CHAPTER 5

A MARKOV SWITCHING REGIME MODEL OF THE SOUTH AFRICAN BUSINESS CYCLE

5.1 INTRODUCTION

According to theory, the behavior of stock market investors and hence the behavior of stock prices is potentially asymmetric conditional on the business cycle (see chapter three). In order to empirically evaluate and estimate this asymmetry, an indicator of the business cycle has to be developed. This indicator should ideally reflect not only whether the economy is in a recession or an expansion, but also the degree of certainty with which investors can regard the economy as being in a recession or expansion. In this chapter, such an indicator will be developed by estimating a Markov switching regime model for the business cycle.

Hamilton (1989) first introduced the Markov switching regime model, a stochastic regime model, to business cycle modeling. He applied it to economic growth and his model has been increasingly used to assist in the dating and forecasting of turning points in the business cycle. The model is conceptually appealing in that over time the variable of interest, such as some appropriate measure of the business cycle, is regarded as having a certain probability of switching abruptly among a number of regimes. In the case of the business cycle, expansions and contractions might be considered as the two regimes, each with unique characteristics such a unique mean and variance. In other words, the business cycle switches between a high-growth and a low-growth regime.

These discrete shifts have their own dynamics, specified as a Markov switching regime process. An attractive feature of the model is that no prior information regarding the dates when the economy was in each regime, or the size of the two growth rates is required. This is in contrast with models such as probit and logit models that requires and depends heavily upon the exact dates of all the regimes in the history of the series. Instead, the probability of being in a particular regime is inferred from the data.

In this chapter, the South African business cycle will be modeled with a Markov switching regime model. The purpose of the Markov switching regime (MS) model is two-fold. First, it estimates the data generating process (DGP) of the variable under consideration in this case economic growth. Second, it can be used to classify each observation into one of two regimes, which can in turn be used to predict turning points in the cycles when a number of observations in one regime is followed by a number of observations in the other regime. In the empirical analysis, the performance of the MS model in each of these two aspects will be compared against other models with the same purpose. Specifically, the performance of the MS model in terms of modeling the growth rate will be compared against an autoregressive model. Likewise, the accuracy of the turning points predicted by the MS model will be compared against the outcomes of a logit model.

It has became increasingly popular to use the yield spread as explanatory or information variable to model business cycle turning points (see e.g. Estrella and Hardouvelis (1991), Bernard and Gerlach (1996), Estrella and Mishkin (1998)). In this chapter, the yield spread will be used as explanatory or information variable in both the Markov switching regime and the logit models (see Appendix 1 for a comparison of the performance of the yield spread and other indicators in predicting business cycle turning points).

The outline of this chapter is as follows: The next section will summarize the theory of the lagged relationship between the yield spread and the business cycle. In section 5.3, the Markov switching regime and logit techniques are exposed. Section 5.4 provides an overview on the empirical literature of modeling the business cycle with the Markov switching regime technique, as well as empirical models of the relationship between the yield spread and the business cycle. The estimation results are presented in section 5.5.

5.2 THE RELATIONSHIP BETWEEN THE BUSINESS CYCLE AND THE YIELD SPREAD

There are two explanations for the relationship between the business cycle and the term structure of interest rates or the yield spread between similar long-term and short-term interest rates (the so-called "yield spread"). For the first explanation, assume that the country is currently enjoying high growth, so that there is a general agreement among investors that the country is heading for a slow-down or recession in the future. Consumers want to hedge against the recession and therefore purchase financial instruments (e.g. long-term bonds) that will deliver pay-offs during the economic slowdown. The increased demand for long-term bonds causes an increase in the price of long-term bonds, in other words a decrease in the yield on long-term bonds. In order to finance these purchases, investors sell their shorter-term assets, which results in a decline in the price of short-term assets and an increase in the yield on short-term assets. In other words, if a recession is expected, long-term interest rates will fall and short-term interest rates will rise. Consequently, prior to the recession, the slope of the term structure of interest rates will become flat (or even inverted), which means that the yield spread declines. Similarly, long-term interest rates rises while short-term interest rates falls when an expansion is expected, so that an upward-sloping yield curve predicts an expansion.

The second explanation is based on the expectations hypothesis of the term structure of interest rates. This hypothesis is based on the assumption that similar financial instruments with different maturities are perfect substitutes, so that an investor will be indifferent between investing in one long-term instrument or several similar consecutive short-term instruments, as long as their expected returns are equal (Mishkin 1998:156). This means that, for similar financial instruments, the long-term yield will be the average of current and future short-term yields. Assume that a central bank tightens monetary policy by raising short-term rates. Economic agents will view this as a temporary shock and therefore they expect future short-term rates to rise by less than the current change in short-term interest rates. Based on the expectations hypothesis of the term structure, long-term rates will rise by less than the current short rate. This will lead to a flatter or even an inverted yield curve. Since monetary policy affects economic activity with a lag of one to two years, the tightening of policy will cause a reduction of future economic activity and an increase in the probability of a recession. Therefore, prior to a recession (expansion), the yield spread will decline (increase).

5.3 THE ECONOMETRIC TECHNIQUES

5.3.1 The Markov Switching Regime Model

(i) The Markov switching regime model with fixed transition probabilities

Assume that there are two regimes, represented by an unobservable process denoted S_t. Let S_t take on the values 0 and 1, depending on the prevailing regime. Then the data generating process of the series being modeled, Y_t, will be different in each regime, for example

$$Y_t = \phi_{0,0} + \phi_{1,0} Y_{t-1} + \ldots + \phi_{p,0} Y_{t-p} + \varepsilon_{t,0} \qquad \text{if } S_t = 0 \tag{5.1}$$

$$Y_t = \phi_{0,1} + \phi_{1,1} Y_{t-1} + \ldots + \phi_{p,1} Y_{t-p} + \varepsilon_{t,1} \qquad \text{if } S_t = 1 \tag{5.2}$$

where $\varepsilon_{t,j} \sim N(0, \sigma_j^2)$.

Following Hamilton (1989), assume that S_t is a first-order Markov-process, which means that the current regime (S_t) depends only on the regime in the preceding period (S_{t-1}). The model is completed by defining the transition probabilities of moving from one regime to another, called the transition probabilities:

$$P(S_t=j|S_{t-1}=i) = p_{ij} \qquad i, j = 0,1 \tag{5.3}$$

Notice that, since $p_{01} = 1 - p_{00}$ and $p_{10} = 1 - p_{11}$, the transition probabilities are completely defined by p_{00} and p_{11}.

Let Ω_{t-1} be the information matrix at time t-1:

$$\Omega_{t-1} = (Y_{t-1}, Y_{t-2}, \ldots Y_1). \tag{5.4}$$

Assuming that ε_t in equations 5.1 and 5.2 are Gaussian, the density of Y_t conditional upon the history Ω_{t-1} and S_t is

$$f(Y_t \mid S_t = j, \Omega_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-(Y_t - \phi'_j X_t)^2}{2\sigma^2} \right\} \tag{5.5}$$

where $X_t = (1, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})'$

$\phi_j = (\phi_{0,j}, \phi_{1,j}, \ldots, \phi_{p,j})'$

$\theta = (\phi_1', \phi_2', p_{00}, p_{11}, \sigma^2)'$

$j = 0,1$

$t = 1, \ldots, n$

n is the sample size.

Since the regime S_t is unobservable, the conditional log likelihood for the *t*th observation $l_t(\theta)$ is given by the log of the density of Y_t conditional only upon the history Ω_{t-1}, that is:

$$l_t(\theta) = \ln f(Y_t \mid \Omega_{t-1}; \theta) \tag{5.6}$$

where

$$f(Y_t \mid \Omega_{t-1}; \theta) = f(y_t, S_t = 0 \mid \Omega_{t-1}; \theta) + f(y_t, S_t = 1 \mid \Omega_{t-1}; \theta)$$

$$= \sum_{j=0}^{1} f(y_t \mid S_t = j, \Omega_{t-1}; \theta).P(S_t = j \mid \Omega_{t-1}; \theta) \tag{5.7}$$

In order to calculate this density, the conditional probability of being in a regime given the history of the process, $P(S_t = j \mid \Omega_{t-1}; \theta)$, has to be quantified. If the regime at time t-1 were known, the optimal forecasts of the regime probabilities would be

$$\hat{\xi}_{t|t-1} = P.\xi_{t-1} \tag{5.8}$$

where

$$\hat{\xi}_{t|t-1} = \begin{bmatrix} P(S_t = 0 \mid \Omega_{t-1}; \theta) \\ P(S_t = 1 \mid \Omega_{t-1}; \theta) \end{bmatrix} \text{ and } P = \begin{bmatrix} p_{00} & 1-p_{11} \\ 1-p_{00} & p_{11} \end{bmatrix}$$

$\xi_{t-1} = (1,0)'$ if $S_{t-1} = 0$

$\xi_{t-1} = (0,1)'$ if $S_{t-1} = 1$.

However, S_{t-1} is unobservable therefore ξ_{t-1} is replaced by an estimate of the probabilities of each regime occurring at time t-1 conditional on all information up to and including observation t-1. Let $\hat{\xi}_{t-1|t-1}$ be the optimal inference concerning the regime probabilities. Then

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \Theta f_t}{1' \hat{\xi}_{t|t-1} \Theta f_t}. \tag{5.9}$$

Given $\hat{\xi}_{1|0}$ and $\hat{\theta}$, the optimal forecast and inference for the conditional regime probabilities can be calculated by iterating on the following two equations:

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \Theta f_t}{1' \hat{\xi}_{t|t-1} \Theta f_t} \tag{5.10}$$

$$\hat{\xi}_{t+1|t} = P.\xi_{t|t} \tag{5.11}$$

where f_t denotes the vector containing the conditional densities for the two regimes, **1** is a 2x1 vector of ones and the symbol Θ indicates element-by-element multiplication. The necessary starting values $\hat{\xi}_{1|0}$ can either be taken to be affixed vector of constants which sum to unity, or can be included as separate parameters that need to be estimated. Hamilton (1994:693) provides an intuitive explanation of why this algorithm works.

Finally, let $\hat{\xi}_{t|n}$ denote the smoothed inference on the regime probabilities, in other words, the estimates of the probability that regime j occurs at time t given all available observations in the sample:

$$\hat{\xi}_{t|n} = P(s_t = j \mid \Omega_n; \theta). \tag{5.12}$$

Kim (1993) developed an algorithm to calculate the smoothed inference probabilities:

$$\hat{\xi}_{t|n} = \hat{\xi}_{t|t} \otimes \left(P' \left[\hat{\xi}_{t+1|n} \div \hat{\xi}_{t+1|t}\right]\right) \tag{5.13}$$

where $\div$ indicates element-by-element division and $\otimes$ indicates element-by-element multiplication. The algorithm runs backwards though the sample, that is, starting with $\hat{\xi}_{n|n}$ from the inference regime probabilities up to $\hat{\xi}_{1|n}$.

It was shown by Hamilton (1990) that the maximum likelihood estimates of the transition probabilities are given by

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{n} P\left(S_{t=j}, S_{t-1} = i \mid \Omega_n; \hat{\theta}\right)}{\sum_{t=2}^{n} P\left(S_{t-1} = i \mid \Omega_n; \hat{\theta}\right)}. \tag{5.14}$$

The maximum likelihood estimates of the transition probabilities satisfy the following first order conditions (Hamilton 1990):

$$\hat{\sigma}_j = \frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{2}\left(y_t - \hat{\phi}'_j x_t\right)^2 P\left(s_t = j \mid \Omega_n; \hat{\theta}\right) \tag{5.15}$$

and

$$\hat{\phi}_j = \left(\sum_{t=1}^{n} x_t(j) x_t(j)'\right)^{-1}\left(\sum_{t=1}^{n} x_t(j) y_t(j)\right). \tag{5.16}$$

In other words, the maximum likelihood estimates of σ^2 and ϕ_j can be obtained by estimating a weighted least squares regression of y_t on x_t, where the weights are given by the square root of the smoothed probability of regime j occurring. Therefore, the maximum likelihood estimate of ϕ_j is the vector of coefficients in a regression of $y_t(j)$ on $x_t(j)$, where

$$y_t(j) = y_t\sqrt{P\left(S_t = j \mid \Omega_n; \hat{\theta}\right)} \tag{5.17}$$

$$x_t(j) = x_t\sqrt{P\left(S_t = j \mid \Omega_n; \hat{\theta}\right)}. \tag{5.18}$$

Putting all the above elements together suggests the following iterative procedure to estimate the parameters of the Markov switching regime model. Start off with an arbitrary initial guess for the value of $\hat{\theta}^{(0)}$, where $\hat{\theta} = \left(\hat{\phi}_1, \hat{\phi}_2, \hat{p}_{11}, \hat{p}_{22}, \hat{\sigma}^2\right)$. This can be used with equations 5.10 to 5.12 to calculate the initial estimates of the smoothed regime probabilities ($\hat{\xi}^{(0)}_{t|n}$). Next, the smoothed regime probabilities are combined with the initial estimates of the transition probabilities ($\hat{p}^{(0)}_{ij}$) to calculate new estimates of the transition probabilities ($\hat{p}^{(1)}_{ij}$). Finally, equations 5.15 and 5.16 can be used to obtain a new set of estimates of the autoregressive parameters ($\hat{\phi}_j$) and the residual variance (σ^2). Combined with the new estimates of the transition probabilities, this gives a new set of estimates for all the parameters in the model, $\hat{\theta}^{(1)}$.

Iterating this process renders estimates for the parameters $\hat{\theta}^{(2)}, \hat{\theta}^{(3)}$, … until convergence occurs, in other words, until the estimates in subsequent iterations are the same. This procedure turns out to be an application of the Expectation Maximization (EM) algorithm developed by Dempster, Laird and Rubin (1977). It can be shown that each iteration of this procedure increases the value of the likelihood function, which guarantees that the final estimates are maximum likelihood estimates (Hamilton 1994: 689).

(ii) The Markov switching regime model with time-varying transition probabilities

The drawback of fixed transition probabilities model set out in the previous section is that it implies that the expected durations of expansions and recessions can differ but are forced to be constant over time. Intuitively, the expected duration of an expansion or contraction is generally thought to vary with the underlying strength of the economy. For example, as the economy exits a relatively deep recession and enters a relatively robust recovery period, the economy is less likely to fall back into the recession at that time (Filardo and Gordon 1998). The assumption that the transition probabilities are time invariant, may be costly from an empirical point of view. With fixed transition probabilities, the conditional expected durations do not vary over the cycle. This implies that exogenous shocks, macroeconomic policies and an economy's own internal propagation mechanisms do not affect the expectation of how long an expansion or recession will last (Filardo and Gordon 1998).

A solution to this problem is to incorporate time-varying transition probabilities (TVTP) into the model, by using a specification for the transition probabilities that reflects information about where the economy is heading. The variations in the transition probabilities will generate variations in the expected durations (Filardo and Gordon 1998).

The time-invariant transition probabilities were

$$P(S_t = s_t \mid S_{t-1} = s_{t-1}, z_t) = \begin{bmatrix} p_{00} & 1-p_{11} \\ 1-p_{00} & p_{11} \end{bmatrix} \qquad (5.19)$$

where $p_{ii} = P(S_t = i|S_{t-1} = i)$.

Instead, the time-varying transition probabilities are

$$P(S_t = s_t \mid S_{t-1} = s_{t-1}, z_t) = \begin{bmatrix} p_{00}(z_t) & 1-p_{11}(z_t) \\ 1-p_{00}(z_t) & p_{11}(z_t) \end{bmatrix} \qquad (5.20)$$

where z_t is the information variable(s) upon which the evolution of the unobserved regime will depend, such as the index of leading indicators, or individual leading indicators such as the term structure of interest rates.

There are three reasons why the time-varying transition probabilities (TVTP) model may be a significant extension of the fixed transition probabilities (FTP) model (Filardo 1994):

- The TVTP model allows the transition probabilities to rise just before a contraction or an expansion begins, while the FTP does not. In an FTP model, the transitions probabilities are constant before, during and after turning points. On the other hand, TVTP models have the flexibility to identify systematic variations in the transition probabilities both before and after turning points.

- The TVTP model is able to capture more complex temporal persistence than an FTP model. Both the FTP and TVTP models can distinguish between two sources of business cycle persistence, namely through the autoregressive (AR) parameters and through the persistence of the phase over time that is reflected in the transition probability matrix. By allowing the transition probabilities to vary over time, the nature of the persistence that can be identified is expanded.

- TVTP are intrinsically linked to the notion of time-varying expected durations in the Markov switching regime framework. In the FTP model, expected duration is constant, while it can vary over time in the TVTP model. Several studies (see e.g. Diebold, Rudebusch and Sichel (1993) and Durland and McCurdy (1993)) have confirmed the intuition that the expected duration of a cycle is not necessarily constant over time and unlike the FTP model, the TVTP model is flexible enough to capture this.

A popular way to model time-varying transition probabilities is to incorporate a simple probit or logit function (see e.g. Filardo and Gordon (1998), Durland and McCurdy (1994) and Bodman (1998)). A probit or logit function can be estimated to measure the transition probability matrix at each time t. This way, the transition probabilities is a function of an economic indicator(s) such as the index of leading indicators (see e.g. Filardo and Gordon (1998)), or an individual leading indicator such as the term structure of interest rates (see e.g. Filardo (1994)). In particular, if a logit function is used the transition probabilities are

$$p_{11} = P(S_t=1 \mid S_{t-1}=1) = \exp(\alpha_1+\beta_1 z_t)/(1+ \exp(\alpha_1+\beta_1 z_t)) \qquad (5.21)$$

$$p_{22} = P(S_t=2 \mid S_{t-1}=2) = \exp(\alpha_2+\beta_2 z_t)/(1+ \exp(\alpha_2+\beta_2 z_t)) \qquad (5.22)$$

The expected duration of a phase is determined by the transition probabilities. This means that variation in z_t and S_{t-1} will affect the expectation of how long a phase will last.

5.3.2 The Logit Model

Several authors have used probit or logit models to model business cycle turning points (see e.g. Estrella and Hardouvelis 1991; Dueker 1997; Dotsey 1998; Estrella and Mishkin 1998; Bernard and Gerlach 1996). The probit or logit form is dictated by the fact that the variable being predicted takes on only two possible values – whether the economy is in a recession or not. The model is defined in reference to a theoretical linear relationship of the form:

$$Y^*_{t+k} = \alpha + \beta * x_t + \varepsilon_t \tag{5.23}$$

where Y^*_t is an unobservable that determines the occurrence of a recession at time t, k is the length of the forecast horizon, ε_t is a normally distributed error term and x_t the value of the explanatory variable at time t. The parameters α and β are estimated with maximum likelihood. The observable recession indicator R_t is related to this model by

$$R_t = 1 \text{ if } Y^*_t > 0 \text{ and } 0 \text{ otherwise} \tag{5.24}$$

The form of the estimated equation is

$$P(R_{t+k} = 1) = F(\alpha + \beta * x_t) \tag{5.25}$$

where F is the cumulative logistic distribution function.

The model is estimated by maximum likelihood. The recession indicator is obtained from the South African Reserve Bank, that is, $R_t = 1$ if the economy is in a recession at time t and 0 otherwise.

5.4 EXISTING MARKOV SWITCHING REGIME BUSINESS CYCLE MODELS

Business cycles have been modeled with different techniques, such as autoregressive integrated moving average (ARIMA) models (e.g. Nelson and Plosser (1982), Beveridge and Nelson (1981) and Campbell and Mankiw (1987)); cointegration techniques (e.g. King, Plosser, Stock and Watson (1991)); and the Kalman filter whereby real gross national product (GNP) is modeled as the sum of unobserved components (e.g. Harvey (1985), Watson (1986), Clark (1987)). These techniques share a potential shortcoming, namely the assumption that the growth rate of real GNP is a linear stationary process. Linear models are incompatible with the asymmetry between expansions and contractions that has been documented by, amongst others, Neftci (1984), Stock (1987), Diebold and Rudebusch (1990) and Sichel (1993).

Hamilton (1989) proposed a Markov switching regime model that models real GNP growth as an autoregressive model of order four (AR(4)), allowing for non-linearity by introducing discrete shifts in the mean between high-growth and low-growth regimes. These discrete shifts have their own dynamics, specified as a two-regime first-order Markov process. The most attractive feature of this model is that no prior information regarding the dates of the two growth periods or the size of the two growth rates is required. In addition, the low-growth rate need not be negative. In this section, a brief overview of the empirical literature on Markov switching regime models for business cycles and on the relationship between the yield spread and the business cycle will be given.

5.4.1 Empirical Markov Switching Regime Business Cycle Models with Fixed Transition Probabilities

Hamilton (1989) developed a Markov switching regime model for dating and forecasting business cycles. He applied this model to the quarterly real GNP of the US for the period 1951 to 1984. In particular, he modeled GNP growth as a AR(4) two regime Markov switching regime (MS) model. In other words, GNP growth switches between two regimes, which each have a unique intercept but he constrained the AR coefficients to be the same across regimes. The MS model calculates the probability that the economy is in a particular regime in a certain period and the econometrician has to devise a dating rule to actually decide from which regime this observation is. Hamilton used a very popular dating rule, which classifies a particular period as a recession (expansion) if the econometrician concludes that the economy is more likely than not to be in a recession (expansion), in other words, when the probability of being in a recession (expansion) is higher than the probability of being in a expansion (recession). The dates of the turning points predicted by his MS model are usually within three months of the dates of the official dates set by the National Bureau for Economic Research (NBER).

Goodwin (1993) used Hamilton's (1989) Markov switching regime model to model the business cycles of eight developed countries. Real GNP growth was allowed to follow an AR(4) process. Hansen's (1992) likelihood ratio test rejected the null hypothesis that the Markov model performs better than linear autoregressive models. However, the filtered and smoothed conditional probabilities indicated business cycle turning points that closely correlate with official turning points. Implicit in much of the research on business cycles going back to Keynes and before, is the notion that business cycles can be characterized as exhibiting sharp drops during contractions followed by gradual movements during expansions. Goodwin tested a closely related idea, namely that contractions have shorter durations than expansions, by comparing the expected durations of expansion and recessions. He rejected the hypothesis of symmetry, in other words that the expected duration of expansion and recessions are equal.

Ivanova, Lahiri and Seitz (2000) used the same technique as Hamilton (1989) and Goodwin (1993), but instead of modeling GNP directly, they modeled a leading indicator and then consider a change in regime as a business cycle turning point signal. In particular, they compared the performance of a number of interest rate spreads as predictors of the German business cycle. They use a two regime, first order Markov switching regime model, in other words they allowed for two regimes where the regime probability in a particular period is only influenced by the regime in the preceding period. They allow the dynamic behavior of the economy to vary between expansions and recessions in terms of duration and volatility. They model the interest rate spread as a univariate Markov switching model with no autoregressive terms, allowing both the intercept and variance to differ across regimes. They define a regime change as the event that the probability of a recession (expansion) is greater than the probability of an expansion (recession). Since the interest rate spread is considered to be a leading indicator of the business cycle, the change in regime is the turning point signal. Their results indicate that the market spreads does follow regimes. None of the bank spreads gave any false signals, but the spread between government and bank bonds of 1-2 years gave multiple false signals. The call rate spread performs slightly inferior to the other spreads, since its predictions lagged the predictions of the other spreads.

Instead of a univariate Markov switching regime model, Kontolemis (1999) used a vector Markov switching regime model to date and forecast US business cycle. In other words, they forced the different indicators to have simultaneous turning points. The four series used in the construction of the coincident index are the index of industrial production, non-agricultural employment, personal income (less transfer payments) and manufacturing and trade sales. Monthly data from 1948 to 1995 was used. Following Hamilton (1989), the rule for dating the business cycle is based on whether the economy is more likely than not to stay in one of the two phases. They imposed a requirement that each cycle is at least 6 months (i.e. two quarters) to eliminate spurious cycles in the monthly series. The estimated probabilities tracked the NBER downturns relatively well. They extended the model to include an autoregressive term, but this model failed to track the NBER reference cycle during the entire sample period. The vector Markov switching model produces more accurate forecasts than a simple univariate Markov switching model specification.

5.4.2 Empirical Markov Switching Regime Business Cycle Models with Time-Varying Transition Probabilities

The models reviewed in section 5.4.1 all assumed constant transition probabilities, which implies that the conditional expected durations are constant as well. Intuitively, however, the expected duration of an expansion or contraction is generally thought to vary with the underlying strength of the economy. For example, as the economy exits a relatively deep recession and enters a relatively robust recovery period, the economy is less likely to fall back into the recession at that time. The time-varying transition probabilities (TVTP) model offers a solution to this problem, by using a specification for the transition probabilities that reflects information about where the economy is heading. The variations in the transition probabilities will generate variations in the expected durations.

Filardo (1994) extended the Markov switching regime model to allow for time-varying transition probabilities. He used a logit function to generate the transition probabilities. He compared different information variables, namely the composite index of leading indicators, the interest rate spread, the Standard and Poor stock index and the short-term interest rate. There was statistically significant evidence that the model supports the two-phase view of the US business cycles, in other words that economic growth switches between a positive growth rate (expansion) and a negative growth rate (recession). In addition, it has been shown that expansions have higher persistence and that of both phases are time-varying. The different leading indicators used contain different information and gave different turning points. His results showed that the business cycle dynamics of this model stem mainly from the variation in the transition probabilities, rather than from a shift in the means.

Durland and McCurdy (1994) also modeled time-varying transition probabilities with a logit function. They modeled the transition probabilities as functions of both the inferred current regime and the associated number of periods the system has been in the current regime. In other words, they allowed the transition probabilities to be duration dependent, so that the probability of staying in, say, a recession, declines the longer the economy is in a recession. They are able to reject the linear model in favor of a duration-dependent parameterization of the regime transition probabilities in a regime-switching model.

Filardo and Gordon (1998) generated the time-varying transition probabilities with a probit function. Specifically, they use the information contained in leading indicator data to forecast the transition probabilities. Their results indicate that the US business cycle can indeed be classified as a two-state model and the turning points predicted by their model are similar to the official turning points.

Probit and logit functions are flexible and have a sensible economic interpretation. However, some studies have reported estimation problems when these functions are applied. In context of smooth transition autoregressive (STAR[1]) modeling, Ocal and Osborn (2000) found exponential STAR (ESTAR) more robust to outlier observations than logistic STAR (LSTAR). Therefore Simpson, Osborn and Sensier (2001) tried to model the time-varying transition probabilities with an exponential function instead of the popular probit or logit functions. The problem with the logistic form is that the interpretation is not as economically intuitive as the logit or probit form and it may not lead to sensible probabilities for certain values of the leading indicator because of its shape. Their results indicate that a constant transition probability Markov switching regime model captures the major recessions of the sample, but the use of leading indicators through the time-varying transition probabilities framework improve this regime recognition. On average, contractions are shorter than expansions.

Layton and Katsuura (2001) compared different techniques to date and forecast US business cycles, using three different composite business cycle indexes. Specifically, they estimated binomial and multinomial probit models, binomial and multinomial logit models and a two-regime Markov switching regime model where the transition probabilities are modeled as logistic functions. All these models estimate the probabilities that the economy is in contraction or expansion. When these probabilities are more than 0.5, the economy are regarded to be in contraction or expansions and, in this way, they date the turning points as derived from the models. They used the R^2, the log likelihood and also the official dates of US business cycles as determined by the NBER as a benchmark for comparison. Their results showed that the MS model performed relatively better than the other models. The MS model overcomes a very real practical and fundamental limitation of the logit and probit specifications as far as their use in real time business cycle phase shift forecasting is concerned. Their estimation requires exact knowledge of the regime of the economy for every observation in the estimation period so as to assign values to the dependent variable in the model.

[1] Like the Markov switching regime model, the STAR model is also a regime switching modeling technique. However, in the STAR model the regime is determined by an observable variable, in contrast with the Markov switching regime model where the regime is determined by an unobservable variable.

5.4.3 The Yield Spread as Predictor of Business Cycles

Estrella and Hardouvelis (1991) were the first to empirically analyze the term structure as a predictor of real economic activity. Their study was based on quarterly data of US GNP growth for the period 1955 to 1988. They used the slope of the yield curve, defined as the difference between the 10-year government bond rate and the 3-month T-bill rate, as explanatory variable. Regressions of future GNP growth on the slope of the yield curve and other information variables showed that a steeper (flatter) slope implies faster (slower) future growth in real output. The estimated constant and coefficient of the yield spread for GNP one to five quarters ahead are approximately 1.70 and 1.30 respectively. The positive constant term implies that a negative slope does not necessarily predict negative future real GNP growth. The forecasting accuracy is the highest five to seven quarters ahead. In addition, they also used a probit model to analyze the predictive power of the term structure on a binary variable that simply indicates the presence or absence of a recession. Their probit model relates the probability of a recession as dated by the NBER during the current quarter to the slope of the yield curve lagged four quarters. The results showed that an increase in the spread between the long- and short-term interest rates implies a decrease in the probability of a recession 4 quarters later.

In addition to the domestic term structure, Bernard and Gerlach (1996) also tested the ability of foreign term structures to predict business cycle turning points in eight industrial countries for the period 1972:1 to 1993:4. Using probit models, they showed that the domestic term spreads are statistically significant in explaining business cycle turning points in all eight countries. The period over which the domestic term spread successfully forecast the turning points vary across countries, but the optimal forecast period range from two to five quarters. In general, downward-sloping (upward-sloping) yield curves have historically been associated with subsequent recessions (expansions).

Estrella and Mishkin (1998) compared the performance of various financial variables, including four term structures of interest rates, stock prices, monetary aggregates, indexes of leading indicators and other economic variables such as GDP, CPI and exchange rates, as predictors of US recessions. They estimated probit models with quarterly data for the period 1959 to 1995. Their results indicated that the yield curve outperforms the other indicators for forecasting beyond one quarter ahead.

The only study on the relationship between the term structure of interest rates and the business cycle in the South Africa economy was done by Nel (1996). Unlike the other studies, he analyzed the contemporaneous relationship with cointegration techniques, instead of the lead-lag relationship dictated by theory. He showed that quarterly real GDP is a positive function of the yield spread between 10-year government bonds and the three month banker's acceptance rate. He found real GDP and the yield spread to be cointegrated and showed that the yield spread is statistically significant in explaining GDP, despite a poor overall fit. While Nel (1996) modeled the level or course of the business cycle, this chapter will focus on predicting only turning points.

5.5 EMPIRICAL ANALYSIS OF THE SOUTH AFRICAN BUSINESS CYCLE

5.5.1 Methodology

The South African business cycle is modeled with linear and non-linear models with data for the period 1978 to 2001. Specifically, the performance of a Markov switching regime model of the South African business cycle will be compared with the performance of a autoregressive model and a logit model. In all the models the leading indicator used as explanatory variable was the yield spread. Like most similar studies (see e.g. Durland and McCurdy (1994), Goodwin (1993) and Simpson, Osborn and Sensier (2001)), the empirical estimation was done on a quarterly basis to avoid the excessive random noise prevalent in monthly data.

Table 5.1 Business Cycle Phases According to SARB Since 1978

Upward phase		Downward phase	
January 1978	August 1981	September 1981	March 1983
April 1983	June 1984	July 1984	March 1986
April 1986	February 1989	March 1989	May 1993
June 1993	November 1996	December 1996	August 1999

Source: South African Reserve Bank, Quarterly Bulletin, various issues.

5.5.2 The Estimated Linear Model

Following the most popular Markov switching regime specification for business cycles, real GDP growth is modeled as an AR(4) process with different intercepts in the two different regimes (see e.g. Hamilton (1989), McCurdy and Durland (1994), Goodwin (1993) and Bodman (1998)). Therefore, in the linear model real GDP growth (Y_t) will be modeled as an AR(4) process.

Table 5.2 Linear Autoregressive Model

Dependent Variable: Y_t				
Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
Y_{t-1}	0.421611	0.108621	3.881486	0.0002
Y_{t-2}	0.080301	0.118026	0.680367	0.4981
Y_{t-3}	-0.021405	0.117935	-0.181498	0.8564
Y_{t-4}	-0.043410	0.107054	-0.405500	0.6861
C	1.136723	0.445845	2.549594	0.0125
R-squared	0.208481	F-statistic	5.728825	
Adjusted R-squared	0.172090	Prob(F-statistic)	0.000386	

Source: Own calculations

In the linear model, only the first autoregressive term is significant. The performance of this model is evaluated in section 6, when it is also compared with the performance of the MS model.

5.5.3 The Estimated Logit Model

Table 5.3 Logit model

Dependent Variable: Rec_t*				
Variable	*Coefficient*	*Std. Error*	*z-Statistic*	*Prob.*
$Spread_{t-2}$	-0.994626	0.204696	-4.859029	0.0000
C	0.514365	0.348941	1.474072	0.1405
S.E. of regression		0.299932	Akaike criterion	0.671411
Sum squared resid		8.096318	Schwarz criterion	0.726232

* Rec_t is a dummy variable that takes on the value 1 if the economy is officially in a recession in period t and 0 if not.

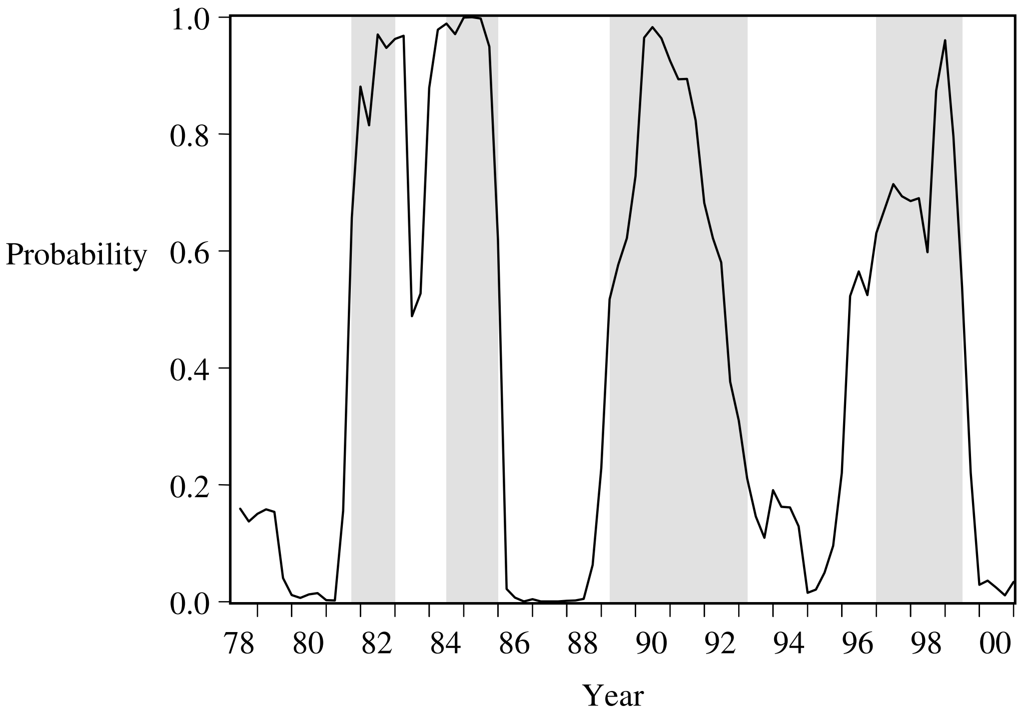
Source: Own calculations

The results in table 5.3 indicate that the probability of a recession in a specific quarter is a negative function of the yield spread lagged two quarters ($spread_{t-2}$). Expressed algebraically

$$P(R_{t+2} = 1) = F(0.514 - 0.995^* x_t) \qquad (5.26)$$

where F is the cumulative logistic distribution, x_t is the yield spread in period t and R_t is a dummy variable that takes on the values one if the economy is in a recession in period 1. In other words, an increase in the spread between the long-term and short-term interest rates implies a decrease in the probability of a recession two quarters later. According to the results in table 5.3, the relationship between the probability of a recession and the yield spread is statistically significant.

Figure 5.1 Recession Probabilities of the Logit Model



Source: Own calculations

Figure 5.1 plots the estimated probability of a recession derived from the historical data on the yield spread lagged two quarters, the parameter estimated in table 5.3 and the cumulative logistic distribution. The shaded areas denote periods of actual recessions as classified by the South African Reserve Bank.

In seven of the eight turning points, the peak of the estimated probability of a turning point preceded the actual turning point by zero to two quarters, in other words the yield spread predicted turning points two to four quarters ahead. The only exception was the upswing in April 1983, when the estimated probability of a recession declined but was higher than with the other upswings. This means that, based on a dating rule that classifies recessions (expansions) as estimated probabilities above (below) 50 percent, the model missed only the upswing in 1983. (However, if the dating rule classifies recessions (expansions) as estimated probabilities above (below) 0.7, the model predicted all the turning points.) If the upswing of 1983 is excluded, the peak of the estimated probability coincided with all the turning points, except for the expansion from June 1993 to November 1996 when it preceded the turning point by two quarters. However, this imperfection should be seen in perspective. For most market participants, the cost of expecting the turning point too early is lower than the cost of expecting the turning point too late. A crucial characteristic of this model is that it did not give any false signals.

5.5.4 The Estimated Markov Switching Regime Model

A first-order, two-regime Markov switching regime model was estimated for the South African business cycle. The model was specified as follows:

$$\begin{aligned} Y_t &= \mu_0(1-S_t)+\mu_1 S_t+\phi_1(Y_{t-1}-(\mu_0(1-S_{t-1})+\mu_1 S_{t-1}))+\phi_2(Y_{t-2}-(\mu_0(1-S_{t-2})+\mu_1 S_{t-2})) \\ &+\phi_3(Y_{t-3}-(\mu_0(1-S_{t-3})+\mu_1 S_{t-3}))+\phi_4(Y_{t-4}-(\mu_0(1-S_{t-4})+\mu_1 S_{t-4}))+\varepsilon_t \end{aligned} \quad (5.27)$$

where $\varepsilon_t \sim N(0,\sigma^2)$

$S_t = 1$ if low-growth regime, 0 otherwise

$P(s_t=j|s_{t-1}=i) = p_{ij,t}$ $\qquad$ $i, j = 0,1.$

Notice that, since $p_{10,t} = 1 - p_{11,t}$ and $p0_{1,t} = 1 - p_{00,t}$, the transition probabilities are completely defined by $p_{11,t}$ and $p_{00,t}$.

Following Filardo (1994), Durland and McCurday (1994), amongst others, the transition probabilities were modeled with a logit function:

$$p_{11,t} = p(S_t = 1 | S_{t-1} = 1) = \exp(\alpha_1 + \beta_1 z_{t-k})/(1+\exp(\alpha_1 + \beta_1 z_{t-k})) \quad (5.28)$$

$$p_{00,t} = p(S_t = 0 | S_{t-1} = 0) = \exp(\alpha_0 + \beta_0 z_{t-k})/(1+\exp(\alpha_0 + \beta_0 z_{t-k})) \quad (5.29)$$

where z_t is the yield spread and α and β the coefficients estimated with maximum likelihood.

Table 5.4 presents significant evidence to support the assumption that two distinct growth-rate phases characterize the business cycle. The point estimates of the regime-dependent means, μ_1 and μ_0, are statistically different. More important, their magnitudes differ significantly and economically. The mean growth rate in the high-growth regime, μ_0, is significantly positive, while the mean growth rate in the low-growth regime, μ_1, is significantly negative. Because the sample dichotomizes into phases that exhibit declining aggregate output and growing aggregate output, each can be labeled as low-growth and high-growth regimes of the economy.

Table 5.4 Parameters of Growth Equation in Markov Switching Regime Model

Growth Model:

$$Y_t = \mu_0(1-S_t) + \mu_1 S_t + \phi_1(Y_{t-1} - (\mu_0(1-S_{t-1}) + \mu_1 S_{t-1})) + \phi_2(Y_{t-2} - (\mu_0(1-S_{t-2}) + \mu_1 S_{t-2})) + \phi_3(Y_{t-3} - (\mu_0(1-S_{t-3}) + \mu_1 S_{t-3})) + \phi_4(Y_{t-4} - (\mu_0(1-S_{t-4}) + \mu_1 S_{t-4})) + \varepsilon_t$$

Parameter	*Coefficient*	*Std. Error*
μ_1	-1.061275	0.287213
μ_0	3.741749	0.313490
ϕ_1	0.332210	0.064285
ϕ_2	0.035363	0.067236
ϕ_3	-0.032597	0.068706
ϕ_4	0.001868	0.067109
σ^2	2.693322	0.293941

Source: Own calculations

According to the results in table 5.5, all the estimated coefficients in the generation process of the transition probabilities are significant. The parameters that govern the time-variation of the transition probabilities, β_1 and β_0, have different signs. This is consistent with the intuition that an increase in the yield spread decreases the probability of remaining in an expansion and increases the probability of remaining in a recession (see section 5.4.3). The parameters α_0 and α_1 determine the unconditional mean duration of recessions and expansions. The estimates capture the potential asymmetry in duration across expansions and recessions.

Table 5.5 Parameters of Transition Probability Equation in Markov Switching Regime Model

TVTP model: $p_{ii,t} = p(S_t = i \mid S_{t-1} = i) = \exp(\alpha_i + \beta_i z_{t-k})/(1+\exp(\alpha_i + \beta_i z_{t-k}))$

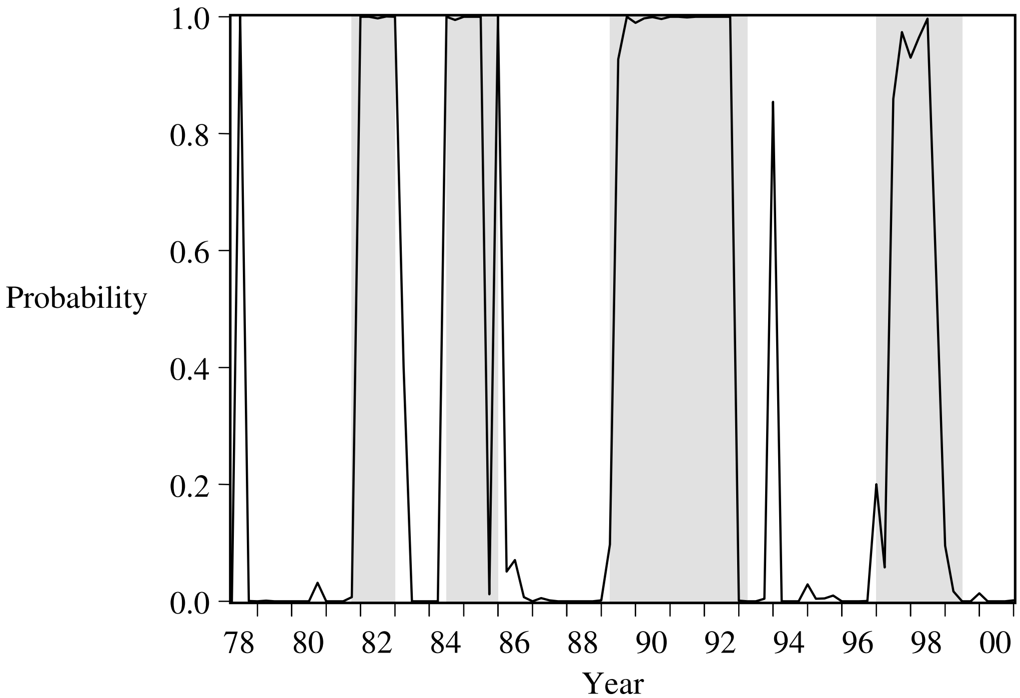
Parameter	*Coefficient*	*Std. Error*	*T-value*
α_1	-0.880836	0.536753	1.64
β_1	-0.784035	0.418566	1.87
α_0	1.250595	0.555241	2.25
β_0	0.388441	0.184527	2.11

Source: Own calculations

Figure 5.2 plots the inferred probability of a low-growth-rate regime given the available data. When above (below) 0.5, the economy is more likely to be in a recession (expansion). The inferred regimes of the FTP model correspond to the official cycles of the SARB. The shaded areas represent the official recessions.

The turning points predicted by the Markov switching regime model are highly correlated with the dates of the official turning points and the regime probabilities are generally very close to 0 or 1, so it is always explicitly indicating one of the regimes. The Markov switching regime model gave “false” signals of an expansion in 1985 and a recession in 1994, but both these signals only lasts for 1 quarter and can therefore be eliminated by applying the common dating rule that a cycle should last for at least 2 quarters. However, instead of regarding these signals as “false” simply because they do not correspond to the official dates, a careful analysis of the periods during which they occurred might show that they were not truly false in the sense of incorrectly indicating the general state of the economy.

Figure 5.2 Markov Switching Regime Model: Time-Varying Transition Probabilities



Source: Own calculations

The definition used by the Reserve Bank is to classify a recession as at least two consecutive quarters of negative economic growth. In other words, if only a single quarter of negative growth is experienced it will not be reflected by the official recessions. For example, during the first quarter of 1994, the economy was contracting by 0.6 percent but since the previous and following quarters both had positive economic growth this was not defined as a recession. The high recession probability in the first quarter of 1994 therefore are reflecting this drop in economic growth rather than giving a false signal. Likewise, the low recession probability in the last quarter of 1985 corresponds to a positive economic growth rate, but since growth was negative during the following quarter the economy was officially still in a recession. This was also the case with the third quarter of 1978. This means that the differences between the Markov switching regime model and the official classification should not be viewed as "false" signals, but should rather be viewed as additional information given by the Markov switching regime model regarding the true state of the economy which are not influenced by an asymmetric classification definition.

5.6 MODEL SELECTION

As stated earlier, the purpose of Markov switching regime model is two-fold, namely to model economic growth, as well as to model the dating of the two regimes. In this section, the two types of results of the Markov switching regime model will be compared with two corresponding types of models. First, the Markov switching regime model's accuracy in modeling economic growth will be compared with two linear models. Second, the Markov switching regime model's accuracy in predicting business cycle turning points will be compared with the turning points predicted by a logit model.

5.6.1 Comparing Linear and Markov Switching Regime Models

Table 5.6 Model Selection Criteria for the Linear and Markov Models

Criteria	Linear model	Markov model
MAPE	1.13	1.48
RMSE	3.31	2.99
MAE	2.46	2.20
Theil's U	0.48	0.36

Source: Own calculations

The mean absolute percentage error (MAPE), the square root of the mean squared error (RMSE), the mean absolute error (MAE) and Theil's inequality coefficient (U) were used to compare the linear and MS models. The Markov switching regime model was preferred to the AR(4) models by all the criteria.

5.6.2 Comparing the Estimated Logit and Markov Switching Regime Models

Criteria:

(i) Number of wrong predictions: $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

(ii) Sum of Squared Residuals (SSR): $\sum_{i=1}^{n}\left(y_i - F\left(x_i'\hat{\beta}\right)\right)^2$

(iii) Sum of Absolute Value of Residuals: $\sum_{i=1}^{n}| y_i - F\left(x_i'\hat{\beta}\right)|$

(iv) Efron's (1978)[2] R^2: $R^2_{Efron} = \dfrac{\sum (y_i - F(x_i'\beta))^2}{\sum (y_i - \hat{y}_i)^2}$

where $\hat{y}_i = 1$ if $F\left(x_i'\hat{\beta}\right) \geq 0.5$ and $\hat{y}_i = 0$ if $F\left(x_i'\hat{\beta}\right) < 0.5$.

The model selection criteria for the logit and Markov switching regime models are given in table 5.7. The preferred model according each criterion is indicated in bold print.

Table 5.7 Model Selection Criteria for Logit and MS Models

Criteria	**MS model**	**Logit model**
Number of wrong predictions	12	**11**
Sum of squared errors	9.58	**8.03**
Efron's R^2	**0.57**	0.50
Sum of absolute errors	**14.58**	16.67

Source: Own calculations

[2] The usual R^2 is calculated as $\sum \hat{y}_i^2 / \sum y_i^2$.

The results in table 5.7 indicate that the Markov model made fewer wrong predictions than the logit model with regards to the inferred regime or regime of the economy. However, this criterion penalizes a model only for the number of times that it is wrong, without taking into account the size of the wrong probability. According to the sum of squared errors, the logit model is preferred to the Markov model. However, since the errors will always lie between zero and one, the larger the error the smaller its square will be. When the sum of the absolute values of the errors is used instead, the Markov model is preferred to the logit model.

It should be kept in mind that the logit model is designed to try to get the best fit for the official turning points. The Markov model, on the other hand, does not use the official turning points in its estimation at all. Against this background, the Markov model actually compares extremely well with the logit model and did make the fewest mistakes.

5.7 CONCLUSION

According to theory, the behavior of stock market investors and hence the behavior of stock prices is potentially asymmetric conditional on the business cycle (see chapter three). In order to empirically evaluate and estimate this asymmetry, an indicator of the business cycle has to be developed. This indicator should ideally reflect not only whether the economy is in a recession or an expansion, but also the degree of certainty with which investors can regard the economy as being in a recession or expansion. In this chapter, such an indicator has been developed by estimating a Markov switching regime model for the business cycle.

The South African business cycle has been modeled with a two-state first-order Markov switching regime with time-varying transition probabilities, with the logit technique and with a autoregressive model. The transition probabilities and the logit model were estimated with the yield spread as explanatory variable. The results indicated that two distinct growth rate phases, a low and a high growth rate phase, characterize the business cycle. It was showed that the Markov switching regime model outperformed both the linear and logit models and even provided more information regarding the state of the business cycle than the official classification of the Reserve Bank. Therefore this indicator is ideal for capturing the state of the business cycle as well as the (un-)certainty regarding this state and can therefore be used in the stock market model to test the influence of these factors.