improved penalization for determining the number of factors in approximate factor model*, *Statistics & Probability Letters* 80: 1806-1813.
- Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L. & Rünstler, G. (2011). *Short-Term Forecasts of Euro Area GDP Growth*, *Econometrics Journal*, 14(1), C25–C44.
- Bai, J. & Ng, S. (2007). *Determining the number of primitive shocks in factor models*, *Journal of Business and Economic Statistics*, 25, pp. 52-60.
- Bai, J. & Wang, P. (2015). *Identification and Bayesian Estimation of Dynamic Factor Models*, *Journal of Business & Economic Statistics*, 33(2), 221-240.
- Barhoumi, K., Ferrara, L., & Darné, O. (2014). “*Dynamic factor models: A review of the literature*”, *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 8(2), 73-107.
- Boivin, J. & Ng, S. (2003). *Are more data always better for factor analysis?* *Journal of Econometrics* 127, 169–194.
- Breitung, J. & U. Pigorsch (2013). *A Canonical Correlation Approach for Selecting The Number of Dynamic Factor*, *Oxford Bulletin of Economics and Statistics*, 75, 1, 23-36.
- Bums, A. F., and Mitchell, W. C. (1947). *Measuring Business Cycles*. New York: National Bureau of Economic Research.
- Dixit, S.S., & Opoku-Afari, M. (2012). *IMF Working Paper: Tracking Short-Term Dynamics of Economic Activity in Low-Income Countries in the Absence of High-Frequency GDP Data*. USA: International Monetary Fund.
- Doz, C., Giannone, D., & Reichlin, L. (2006). *A quasi maximum likelihood approach for large approximate dynamic factor models*, Working paper series, European Central Bank.
- Eklund, J. & Kapetanios, G. (2008). *A review of forecasting techniques for large datasets*, *National Institute Economic Review*, 203(1), 109-115.
- Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000). *The generalized dynamic-factor model: Identification and estimation*, *Review of Economics and Statistics*, 82, 540-554.

- Giannone, D., Reichlin, L., & Small, D. (2008). *Now-casting: The real-time informational content of macroeconomic data*, *Journal of Monetary Economics*, 55(4), 665–676.
- Golinelli, R. & Parigi, G. (2007). *The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries*, *Journal of Forecasting*, 26, pp. 77–94.
- Gupta, R., & Kabundi, A. (2011). *Forecasting macroeconomic variables using large datasets: Dynamic factor model versus large-scale BVARs*, *Indian Economic Review*, 23-40
- Hallin, M., and Liška, R., (2007). *The Generalized Dynamic Factor Model: Determining the Number of Factors*, *Journal of the American Statistical Association*, 102, 603-617.
- Kapetanios, G. (2010). *An alternative method for determining the number of factors in factor models with large datasets*, *Journal of Business and Economic Statistics*, 28, 397-409.
- Liu, P., Romeu, R., & Matheson, T. (2011). *Real-time Forecasts of Economic Activity for Latin American Economies*, Washington: International Monetary Fund.
- Stock, J. H. , & Watson, M. W (2002). *Macroeconomic forecasting using diffusion indexes*, *Journal of Business and Economic Statistics*, 20, pp. 147-162.
- Stock, J. H. , & Watson, M. W. (2011). *Dynamic factor models*. *Oxford Handbook of Economic Forecasting*, 1, 35-59.