A MULTIPLE MARKOV SWITCHING MODEL FOR ACTUARIAL USE IN SOUTH AFRICA

By AJ Maitland

ABSTRACT

This paper introduces a new class of Markov switching models where switches in variables are not perfectly correlated. Maximum-likelihood estimates of the parameters are derived and shown to require only the smoothed inferences obtained from a univariate analysis of the variables. The framework is used to estimate a multiple Markov switching (MMS) model of South African financial and economic variables, which can be used for various actuarial applications, especially those involving long-term projections. Users may wish to set certain parameters in relation to future expectations rather than simply using estimates based on past data, but that process is not covered in this paper.

CONTACT DETAILS

School of Statistics and Actuarial Science, University of the Witwatersrand, P.O. Box 2484, Parklands, 2121, South Africa. Tel: +27 11 486 1946. E-mail: james@stochasticsolutions.co.za

1.   INTRODUCTION

1.1   Early actuarial stochastic models assume that the dynamic process for various economic and financial variables is linear. For example, Wilkie's (1986) model for UK inflation uses an AR(1) process to describe the data. Thomson (1996) uses a linear transfer function model to model inflation, with equity dividend growth as the input. Other linear models include those of Carter (1991), Claasen (unpublished) and Sherris *et al* (1999).

1.2   One of the primary assumptions of such models is that certain key variables are stationary. For example, both Wilkie (1986, 1995) and Thomson (1996) assume that various yields and inflation are stationary, although standard unit-root tests suggest that these variables may be integrated (see Maitland, unpublished b). The implication of a unit root in a time series is that shocks to the system are permanent, trends are stochastic and forecast variances increase linearly as the lead time of the forecast increases. Hence, stationarity is a necessary assumption for producing reasonable long-term projections.

1.3   Maitland (unpublished a) shows that the Thomson (1996) transfer-function models suffer from a number of statistical problems and estimation errors. Mean reversion in certain variables of the model creates risk-adjusted returns that are unrealistic and gives rise to predictability that violates the efficient market hypothesis (EMH). Maitland (unpublished a, b) also shows that the Thomson (1996) model suffers from parameter instability and bias, which also makes it problematic for use in long-term projections. Some authors argue that the EMH is unrealistic, but a more complete discussion of actuarial models and EMH is beyond the scope of this paper.

1.4   Later models include non-linear effects through the use of autoregressive conditional heteroscedasticity (ARCH) models, which were introduced by Engle (1982). Such models include the inflation model of Wilkie (1995) and Hua (unpublished). Whitten & Thomas (1999) extend Wilkie's (1995) UK inflation model with further analysis using ARCH and threshold autoregressive (TAR) models. Harris (1994) defines an exponential regressive conditional heteroscedasticity (ERCH) model for Australian data.

1.5   Harris (1996) fits a Markov switching model to quarterly share-price returns and inflation. Markov switching models form another class of non-linear models and were first introduced by Goldfeld & Quandt (1973). They were popularised by the pioneering work of Hamilton (1989, 1990), who describes the likelihood function, regime inferences and an efficient estimation technique for fitting such models.

1.6   Krolzig (1997) develops a comprehensive framework for Markov switching vector autoregressions in which switches between the various components of the vector are perfectly correlated. In Harris (1999), an alternative switching vector autoregression framework is developed, and a vector switching model is estimated for Australian data. However, a vector switching model is not useful if switches between parameter values relating to the individual components of the vector of variables can occur at different points in time. In this case, neither framework provides a useful approach for jointly modelling the variables because parameter values of one variable may switch from one state to another without simultaneous switches in other variables.

1.7   This paper generalises the Markov switching framework by allowing the parameter values of individual series to switch at different times, while allowing for the joint modelling of the variables and state switching. It provides a new and

parsimonious framework that allows individual variables to switch from one state to another without all variables switching at the same time. The model is shown to provide a reasonable description of South African data. The framework presented also allows for easy application of normative assumptions (see Thomson, 2006) while retaining those descriptive aspects of the model that are still believed to be relevant to the future.

1.8   The model presented in this paper was initially presented to the Actuarial Society of South Africa in 1999 and based on data from 1960Q1 to 1998Q4. Those parameter estimates are presented together with updated parameters based on data from the period 1960Q1 to 2006Q2.

2.   VARIABLES MODELLED AND TIME INTERVALS

This section considers some of the data requirements for a stochastic asset–liability model. Many of the issues have been extensively covered by Thomson (1996), so this section focuses mainly on aspects where that approach has been modified or extended.

2.1   VARIABLES MODELLED
2.1.1   As explained by Thomson (1996:768),
"sufficient variables should be modelled to enable the assets and liabilities of the financial institution to be simulated in such a way as to facilitate decision making."

Uncertainty in the liabilities may be due to a large number of random elements. For example, for a defined-benefit pension scheme, wage inflation, price inflation and demographic effects are uncertain. However, since the demographic effects "have a lesser effect on the finances of the scheme and because they are not as strongly correlated with the variables used for simulating asset cash flows," (Thomson, 1996:770), their inclusion increases the dimension of the model unnecessarily. By marginalising the distribution of the liabilities with respect to the demographic effects, we reduce the dimension of the model with only a small loss of information.
2.1.2   The variables specified in this paper facilitate a market-based approach to the valuation of liabilities. Section 9.1.1 of PGN 201 of the Actuarial Society of South Africa[1] states that "the basis used to value the assets must be consistent with that used to value the liabilities …" If, instead of using discounted cash-flow techniques, the market value of liabilities is used to measure the liabilities and assets are taken at market value, it is not necessary to model dividend yields and dividend growth rates.
2.1.3   In an attempt to minimise the dimension of the model and to simplify the analysis, this paper considers a model using only the following four variables:
— the inflation rate;
— the zero-year nominal yield;
— the 20-year nominal par yield; and

1   PGN 201   Retirement funds – actuarial valuation reports and related topics. Actuarial Society of South Africa, 2003

— the total return on equities.

Maitland (2002) shows how to construct a full arbitrage-free yield curve given a model of the zero- and 20-year nominal par yields. Property, wage inflation and offshore asset classes have been excluded as well as inflation-linked yields, although the latter can be inferred from the inflation rate and the nominal yield curve derived from the model. Hence, the variables in this model represent only a subset of the variables required for a comprehensive asset–liability modelling exercise. However, it is believed that the model presented can form the basis for a more comprehensive model of the assets and liabilities of a financial institution.

2.2   TIME INTERVALS

2.2.1   The intended purpose of a stochastic model largely dictates the minimum time interval between forecasts. The Thomson (1996:772) model is developed for annual liability cash flows produced from demographic models based on annual age intervals and for comparison with revenue accounts prepared on an annual basis. Wilkie (1995) and the Finnish Group (Ranne, unpublished) also use annual intervals although Wilkie (*op. cit.*) presents some results for quarterly and monthly intervals as well.

2.2.2   Sherris *et al* (1999:238) consider annual cash flow projections to be a crude approximation to the timing of cash flows and hence prefer a quarterly model. For resilience reserving, capital adequacy and solvency testing, an annual model will tend to understate insolvency probabilities for two reasons. First, solvency can only be assessed annually so that insolvency in the interim will not be detected if the fund has recovered by the following assessment. Secondly, since temporal aggregation tends to reduce excess kurtosis, an annual model may not capture large fluctuations such as equity-market crashes and interest-rate hikes that occur within the year (see Harris 1994:36–8).

2.2.3   Thomson (1996:772) states that

"the use of quarterly data in the development of the model tends to accentuate the short-term relationships at the expense of longer-term relationships."

However, this is not inevitable if the model structure and span of the data allow for longer-term relationships. In the context of the testing of the stationarity of dividend yields, Wilkie (1995:825–6) points out that, even with a large number of frequently sampled observations, a stationary process with high autocorrelation may appear to be non-stationary if the observation period (span) is too short. However, the problem in this context is that the span of the data is too short, not that the sampling frequency of the data is too high. The point Wilkie makes is that an increase in the number of observations by more frequent sampling leads only to a marginal increase in power of unit-root tests, whereas an increase in the span of the data significantly increases the power of these tests (see Perron, 1991). Nonetheless, from this perspective, a model developed using quarterly data from 1960 onwards should be no worse than one developed using annual data over the same period.

2.2.4   As Thomson (1996:772) points out, a quarterly model can be used for comparison with investment performance results, which are often reviewed quarterly.

Furthermore, for some defined-contribution funds, interim bonuses are declared for the following quarter based on investment returns to date and expected returns for the remainder of the year. The fund rules may not allow negative bonuses to be declared so that an investment reserve is required to cover shortfalls. In such cases a quarterly model is required to estimate an appropriate investment reserve level and assess the effects of various bonus strategies. Another advantage of using quarterly intervals over annual intervals is that more data points lead to better parameter estimates. Also, annual figures can be derived from a quarterly model but quarterly figures cannot be derived from an annual model. Consequently, a quarterly time interval is preferred to an annual time interval and so quarterly data are used in this paper.

2.2.5   Possible complications from the use of quarterly intervals instead of annual intervals are that quarterly data may exhibit relatively high kurtosis and may contain seasonal effects. As discussed above, for some applications it is important to capture high kurtosis in the data and so this should be modelled.

2.2.6   Financial series do not usually exhibit seasonal effects but such effects are likely in economic series such as the consumer price index (CPI). This seasonality may be caused by the use of interim price estimates for certain index constituents when actual prices are only available at the end of each year.

2.2.7   Since seasonal effects are of little interest in the current context, the modelling of seasonality requires unnecessary additional parameters and model complexity. Hence, quarter-on-quarter forces of inflation have been seasonally adjusted using the X-12-ARIMA method developed by the U.S. Census Bureau.[2] The model used for the X-12-ARIMA seasonal adjustment is an ARIMA(1,0,1)×(1,0,0)4 model. The seasonally adjusted and annualised force of inflation is the inflation series modelled in this paper.

2.3   TRANSFORMATIONS
2.3.1   The main purpose of transforming data is to enable the use of a simple model form rather than a more complicated one in the original data. The overriding consideration in the choice of transformation is that of linearity. If a non-linear model can be expressed, by suitable transformation of the variables, in linear form, it is said to be intrinsically linear (see Draper & Smith, 1981:222).

2.3.2   Even if relations between variables turn out to be non-linear, linear modelling frameworks such as the transfer-function model with autoregressive integrated moving average terms (ARIMAX) and the vector autoregressive moving average (VARMA) model classes provide a simple and parsimonious framework for model development and should be considered before moving to non-linear modelling frameworks. For this reason, functions that admit a linear relation between variables are highly desirable. For example, rates of growth are multiplicative whereas forces are additive and hence more linear. In this context, a logarithmic transformation of the rates

---

2   For further details see: X-12-ARIMA Reference Manual. Statistical Research Division, U.S. Census Bureau, Washington D.C., 2000

is appropriate. In addition, a logarithmic transformation changes the range to $(-\infty, \infty)$, allowing certain variables to be modelled with standard normal distributions. A broad literature review and discussion of the most appropriate functions to use for the variables modelled in the Thomson Model can be found in Thomson (1996:773–7).

2.4 TIME SERIES MODELLED

2.4.1 The quarter-on-quarter force-of-inflation series is constructed by taking the natural logarithm of the ratio of the all-items CPI at quarterly intervals. This series is seasonally adjusted (as discussed above) and then annualised to give the seasonally adjusted and annualised force of inflation series, $INFL_t$.

2.4.2 Figure 1 shows $INFL_t$ together with the year-on-year force of inflation, $INFL\text{-}YY_t$ at quarterly intervals from 1960Q1 to 2006Q2. It should be noted that the modelling of the year-on-year force of inflation at quarterly intervals is problematic in that it has a tendency to increase the autocorrelation between successive periods and to obscure temporal dependence in the series. This comparison is shown only for illustrative purposes and for comparison with the more familiar year-on-year figures often published.

2.4.3 For long-term interest-bearing securities, the use of an average annual force of interest is well motivated by Thomson (1996:776). The variable modelled is:

$$LINT_t = 2*\ln(1+JAYC20_t/200);$$

where $JAYC20_t$ is the JSE-Actuaries 20-year nominal bond yield, convertible half-yearly, at time t, as quoted under code $JAYC20$ by INET.[3]

Figure 1: $INFL_t$ ("SAdj-X12 INFL) and $INFL\text{-}YY_t$ ("Y-on-Y INFL")

---
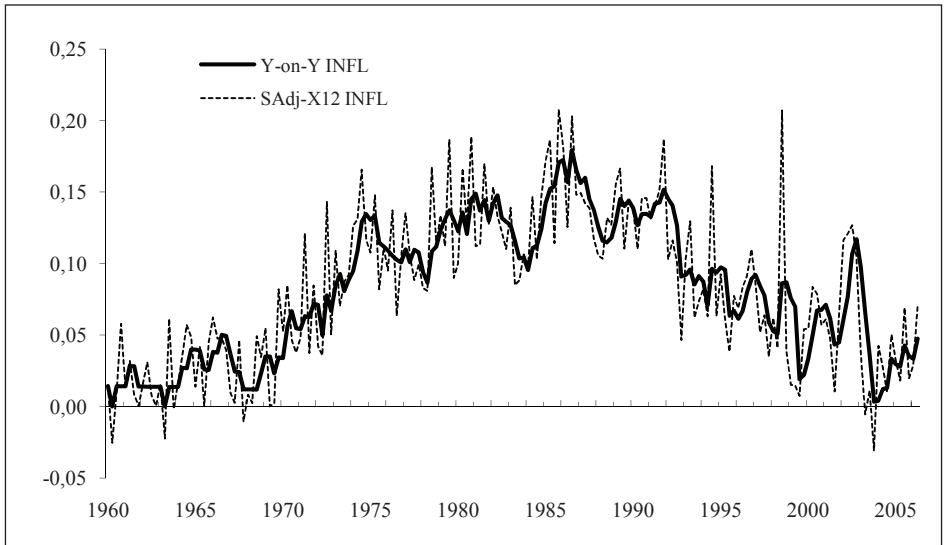
3   INET Graphics and Database Service. Intelligent Network (Pty) Ltd., 2006

2.4.4   For money-market instruments, Thomson (1996:776) models the annual force of return on the Alexander Forbes money-market index, as quoted under code *GMC*1 by INET.[4] However, that index is constructed from the average monthly return on a portfolio consisting of 3-month negotiable certificates of deposit with 1, 2 and 3 months to maturity. All information contained in *GMC*1 at time *t* is available before that time, so the value of *GMC*1 at time *t* does not belong in the information set at that time: unlike the yield curve, *GMC*1 does not reflect rates available at the start of the period.

2.4.5   Ideally, the force of interest on 3-month Treasury bills at time *t* should be used to reflect the risk-free rate available at that time for the quarterly period $(t, t + 1)$. However, since there is only a short history of Treasury-bill rates, the zero-year nominal yield, as quoted under code *JAYC*00 by INET[5] is used as a proxy. The short-term interest rate modelled is defined as:

$$SINT_t = 2\ln(1 + JAYC00_t/200) \text{ for } t \text{ in } \{1986Q1,\ldots,2006Q2\}.$$

2.4.6   Since *JAYC*00 is available only from 1986 onwards, the annualised force of change in *GMC*1 from time *t* to time $t + 1$ is used as a proxy for $JAYC00_t$, as follows:

$$SINT_t = 4\ln(GMC1_{t+1}/GMC1_t) \text{ for } t \text{ in } \{1960Q1,\ldots,1985Q4\}.$$

2.4.7   Maitland (2002) shows that, in constructing a full arbitrage-free yield curve, the zero- and 20-year nominal par yields are the best yields to model to minimise the forecast error of the full yield curve.

2.4.8   For equities, it is preferable to model the excess equity return above the return on a risk-free asset (as modelled by the risk-free rate of interest) rather than the nominal equity return because risk-averse investors are typically interested in the additional returns they receive for taking on risk.

2.4.9   It could be argued that the real yield on a three-month CPI-linked bond is the appropriate risk-free hurdle rate for investors interested in accumulating real wealth. However, since such an instrument does not exist in South Africa, this approach is not particularly helpful. Arguably, for such investors, the three-month Treasury bill is the best proxy we have for a risk-free investment over the short term.

2.4.10   It could also be argued that, for an investor with longer-term liabilities, the return on an immunising portfolio of longer-dated bonds is the relevant hurdle rate (see Maitland (2001) for the immunisation framework that is mathematically optimal). This is indeed true. However, for the purposes of simplicity and without further knowledge of the segmentation of investor objectives, $SINT_t$ is used as a proxy for the three-month Treasury-bill rate, and this is assumed to be the risk-free hurdle rate for each quarterly period.

2.4.11   The total return index for equities, $EQTRI_t$, is taken as the FTSE–JSE all-share index, as quoted under code *J*203*TRI* by the JSE.[6] This is available from

---

4   supra, 2006

5   supra, 2006

6   FTSE / JSE End of Day Indices Dissemination User Manual, JSE Information Products Sales Division, 30 October 2005

30 September 1995 onwards. Before this, a total-return index for equities is constructed assuming dividends are uniformly distributed over the calendar year and using the JSE-Actuaries all-share index and the dividend yield per cent on that index, as quoted under codes *CI*01 and *ADY* respectively by INET.[7]

2.4.12   The total, annualised, quarterly force of return on equities in excess of the risk-free rate, $XSEQ_t$, is then constructed as follows:

$$XSEQ_t = 4*\ln(EQTRI_t/EQTRI_{t-1}) - SINT_{t-1}$$

2.4.13   The variables $SINT_t$, $LINT_t$ and $XSEQ_t$ are shown in Figures 2 to 4. The analysis in this paper uses data over the period from 1960Q1 to 2006Q2.

3.   UNIT-ROOT TESTS

3.1   In the building of a multivariate time-series model, the purpose of developing a univariate model for each of the variables is to guide subsequent multivariate modelling. How best to proceed hinges on knowing whether the individual series are stationary or non-stationary. Conventional time-series estimation techniques based on classical assumptions about the distribution of the error terms can lead to incorrect inferences if the series are non-stationary. For example, if classical ordinary least squares are used to estimate the relationship between two non-stationary variables, each containing a unit root, standard test statistics produce misleading inferences. This is known as the spurious regression problem (see Granger & Newbold, 1974).

Figure 2: $SINT_t$

---
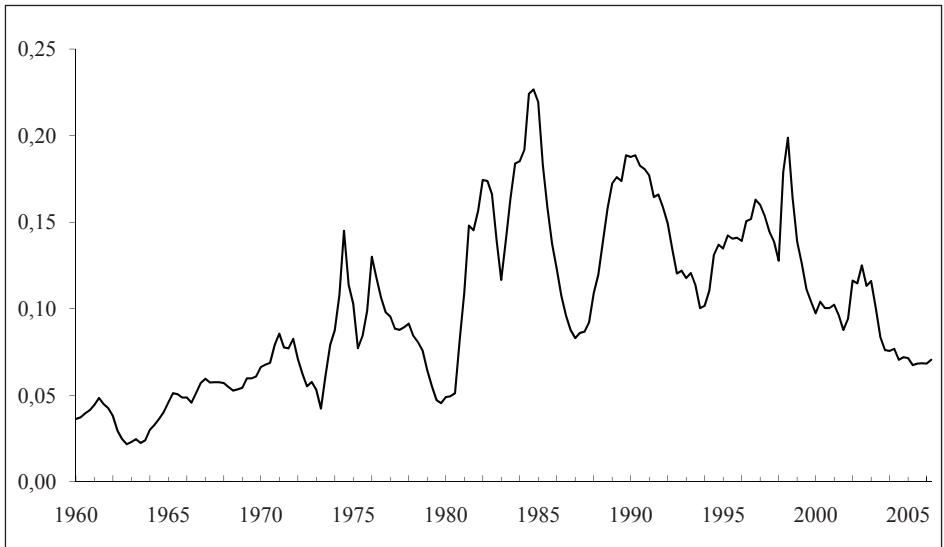
7   supra, 1998

3.2   Standard unit-root tests generally test the null hypothesis of a unit root against the one-sided alternative of no unit root (see, e.g., Hamilton, 1994:Chapter 17). The results of some standard unit-root tests are shown in Table 1, viz. the augmented Dickey–Fuller
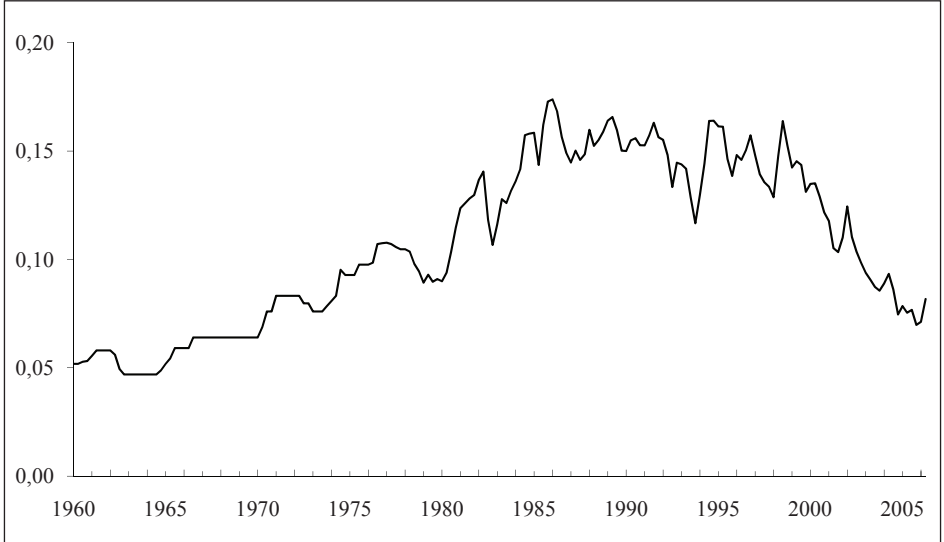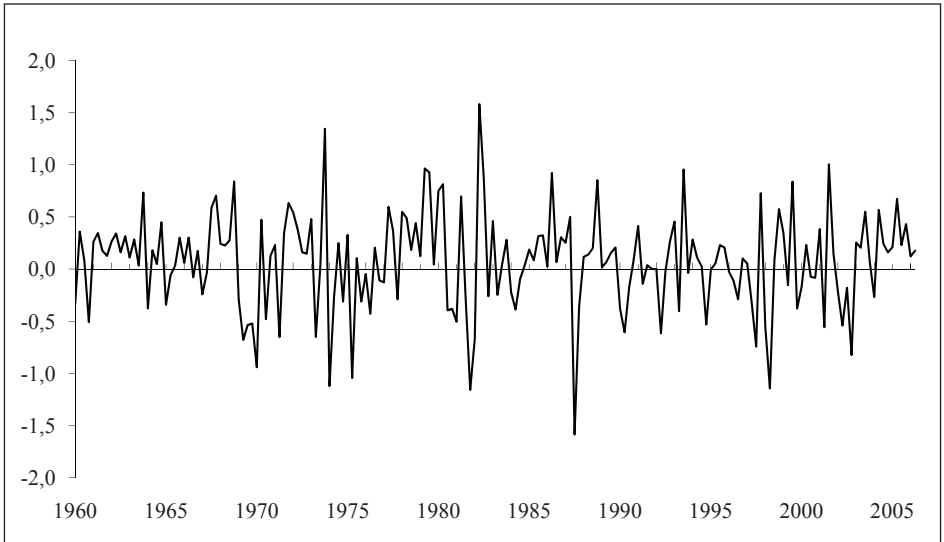
Figure 3: $LINT_t$

Figure 4: $XSEQ_t$

test (ADF), the Phillips–Perron test (PP) and the Kwiatkowski–Phillips–Schmidt–Shin test. For further details, see Maitland (unpublished b), Dickey & Fuller (1979, 1981), Phillips & Perron (1988) and Kwiatkowski, Phillips, Schmidt & Shin (1992). These tests all include an intercept in the test equation and test for a unit root in the level series.

Table 1: Unit-root tests

Variable	ADF	PP	KPSS
	t-statistic (*p*-value)	*t*-statistic (*p*-value)	LM-statistic [5% critical value]
$INFL_t$	−2,982 (0,039)	−6,017 (0,000)	0.504 [0,463]
$SINT_t$	−2,857 (0,053)	−2,441 (0,132)	0.891 [0,463]
$LINT_t$	−1,534 (0,514)	−1,501 (0,531)	1.067 [0,463]
$XSEQ_t$	−13,202 (0,000)	−13,209 (0,000)	0.058 [0,463]

3.3   For both the ADF and PP tests, the null hypothesis is that the series is non-stationary, and only if the series is sufficiently stationary is this assumption rejected. These results suggest that the null hypothesis cannot be rejected at the 5% level for $SINT_t$ and $LINT_t$. The results also suggest that $INFL_t$ and $XSEQ_t$ do not contain a unit root since the null hypothesis is rejected at the 5% level for these two variables.

3.4   The KPSS test is a Lagrange-multiplier (LM) test that evaluates the null hypothesis that the series is stationary against the alternative that it is non-stationary. As a result, the KPSS test reverses the usual burden of proof. An LM-statistic that is greater than the 5% critical value rejects the null hypothesis at the 5% level. The KPSS result supports the findings of the ADF and PP tests that $XSEQ_t$ does not contain a unit root. The results also suggest that the null hypothesis of stationarity can be rejected at the 5% level for $INFL_t$, $SINT_t$ and $LINT_t$. This supports the finding of the ADF and PP tests that $SINT_t$ and $LINT_t$ each contain a unit root but contradicts the earlier result for $INFL_t$.

3.5   The ADF and PP results for $INFL_t$ in Table 1 above also contrast with the standard unit-root test results shown in Maitland (unpublished b), which used annual data for the force of inflation over the period 1960 to 1993. A rerun of the ADF and PP tests for $INFL_t$ for the sub-period 1960Q1 to 1993Q4 shows that the null of a unit root cannot be rejected at the 5% level for the ADF test statistic but that it can be rejected for the PP statistic. Such mixed results are symptomatic of non-linear effects in the data.

3.6   The implication of a unit root in a time series is that shocks to the system are permanent, trends are stochastic and forecast variances increase linearly as the lead time of the forecast increases. Hence, whether or not a variable contains a unit root is critical for projection purposes.

3.7   The above results using data from 1960Q1 to 2006Q2 might suggest a multivariate model with $INFL_t$ modelled as a stationary variable and $SINT_t$ and $LINT_t$ as non-stationary variables. However, such a mixed model would not make sense, particularly as $LINT_t$ reflects the market's expectation of future inflation, the real rate of interest and an inflation risk premium. If inflation is stationary, a non-stationary $LINT_t$ would imply a non-stationary inflation risk premium and real interest rate. However, it is unreasonable to assume that real interest rates can wander off to any level, as implied by a forecast variance that increases linearly with time.

3.8   As discussed in section 2.2, an increase in the span of the data significantly increases the power of unit-root tests (see Perron, 1991 and Wilkie, 1995:825–6). Homer & Sylla (2005) show that for around 4000 years interest rates have remained around 5% a year (in non-inflationary times). This suggests that with an increased span of data, standard unit-root tests on the above series will most likely indicate that they are stationary. However, even with the short span of data available, certain tests lead to more reasonable models than suggested by the above unit-root test results.

3.9   Perron (1989) shows that standard unit-root tests that do not allow for the presence of a structural break have little power against the alternate of no unit root when the underlying series has a structural break but no unit root. The power of these tests decreases as the magnitude of the intervention variables increases.

3.10   Using Perron's (1989) framework, Maitland (unpublished b) shows that the null hypothesis of a unit root can be rejected once the possibility of a structural break is considered. However, Perron's framework does not entertain the possibility of multiple structural breaks at unknown times. For this reason, unit-root tests that allow for the possibility of a deterministic structural break are not considered further in this paper. Instead, a modelling framework allowing for multiple structural breaks at unknown times is presented in the following section.

4.   UNIVARIATE MARKOV SWITCHING MODELS
4.1   INTRODUCTION
4.1.1   Economic and financial time series can exhibit dramatic breaks in their behaviour, associated with events such as oil-price shocks, changes in government policy, financial crises and shifts in investor expectations. If the behaviour for particular periods can be adequately described by autoregressive (AR(p)) models of the form:

$$y_t = c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t \tag{1}$$

with $\varepsilon_t \sim N(0,\sigma^2)$, then it would be reasonable to allow the parameters $c, \phi_1, \ldots, \phi_p, \sigma$ of this model to change to accommodate such breaks. The encompassing model could then be described as:

$$y_t = c_{S_t} + \phi_{1,S_t} y_{t-1} + \ldots + \phi_{p,S_t} y_{t-p} + \varepsilon_t \tag{2}$$

with $\varepsilon_t \sim N(0, \sigma_{St}^2)$, where s_t denotes the regime or state of the process at time t.

4.1.2   Since the determinants of these changes may be unobservable (as, for example, with a shift in investor expectations), or because one may simply not wish to include such determinants as factors in the model (the causes of financial crises are varied and inflation is not only influenced by oil-price shocks), it is preferable to consider a probabilistic model to describe the occurrence of such breaks that give rise to changes in the parameters $c, \phi_1, \ldots, \phi_p, \sigma$.

4.1.3   The simplest specification is that s_t is the realisation of a Markov chain with the probability of a switch from state i to state j ($i, j = 1, 2, \ldots, M$) being:

$$\Pr\left(s_t = j \mid s_{t-1} = i\right) = p_{ij} \qquad (3)$$

where $p_{i1} + p_{i2} + \ldots + p_{iM} = 1$ for all $i \in \{1, \ldots, M\}$. This assumes that the probability of a change in state or regime depends on the past only through the value of the most recent regime. The regimes are not observed directly but can be inferred through the observed behaviour of y_t.

4.1.4   The specification in equations (2) and (3) is non-linear and is referred to as a Markov-switching (MS) model. Markov-switching regressions were introduced by Goldfeld & Quandt (1973), and the likelihood function was first correctly calculated by Cosslett & Lee (1985). The pioneering work of Hamilton (1989, 1990) describes the likelihood function, regime inferences and an efficient estimation technique for fitting such models. Krolzig (1997:Chapter 7) provides a useful discussion on model selection and model checking procedures for MS models.

4.1.5   A model specification search is undertaken using the Schwartz information criterion (SC) and a likelihood ratio (LR) test. Given the number of regimes, standard asymptotic distribution theory holds for the SC concerning the number of auto-regressive parameters and heteroscedasticity. (SC provides the most parsimonious model specification amongst the widely used Akaiki information criterion (AIC), Hannan–Quinn (HQ) criterion and SC.) See Hamilton (1994) for details.

4.1.6   The LR test concerns the appropriate number of states in equation (2), and follows a non-standard distribution. Equivalence in all regimes of the parameters $c, \phi_1, \ldots, \phi_p, \sigma$ of equation (2) implies that the Markov chain parameters p_{ij} are not identified under the null hypothesis of a single state ($M = 1$).

4.1.7   As discussed in Garcia (1998), testing for the number of states in a regime switching framework is complex. Given some $M \geq 2$, the problem is that under any number of regimes smaller than M some transition probability ('nuisance') parameters of the unrestricted model may take any value and are hence unidentified. The result is that the LR test fails to have a standard chi-square distribution with number of degrees of freedom equal to the number of restrictions imposed.

4.1.8   To overcome these complexities, the bounded likelihood ratio test proposed by Davies (1977) and recommended by Krolzig (1997) is used to test the null hypothesis of a single state $M = 1$ (i.e. the 'linear' model) against the alternative of two or more states ($M \geq 2$). This circumvents the problem of estimating nuisance parameters under the alternative hypothesis and derives instead an upper bound for the significance level of the LR test:

$$\Pr\left(LR > x\right) \leq \Pr\left(\chi_q^2 > x\right) + x^{q/2} \exp\left(-x/2\right) \cdot 2^{1-q/2} \cdot \left[\Gamma\left(q/2\right)\right]^{-1}; \qquad (4)$$

where $\Gamma(\bullet)$ is the standard gamma function and q is the number of nuisance parameters. (Note that for $M = 2$, $q = 2$.)

4.2 $INFL_t$

4.2.1   Maitland (unpublished a) shows that the Thomson (1996) model suffers from parameter instability. In particular, he shows that the autoregressive parameter for inflation is much lower than that suggested by the Thomson model and the means quite different when estimated from the sub-periods 1960–1975 and 1976–1993.

4.2.2   Maitland (unpublished b) also shows that it is this parameter instability that gives rise to the apparent unit root in the inflation series, and that the null hypothesis of a unit root can be rejected once the possibility of a single structural break is considered.

4.2.3   This assumption of a single structural break is unsatisfactory as a probability law that could have generated the inflation series. Furthermore, such features are not desirable for projection purposes and are not likely given the current framework of inflation targeting in South Africa.[8] Instead, an MS model, which allows for multiple structural breaks at unknown times, is a more appropriate framework for modelling $INFL_t$.

4.2.4   A number of first-order MS models with $M=2$ states have been estimated for $INFL_t$, allowing for switching in any combination of the intercept term (I), the autoregressive terms (A) and the variance of the residuals (H). The results are shown in Table A.1 of Appendix A. The null of a single state model ($M=1$) is rejected in favour of $M=2$, the LR statistic being highly significant, even after applying Davies's (1977) correction. $M=3$ is rejected. In terms of the SC, the best model is the autoregressive model with switching only in the intercept term:

$$INFL_t = c_{S_t} + \phi_1 INFL_{t-1} + \varepsilon_t, \tag{5}$$

with $\varepsilon_t \sim N(0, \sigma^2 = 0{,}033^2)$, $c_1 = 0{,}0299$ (0,005), $c_2 = 0{,}0944$ (0,013) and $\phi_1 = 0{,}24$ (0,097). Standard errors are shown in parentheses. $s_t = 1$ corresponds to a low-mean regime with a mean force of inflation, $c_1/(1-\phi_1)$, of about 4% a year, while $s_t = 2$ corresponds to a high-mean regime with a mean force of inflation of about 12% a year. The probability of remaining in state 1 given that the process is already in state 1 is $p_{11} = 0{,}968$, while the probability of remaining in state 2 given that the process is already in state 2 is $p_{22} = 0{,}970$. The autoregressive parameter and the variance remain stable across regimes. The ergodic (unconditional) probabilities or stable-state probabilities (see, e.g., Hamilton, 1994:Chapter 22) are both 0,5 for the low- and high-mean states, while the expected durations for the low- and high-mean states are both about 8 years.

4.2.5   When one compares this MS model with the linear AR(1) model shown in Table A.1 of Appendix A, the problem of parameter bias becomes apparent. It can be seen that the autoregressive parameter $\phi_1 = 0{,}67$ for the AR(1) model is much higher than $\phi_1 = 0{,}24$ for the MS model. The bias is caused by the changing level of the series, as discussed by Maitland (unpublished b).

---

8   For details see: A new monetary policy framework. Statement of the Monetary Policy Committee of South Africa, T.T. Mboweni, 6 April 2000.

4.2.6   The estimated probability (conditional on all the data) that the regime was in the high-inflation regime each quarter is shown in Figure 5. This should be compared with Figure 1 where it can be seen that a high-inflation regime corresponds roughly to the periods 1973–1994, 1998:3 and the year 2002.

4.2.7   The year 1973 corresponds to the first oil-price shock, which led to a twenty-year period of entrenched inflation, ending with the end of the apartheid era and the dismantling of many trade barriers, which led to increased international trade, decreased market power of domestic companies and downward pressure on real wages (see Aron & Muellbauer, 2000).

4.2.8   The third quarter of 1998 corresponds to the Russian debt crisis, and 2002 follows the dramatic fall in the rand in 2001 following the Zimbabwe crisis and fears of contagion in the region.

4.2.9   With the introduction of inflation targeting in 2000,[9] periods of persistent high inflation such as were experienced from 1973–1994 are arguably less likely to occur in future. In using the MS model for projection purposes, the user is likely to lower the value of p_{22}, the probability of remaining in state 2.

4.2.10   However, the possibility of future inflation shocks (resulting from oil shocks, currency shocks, political crises, financial crises, wage pressures etc.) is not removed with the introduction of inflation targeting; nor can we rule out the possibility of a weak monetary policy regime at some time in the future. By simply adjusting the transition probabilities, the user is able to mimic stochastic projections under such scenarios.

4.2.11   Although certain aspects of the past are unlikely to repeat themselves in the future, more stable aspects might still prove useful. For example, the user is able to retain the estimates for c_1, c_2, ϕ_1 and σ, unless more plausible values can be justified.

4.2.12   The values of ϕ_1 and σ are stable across both regimes and are therefore estimated from the full sample of data, so our confidence in these estimates should be higher. While inflation expectations from the real and nominal yield curves, and the inflation target band of between 3% and 6% a year, might suggest slightly different

Figure 5: Probability that $INFL_t$ was in the high-inflation regime

---
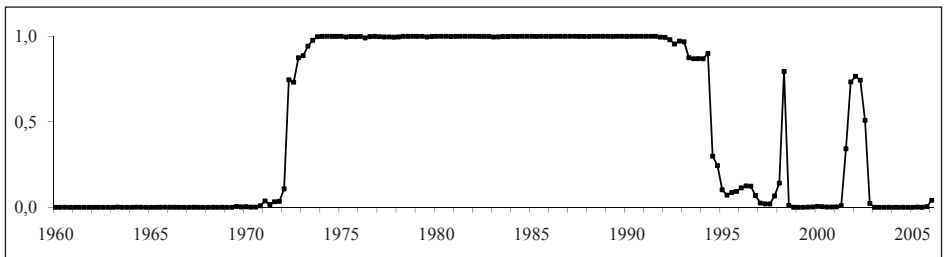
9   For details see Mboweni, supra.

values for c_1 and σ, the value of ϕ_1 can be retained, unless the user has reason to justify how inflation targeting might alter this dynamic.

4.2.13   In contrast to the MS model for $INFL_t$, none of the parameters for $INFL_t$ estimated in the Thomson (1996) model are useful for projection purposes. As discussed by Maitland (unpublished a), Thomson's $INFL_t$ autoregressive parameter is biased; the forecast mean tends to the arbitrary level of 9.5% a year, and clearly depends on the period of data used to estimate the parameters; and the forecast variance is inflated because of regime switching in the underlying series.

4.2.14   The MS model appears to be relatively stable when estimated over the sub-periods 1960:1–1998:4 and 1970:1–2006:2. In both cases, the problem of parameter bias is again apparent in the corresponding linear models (Tables A.1.2–3 of Appendix A).

4.2.15   For the MSIH(2)–AR(1) model estimated over the sub-period 1960:1–1998:4, the parameter estimate for p_{22} is equal to one (see Table A.2 in Appendix A). This implies a permanent switch to the high-mean inflation regime, which is clearly unrealistic. The problem is that that span of data does not include periods during which a switch from the high-mean regime to the low-mean regime occurred (or at least not with sufficient clarity to distinguish this given the higher volatility of the high-mean regime in this model). However, we know inflation has come down since then, and even if it had not, this is always a possibility. Hence, the parameter estimate $p_{22} = 1$ from this subset model, while clearly a reasonable estimate given the data from that sub-period, is clearly not appropriate for forecasting. This illustrates the importance of applying judgement when setting parameters for projection purposes.

4.3   *SINT_t*

4.3.1   The results of fitting various first-order MS models to $SINT_t$ are shown in Table A.4 of Appendix A. The null of a single-state model ($M = 1$) is rejected in favour of $M = 2$, the LR statistic being highly significant, even after applying Davies's (1977) correction. $M = 3$ is rejected. In terms of the SC, the best model is the autoregressive model with switching in both the intercept term and the residual variance:

$$SINT_t = c_{S_t} + \phi_1 SINT_{t-1} + \varepsilon_t; \tag{6}$$

where $\varepsilon_t \sim N(0, \sigma_{St}^2)$, $c_1 = 0{,}0072$ (0,002), $c_2 = 0{,}0234$ (0,006) and $\phi_1 = 0{,}866$ (0,037). Standard errors are shown in parentheses. $s_t = 1$ corresponds to a low-mean regime with a mean force of interest, $c_1/(1-\phi_1)$, of about 5,5% a year, while $s_t = 2$ corresponds to a high-mean regime with a mean force of interest of about 17,5% a year. $\sigma_1 = 0{,}0051$ and $\sigma_2 = 0{,}0155$, so the high-short-term-interest regime is much more volatile, its volatility being three times that of the low-short-term-interest regime. The autoregressive parameter is stable across regimes.

4.3.2   The probability of remaining in state 1 given that the process is already in state 1 is $p_{11} = 0{,}947$ while the probability of remaining in state 2 given that the process is already in state 2 is $p_{22} = 0{,}920$. The ergodic probabilities or stable-state probabilities are 0,601 for the low-mean state and 0,399 for the high mean state. The expected duration for the low-mean state is about 5 years while that for the high-mean state is about 3 years. (Because of the probabilistic nature of the MS process, the series

may remain in either state for as little as one quarter or much longer than the expected duration. Such asymmetry is not well captured by linear models.)

4.3.3   When one compares this MS model with the linear AR(1) model shown in Table A.4 of Appendix A, the problem of parameter bias in the linear model again becomes apparent. The autoregressive parameter $\phi_1 = 0{,}963$ for the AR(1) model is much higher than $\phi_1 = 0{,}866$ for the MS model, the bias again being induced by the changing level of the series. The estimated probability (conditional on all the data) that the regime was in the high-inflation regime each quarter is shown in Figure 6.

4.3.4   The MSIH model for $SINT_t$ appears to be very stable when estimated over the sub-periods 1960:1–1998:4 and 1970:1–2006:2. In both cases, the problem of parameter bias in the autoregressive parameter is again apparent in the linear models.

4.3.5   The residuals from the AR(1) model for $SINT_t$ exhibit a very high kurtosis of 5,4 and a Jarque–Bera statistic of 65,3, suggesting that the null hypothesis that the residuals are normally distributed can be rejected at the 99,999% level. Also, the Ljung–Box test statistics on the squared residuals indicate significant serial correlation structure in the volatility. This suggests fitting a GARCH model (see Engle (1982) and Bollerslev (1986) for details) to $SINT_t$.

4.3.6   The GARCH(1,1) model successfully removes all serial correlation in the squared residuals. However, the model suffers from extreme upward bias in the autoregressive parameter ϕ_1, which is estimated to be 0,999. The kurtosis of the residuals increases to 7,1 and the Jarque–Bera statistic of 182,8 indicates an even more severe departure from normality than with the residuals from the AR(1) model. In contrast, the assumption of normality in the residuals from the MS model in equation (6) cannot be rejected.

4.3.7   The log-likelihood for the GARCH(1,1) model is 585,22, which is considerably less than the log-likelihood of 604,65 for the MS model in equation (6). However, the standard likelihood-ratio test cannot be used to compare these two models because they are not nested.

4.4 $LINT_t$

4.4.1   The results of fitting various first-order MS models to $LINT_t$ for the full period 1960:1–2006:2 as well as the sub-periods 1960:1–1998:4 and 1970:1–2006:2 are

Figure 6: Probability that $SINT_t$ was in the high-interest regime
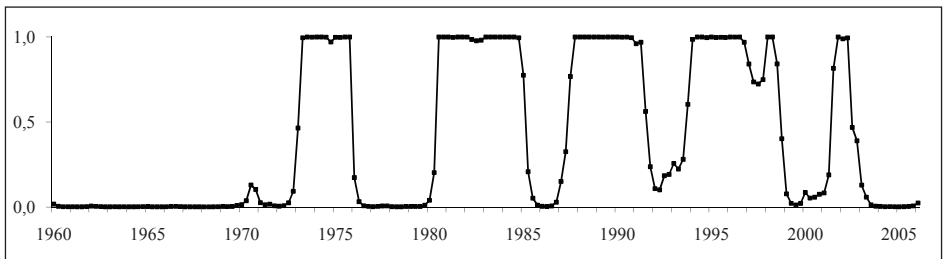
shown in Tables A.7–9 of Appendix A. For both periods including the 1960s, the MSIH model is numerically unstable. Furthermore, the autoregressive parameter is very close to one when data from the 1960s is included in the estimation but drops when this period is excluded.

4.4.2   Examination of the data in Figure 3 reveals the cause of the instability. The earlier data are characterised by long stretches during which the long-bond yield remains unchanged, interspersed with occasional jumps. For example, from December 1962 to September 1964, $LINT_t$ is constant at 0,0469, while from September 1966 to March 1970 it remains constant at 0,064. The series was constructed by Dr J Greener "based on the coupon of the bonds issued in the primary market" using long-dated government bonds issued by the South African Reserve Bank for government. Since there were very few issues in the 1960s, the yield remained constant for long stretches at a time. These bonds were simply bought and held, largely by life offices, pension funds and particularly the Government Employees Pension Fund, which were all subject to prudential regulations forcing them to hold large volumes of government bonds. An active secondary market for these bonds did not develop until the early 1980s.

4.4.3   Such dynamics are not characteristic of ARIMAX, VARMA or MS stochastic processes and bias the autoregressive parameter ϕ_1 towards one. They are unlikely to be repeated in future if market forces continue to determine yields, and hence are not useful for projection purposes. Hence, data from the 1960s are excluded for estimation of the $LINT_t$ MS model parameters.

4.4.4   Using data for the period 1970:1–2006:2, the null of a single state model ($M = 1$) is rejected in favour of $M = 2$, the LR statistic being highly significant, even after applying Davies's (1977) correction. $M = 3$ is rejected. In terms of the SC, the best model is the autoregressive model with switching in both the intercept term and the residual variance:

$$LINT_t = c_{S_t} + \phi_1 LINT_{t-1} + \varepsilon_t; \qquad (7)$$

where $\varepsilon_t \sim N(0, \sigma_{St}^2)$, $c_1 = 0{,}0129$ (0.004), $c_2 = 0{,}0214$ (0.006) and $\phi_1 = 0{,}852$ (0.041). Standard errors are shown in parentheses. $s_t = 1$ corresponds to a low-mean regime with a mean force of interest, $c_1/(1-\phi_1)$, of about 8.5% a year, while $s_t=2$ corresponds to a high-mean regime with a mean force of interest of about 14,5% a year. $\sigma_1 = 0{,}0041$ and $\sigma_2 = 0{,}0083$, so the high-long-term-interest regime is much more volatile, its volatility being more than double that of the low-long-term-interest regime. The autoregressive parameter $\phi_1 = 0{,}852$ is constant across regimes.

4.4.5   The probability of remaining in state 1 given that the process is already in state 1 is $p_{11} = 0{,}973$, while the probability of remaining in state 2 given that the process is already in state 2 is $p_{22} = 0{,}983$. The ergodic or stable-state probabilities are 0,391 for the low-mean state and 0,609 for the high mean state. The expected duration for the low-mean state is about 9 years while that for the high-mean state is about 14 years.

4.4.6   When one compares this MS model with the linear AR(1) model shown in Table A.6 of Appendix A, the problem of parameter bias in the linear model again becomes apparent. The autoregressive parameter $\phi_1 = 0{,}965$ for the AR(1) model is much higher than $\phi_1 = 0{,}852$ for the MS model, the bias again being induced by the

changing level of the series. The estimated probability (conditional on all the data) that the regime was in the high-inflation regime each quarter is shown in Figure 7. (These inferences are based on the parameters estimated from the period 1970:1–2006:2.)

4.4.7   The transition probability estimates for $LINT_t$ are based on very few switches, so the precision of these estimates can be expected to be low. However, with the introduction of inflation targeting in 2000,[10] periods of persistent high inflation such as were experienced from 1973 to 1994 would be expected to occur less often in future, and this would be expected to concentrate 20-year bond yields in the low-mean regime. Hence, in using the MS model for projection purposes, the user is likely to decrease the values of p_{12} and p_{22}, rather than rely on these parameter estimates.

4.4.8   The residuals from the AR(1) model for $LINT_t$ for the period 1970:1–2006:2 exhibit kurtosis of 3,5, a Jarque–Bera statistic of 3 and a *p*-value of 0,216, suggesting that the assumption of normality cannot be rejected. However, the Ljung–Box test statistics on the squared residuals indicate significant serial correlation structure in the volatility. This suggests fitting a GARCH model to $LINT_t$.

4.4.9   The GARCH(1,1) model successfully removes all serial correlation in the squared residuals. The Jarque–Bera statistic of 2,3 for the residuals suggests that the assumption of normality cannot be rejected. The log-likelihood for the GARCH(1,1) model is 521,36, which is slightly lower than the log-likelihood of 524,41 for the MS model in equation (7). However, the GARCH model suffers from upward bias in the autoregressive parameter ϕ_1, and so cannot be recommended.

4.5 $XSEQ_t$

4.5.1   One of the earliest and most enduring models of the behaviour of equity price trajectories is the random-walk model. Such a model implies that equity returns are independent and identically distributed (i.i.d.). The Black–Scholes option-pricing theory extends this model by assuming that returns over any discrete time interval follow a lognormal distribution. These standard models are accommodated by modelling $XSEQ_t$ as an i.i.d. normal random variable, i.e. the AR(0) model.

Figure 7: Probability that $LINT_t$ was in the high-interest regime

---
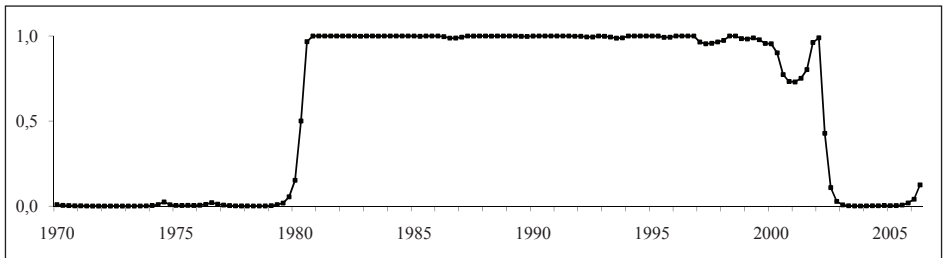
10 For details see Mboweni, supra.

4.5.2   There is now a vast literature suggesting that the standard lognormal model is inadequate. Empirical studies of equity returns provide evidence of time-varying volatility that the standard lognormal model is unable to capture. Engle (1995:xii) states that "[t]he GARCH(1,1) model is [now] the leading generic model for almost all asset classes of returns," and presents a collection of papers on variants of the ARCH model used in finance.

4.5.3   More recently, MS models have been successfully applied to the modelling of equity returns. Harris (1996, 1999) introduced the regime-switching lognormal model for equity returns, Bollen (1998) prices American and European options under this model, and Hardy (2001) successfully applies the approach of Harris (1999) and Bollen (*op. cit.*) to US and Canadian data.

4.5.4   The standard Jarque–Bera test is used to test the null hypothesis that $XSEQ_t$ follows a normal distribution. $XSEQ_t$ has a negative skewness of –0,24 and a kurtosis of 4,1. The Jarque–Bera statistic is 10,410 and has a *p*-value of 0,005, which indicates a severe departure from normality. Superficially, this suggests fitting a model with a fat-tailed residual distribution.

4.5.5   The Ljung–Box test statistics on the squared residuals of the AR(0) model, however, indicate significant serial correlation structure in the volatility. This indicates the need for a model incorporating time-varying volatility rather than simply a model with a fat-tailed residual distribution. For this purpose, the ARCH(1) model turns out to be a better model for $XSEQ_t$, the GARCH(1,1) model being over-parameterised. The log-likelihood for the ARCH(1) model is –118,07.

4.5.6   The results of fitting various MS models to $XSEQ_t$ are shown in Table A.10 of Appendix A. The first-order autoregressive term is not significant, as would be expected under the efficient market hypothesis, and the results from this class of models are not shown. The null of a single-state model ($M = 1$) is rejected in favour of $M = 2$, with a significant LR statistic of 17,16 and a *p*-value of 0,0034 after applying Davies's (1977) correction. $M = 3$ is rejected. In terms of the SC, the best model is the autoregressive model with switching in both the intercept term and the residual variance:

$$XSEQ_t = c_{S_t} + \varepsilon_t ; \qquad (8)$$

where $\varepsilon_t \sim N(0, \sigma_{St}^2)$, $c_1 = -0,0333$ (0,093) and $c_2 = 0,1404$ (0,041). Standard errors are shown in parentheses. $\sigma_1 = 0,620\ 57$ and $\sigma_2 = 0,305\ 81$, so s_t=1 corresponds to a volatile, low-mean-return regime with an effective annual volatility of 31%, while $s_t = 2$ corresponds to a stable, high-mean-return regime with an effective annual volatility of 15%. The model is stable when estimated over the sub-periods 1960:1–1998:4 and 1970:1–2006:2. The log-likelihood for the MS model in equation (8) is –112,9, which is considerably higher than the log-likelihood of –118,0685 for the corresponding ARCH(1) model. However, the standard likelihood ratio test cannot be used to compare these two models because they are not nested.

4.5.7   The probability of remaining in regime 1 given that the process is already in regime 1 is $p_{11} = 0,8259$ while the probability of remaining in regime 2 given that the process is already in regime 2 is $p_{22} = 0,8853$. The ergodic or stable-state probabilities are 0,3972 for the volatile, low-mean-return state and 0,6028 for the stable,

high-mean-return state. The expected duration for the volatile, low-mean-return state is about 6 quarters while that for the high-mean-return state is about 9 quarters. The estimated probability (conditional on all the data) that the regime was in the volatile, low-mean return regime each quarter is shown in Figure 8.

4.5.8   Figure 9 shows the unconditional density function for $XSEQ_t$ and the joint density functions for each of the unobserved regimes, scaled by the probability of being in each of those regimes. Although the distribution of $XSEQ_t$ conditional on each regime is Gaussian, the mixing distribution exhibits higher kurtosis and negative skewness. However, unlike the unconditional distributions for $INFL_t$, $SINT_t$ and $LINT_t$, the unconditional distribution for $XSEQ_t$ is unimodal.

4.5.9   The MSIH(2)–AR(0) model of equation (8) not only captures the time-varying volatility of quarterly equity returns but also captures a time-varying risk premium. Although the unconditional equity risk premium is about 7% a year, the risk premium in the volatile-return regime is negative while that in the stable- (less volatile-) return regime is strongly positive. This suggests that investors have been poorly rewarded for taking on risk when the equity market is in the volatile regime.

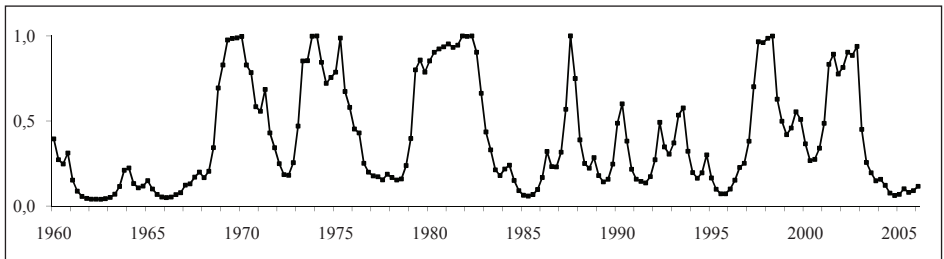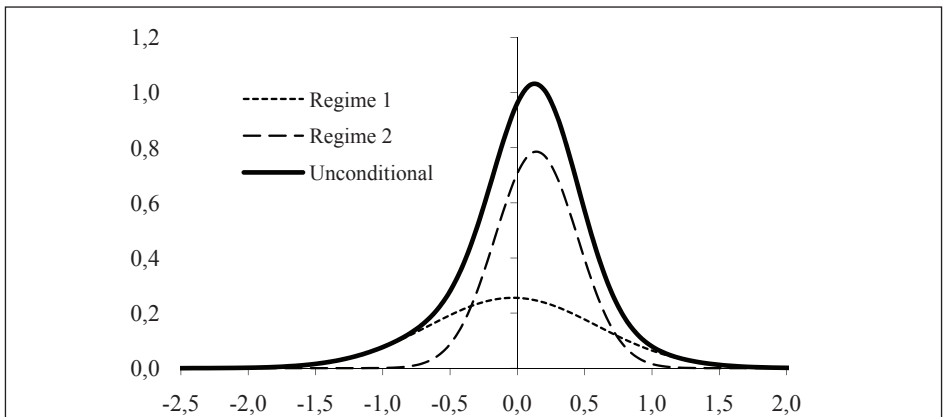Figure 8: Probability that $XSEQ_t$ was in the volatile regime

Figure 9: Unconditional and scaled conditional distributions for $XSEQ_t$

5.   A MULTIVARIATE MARKOV SWITCHING FRAMEWORK
5.1   INTRODUCTION
5.1.1   Harris (1996) fits univariate MS models to quarterly equity returns and inflation data in Australia. In Harris (1999), he develops a regime-switching vector autoregression model, in which a vector, comprising gross domestic product, inflation, share-price returns and changes in the long-bond yield, switches between two unobservable states. Although switching between states in Harris' (1996) univariate MS models is not perfectly correlated, the correlation is high. Hence, this vector-switching simplification provides a reasonable description of the joint process.

5.1.2   The Markov-switching vector autoregression (MSVAR) framework developed by Krolzig (1997) is restricted mainly to vector switching models in which switching between each univariate component of the vector is contemporaneously perfectly correlated. Krolzig (1997:127) also considers a bivariate model in which switching between each univariate series is not simultaneous, but where the lag in switching is fixed, which he refers to as 'intertemporally perfectly correlated regime shifts'.

5.1.3   A problem arises when switching between states in the univariate MS models is poorly correlated (see Figures 5–8 above). Suppose the number of variables is N and each variable is indexed $n \in \{1, 2,\dots, N\}$. Suppose also that the nth variable has possible states $s_t^n = j \in \{1, 2,\dots, M_n\}$ at time t. Then, if switching between states is less than perfectly correlated, the total number of states in the multivariate model is $M^{\text{Total}} = M_1 \cdot M_2 \cdot \dots \cdot M_N$.

5.1.4   Parsimony with respect to the number of regimes is extremely desirable in vector switching models since the number of observations feasible for the estimation of regime-dependent parameters drops as the number of regimes increases.

5.1.5   For example, a four-variable system with each univariate series containing two regimes requires a total of sixteen regimes. In the framework presented by Hamilton (1990) and Krolzig (1997), parameter estimates would be required for each of sixteen autoregressive models (as described by equation (2)), or worse, each of sixteen vector autoregressive models. Clearly, such an approach is not feasible.

5.2   A TRACTABLE MULTIPLE MARKOV-SWITCHING MODEL FRAMEWORK
5.2.1   Let S_t^* denote the joint state of the multivariate system at time t, and index this state as follows:

$S_t^* = 1$ if $s_t^1 = 1, s_t^2 = 1,\dots, s_t^{N-1} = 1, s_t^N = 1;$

$S_t^* = 2$ if $s_t^1 = 1, s_t^2 = 1,\dots, s_t^{N-1} = 1, s_t^N = 2;$

…

$S_t^* = M_1$ if $s_t^1 = 1, s_t^2 = 1,\dots, s_t^{N-1} = 1, s_t^N = M_N;$

$S_t^* = M_1 + 1$ if $s_t^1 = 1, s_t^2 = 1,\dots, s_t^{N-1} = 2, s_t^N = 1;$

…

$S_t^* = M^{Total}$ if $s_t^1 = M_1, s_t^2 = M_2,\dots, s_t^{N-1} = M_{N-1}, s_t^N = M_N.$ (9)

5.2.2 Let $c_{S_t^*}^n$ denote the parameter $c_{S_t^*}$ for the nth variable in state S_t^*, with similar notation for the autoregressive and residual standard-deviation parameters $\phi_{1,S_t^*}^n, \ldots, \phi_{p,S_t^*}^n \sigma_{S_t^*}^n$ of equation (2). Then place the following across-regime restrictions on the $c_{S_t^*}^n$ parameters and collect these into the reduced set $\theta^n = \left(c_{1,S_t^*}^n \phi_{1,S_t^*}^n, \ldots, \phi_{p,S_t^*}^n \sigma_{S_t^*}^n \right)$:

$$c_{S_t^*=j_n}^n = c_{S_t^*=k}^n \quad \text{iff } j \equiv k \left(\mathrm{mod} \prod_{l=n}^N M_l \right) \text{ or } \mathrm{int}\left(\frac{j-1}{\prod_{l=n+1}^N M_l} \right) = \mathrm{int}\left(\frac{k-1}{\prod_{l=n+1}^N M_l} \right) ; \quad (10)$$

with similar restrictions for the autoregressive and residual standard-deviation parameters of equation (2). This reduces the number of parameters in equation (2) to the same total number as would be estimated in the estimation of univariate MS models for each of the variables.

5.2.3 Let S_t^* be the realisation of a Markov chain with the probability of a switch from state i to state j $(i, j = 1, 2, \ldots, M^{\text{Total}})$ being:

$$\Pr\left(S_t^* = j \mid S_{t-1}^* = i\right) = p_{ij}^*. \quad (11)$$

From the indexation in (9), we can find $i_n, j_n \in \{1, 2, \ldots, M_n\}$ and $n \in \{1, \ldots, N\}$, such that:

$$\Pr\left(S_t^* = j \mid S_{t-1}^* = i\right) = \Pr\left(s_t^1 = j_1, \ldots, s_t^N = j_N \mid s_{t-1}^1 = i_1, \ldots, s_{t-1}^N = i_N\right). \quad (12)$$

5.2.4 Let Y_t^n represent the subset of our sample for the nth variable $(y_1^n, \ldots, y_t^n)$ with $t \in \{1, 2, \ldots, T\}$. Given the data to time t for all variables, $Y_t = \left\{ Y_t^1, \ldots, Y_t^N \right\}$, our joint inference about the unobserved states at times t and $t-1$ is:

$$\Pr\left(S_t^* = j, S_{t-1}^* = i \mid Y_t; \lambda\right) = \Pr\left(s_t^1 = j_1, \ldots, s_t^N = j_N, s_{t-1}^1 = i_1, \ldots, s_{t-1}^N = i_N \mid Y_t^1, \ldots, Y_t^N; \lambda\right); \quad (13)$$

where $\lambda = \left\{ P^*, \theta^1, \ldots, \theta^N, \rho \right\}$, P^* is the multiple-switching transition matrix of probabilities p_{ij}^*, and ρ is the vector of initial state probabilities across all variables (see Hamilton (1990) for further details).

5.2.5 If we assume that the inference in equation (13) is independent for each of the variables, then, as shown in Appendix B, the maximum-likelihood estimates for the transition probabilities satisfy:

$$\hat{p}_{ij}^* = \frac{\sum_{t=2}^T \prod_{n=1}^N \Pr\left\{ s_t^n = j_n, s_{t-1}^n = i_n \mid Y_T^n; \lambda^n \right\}}{\sum_{t=2}^T \prod_{n=1}^N \Pr\left\{ s_{t-1}^n = i_n \mid Y_T^n; \lambda^n \right\}}; \quad (14)$$

where $\lambda^n = \left\{ P^n, \theta^n, \rho^n \right\}$ and, for the nth variable, ρ^n is the vector of initial state probabilities and P^n is the transition matrix. The maximum-likelihood estimates for the parameters

$$c_{S_t^*}^n, \phi_{1,S_t^*}^n, ..., \phi_{p,S_t^*}^n \sigma_{S_t^*}^n \text{ and } \rho = \left\{ \rho^1, ..., \rho^N \right\}$$

are those obtained from estimation of the univariate MS model for the nth variable.

5.2.6   The maximum-likelihood estimate of the residual covariance matrix is the cross-product of the residuals from each of the variables in each state, each factor being weighted by the corresponding smoothed inferences (shown in the subject of the product in the denominator of equation (14)). For details, see Appendix B.

5.2.7   The assumption of independence in the smoothed inferences for each of the variables in equation (14) is not restrictive since it is exactly the assumption we made in fitting the univariate MS models. However, this assumption does not imply that switching in any one variable is independent of switching in the other variables, and by reference to equation (12) we find in general that:

$$\Pr\left(S_t^* = j \mid S_{t-1}^* = i\right) \neq \prod_{n=1}^{N} \Pr\left(s_t^n = j_n \mid s_{t-1}^n = i_n\right). \tag{15}$$

6.   EMPIRICAL ESTIMATION OF THE MMS MODEL

6.1   Table 2 summarises the univariate MS-model parameters obtained in section 4 and the ordering, n, of the variables (from left to right) that are used in the MMS model.

Table 2: MMS model parameters for the individual variables

	$INFL_t$	$SINT_t$	$LINT_t$	$XSEQ_t$
c_1	0,029 9	0,007 2	0,012 9	−0,033 3
c_2	0,094 4	0,023 4	0,021 4	0,140 4
ϕ_1	0,241 5	0,865 9	0,852 3	–
σ_1	0,032 995	0,005 074	0,004 082	0,620 57
σ_2	–	0,0154 66	0,008 319	0,305 81
p_{11}	0,968 2	0,947 2	0,9730	0,825 9
p_{22}	0,968 8	0,920 4	0,982 7	0,885 3
ℓ	350,403 7	604,650 1	524,408 8	−112,935

6.2   The calculation of the maximum-likelihood estimate of the covariance matrix of the residuals is described in Appendix B. The corresponding correlation matrix of the residuals is shown in Table 3.

6.3   Table 4 shows the maximum-likelihood estimate of the transition matrix P, using joint inferences for the period 1960:1–2006:2. The joint inferences for $LINT_t$ are based on parameter estimates for the period 1970:1–2006:2 (see section 4.4 for further details).

Table 3: Residual correlation matrix for the MMS model

	$INFL_t$	$SINT_t$	$LINT_t$	$XSEQ_t$
$INFL_t$	1,000			
$SINT_t$	0,147	1,000		
$LINT_t$	0,211	0,179	1,000	
$XSEQ_t$	0,064	−0,157	−0,082	1,000

6.4   In keeping with Hamilton's (1990) approach, the format of the transition matrix P in Table 4 is that the transition probability p_{ij} occurs in row j and column i, where p_{ij} is defined as in equation (11). Hence, the columns of P must sum to one.

Table 4: Multiple-switching transition matrix P for the MMS model

State	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0,75	0,07	0,02	0,00	0,26	0,04	0,00	0,00	0,04	0,00	0,00	0,00	0,01	0,00	0,00	0,00
2	0,21	0,91	0,01	0,03	0,12	0,37	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3	0,00	0,00	0,65	0,17	0,00	0,00	0,08	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00
4	0,00	0,00	0,12	0,73	0,00	0,00	0,02	0,05	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
5	0,01	0,00	0,00	0,00	0,44	0,14	0,03	0,00	0,00	0,00	0,00	0,00	0,03	0,01	0,00	0,00
6	0,00	0,00	0,00	0,00	0,08	0,35	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00
7	0,00	0,00	0,13	0,02	0,04	0,01	0,60	0,12	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,00
8	0,00	0,00	0,01	0,03	0,00	0,03	0,10	0,79	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,03
9	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,64	0,12	0,00	0,00	0,06	0,02	0,00	0,00
10	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,14	0,81	0,00	0,00	0,02	0,17	0,00	0,00
11	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,10	0,01	0,56	0,13	0,00	0,00	0,01	0,00
12	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,12	0,76	0,00	0,00	0,00	0,07
13	0,00	0,00	0,00	0,00	0,05	0,03	0,00	0,00	0,05	0,02	0,00	0,00	0,79	0,32	0,03	0,00
14	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,02	0,00	0,00	0,09	0,45	0,00	0,00
15	0,00	0,00	0,04	0,00	0,01	0,00	0,16	0,00	0,01	0,00	0,26	0,03	0,01	0,00	0,74	0,08
16	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,02	0,00	0,00	0,04	0,06	0,00	0,00	0,17	0,80

6.5   The MMS model is a discrete Markov process at discrete time points $t = 1, 2, 3,\ldots$ (in our case, quarter-ends) and these states are represented by S_t in equation (9). Only one state change is possible from one discrete time to the next. For example, state 1 at time $t-1$ can switch to any one (but only one) of the 16 states at time t, i.e. given $S_{t-1} = 1$, $S_t = I$, where I is any integer from 1 to 16.

6.6   Note that S_t also maps to one of the 16 binary sets represented by the four univariate states $(s_t^1, s_t^2, s_t^3, s_t^4)$. Hence it is possible for more than one of these univariate states

to switch at the same time. An example of this can be seen in Table 4 where state 1, with binary representation (1, 1, 1, 1), can go to states 2 (1, 1, 1, 2), 3 (1, 1, 2, 1), 5 (1, 2, 1, 1), 9 (2, 1, 1, 1) and 10 (2, 1, 1, 2), where it can be seen that more than one of the univariate states has switched from 1 to 2.

6.7   The ergodic or stable-state probabilities and the expected duration (in quarters) for each state are shown in Table 5.

Table 5: States, ergodic probabilities and durations

MMS state S^*	s^1	s^2	s^3	s^4	Ergodic probability	Duration
1	1	1	1	1	0,084	3,9
2	1	1	1	2	0,235	11,0
3	1	1	2	1	0,030	2,9
4	1	1	2	2	0,032	3,7
5	1	2	1	1	0,008	1,8
6	1	2	1	2	0,003	1,5
7	1	2	2	1	0,037	2,5
8	1	2	2	2	0,053	4,7
9	2	1	1	1	0,037	2,8
10	2	1	1	2	0,068	5,3
11	2	1	2	1	0,033	2,3
12	2	1	2	2	0,064	4,2
13	2	2	1	1	0,057	4,7
14	2	2	1	2	0,014	1,8
15	2	2	2	1	0,116	3,8
16	2	2	2	2	0,130	4,9

6.8   For example, row 2 in Tables 4 and 5 corresponds to the joint state with low expected inflation, low expected short- and long-term interest rates and stable excess equity returns. Historically, this state had the highest duration and the system remained in this state the longest.

6.9   Row 16 in Tables 4 and 5 corresponds to the joint state with high expected inflation, high expected short- and long-term interest rates and stable excess equity returns. This state had the next highest duration and was the second most persistent state in the system.

6.10   Intuition might suggest that $SINT_t$ and $LINT_t$ should be in the same state at any one time. From the ergodic probabilities in Table 5, the probability that $SINT_t$ and $LINT_t$ are in the same state is calculated by summing rows 1, 2, 9 and 10, as well as rows 7, 8, 15 and 16. An alternative might be to model $LINT_t$ as a base variable, and then to model

a variable $SPREAD_t = SINT_t - LINT_t$. However, using the tests outlined in section 4, the null hypothesis of a single state for $SPREAD_t$ is rejected in favour of a two-state MS model. Similar results hold for the spreads of $SINT_t$ over $INFL_t$ and $LINT_t$ over $INFL_t$. Hence, no reduction in states is possible from this approach.

6.11   For $LINT_t$ and $INFL_t$ another possible approach to modelling the dynamics of these variables is to model a series in which $INFL_t$ jumps first and then $LINT_t$ jumps with a variable lag. Such a model has been proposed by Durland & McCurdy (1994). However, from a financial perspective, while past expectations in the bond market failed to adequately reflect future inflation, one would not want to force such a structure on a stochastic model as it would not permit instances where the bond market correctly anticipates future inflation and inflation shocks.

6.12   With inflation targeting and tight monetary policy, one might expect short-term and possibly long-term interest rates to rise with, or following, an increase in inflation. However, if the system is in state 9 or 10 (representing high expected inflation and low expected short- and long-term interest rates), it is more likely to remain in that state than to move to state 13 or 14, which would equate to an increase in short-term interest rates. (See columns 9 and 10 of Table 4, where it can be seen that the transition probabilities in rows 9 and 10 are much higher than those in rows 13 and 14.)

6.13   Clearly, the estimated transition matrix parameters are purely a description of past experience and are not appropriate for projection purposes, especially given the current framework of inflation targeting in South Africa.

6.14   However, while certain aspects of these dynamics may not be useful for projection purposes, it is possible to adjust these aspects while conditioning on other aspects that may still be relevant. It is also possible to adjust the multiple-switching transition matrix and regime-specific parameters independently of one another. This makes the proposed framework very powerful as a tool for extracting those aspects of the past that one believes may be relevant to the future. It also makes market-consistent projections and stochastic scenario testing relatively simple.

6.15   Parameterisation of the model for projection purposes is beyond the scope of this paper. Also, the effect of parameter uncertainty on projection distributions cannot be assessed in the proposed framework. Joint parameter uncertainty could be modelled using the Markov-chain Monte Carlo approach, as in Harris (1999), but this is left for future research.

7.   CONCLUSION
7.1   The MMS framework presented in this paper allows the modelling of Markov-switching variables where switches in variables are not perfectly correlated.

7.2   The approach to modelling MS follows Krolzig (1997) in recommending a bottom-up approach in which univariate MS models are used to identify states. The correlation, or otherwise, of switching between these states can then be used to guide the choice of vector-switching or multiple Markov switching for the multivariate model.

7.3   Maximum-likelihood estimation of the parameters is shown to be relatively simple once the univariate Markov-switching models have been estimated.

7.4   The framework is used to estimate an MMS model of South African financial and economic variables. The variables estimated are by definition descriptive of a past that is unlikely to repeat itself. However, while certain dynamics may not be useful for projection purposes, it is possible to adjust these aspects while conditioning on other features that may be still be relevant for projection purposes.

7.5   As part of such adjustments, it is also possible to include current market conditions so that projections are market-consistent. Together with the framework suggested in Maitland (2002), it is possible to construct an arbitrage-free model of the local market. It is suggested that that framework can be used for various actuarial applications, especially those involving long-term projections.

REFERENCES

Aron, J & Muellbauer, J (2000). Inflation and output forecasting for South Africa: monetary transmission implications. *Working Paper Series*, 2000–23, Nuffield College, Oxford University

Bollerslev, T (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**, 307–27

Bollen, NPB (1998). Valuing options in regime switching models. *Journal of Derivatives* **6**, 38–49

Carter, J (1991). The derivation and application of an Australian stochastic investment model. *Transactions of the Institute of Actuaries of Australia* **1**, 315–428

Cosslett, SR & Lee, L (1985). Serial correlation in discrete variable models. *Journal of Econometrics* **27**, 79–97

Claasen, MS (unpublished). A systematic approach to the financial management and risk control of employee benefit funds. 3rd AFIR International Colloquium, 1993

Davies, R (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–54

Dickey, DA & Fuller, WA (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–31

Dickey, DA & Fuller, WA (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49**, 1057–72

Draper, NR & Smith, H (1981). *Applied Regression Analysis*. 2nd edn., Wiley

Durland, JM & McCurdy, TH (1994). Duration dependent transitions in a Markov Model of U.S. GNP Growth," *Journal of Business and Economic Statistics* **12**, 279–88

Engle, RF (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007

Engle, RF (1995). *ARCH Selected Readings*, Oxford University Press.

Garcia, R (1998). Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review* **39**, 763–88

Granger, CWJ & Newbold, P (1974). Spurious regressions in economics. *Journal of Econometrics* **2**, 111–20

Goldfeld, SM & Quandt, RE (1973). A Markov model for switching regressions. *Journal of Econometrics* **1**, 3–16

Hamilton, JD (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357–84

Hamilton, JD (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics* **45**, 39–70

Hamilton, JD (1994). *Time Series Analysis*. Princeton University Press

Hardy, MR (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal* **52**, 41–53

Harris, G (1994). On Australian stochastic share return models for actuarial use. *The Institute of Actuaries of Australia Quarterly Journal*, September, 34–54

Harris, G (1996). Market phases and cycles? A regime switching approach. *The Institute of Actuaries of Australia Quarterly Journal*, December, 28–44

Harris, G (1999). Regime switching vector autoregressions: a Bayesian Markov chain Monte Carlo approach. *ASTIN*, **29**, 1, May 1999

Homer, S & Sylla, R (2005). *A History of Interest Rates*. 4th edn., Wiley

Hua, G (unpublished). Modelling equity returns using a simple ARCH model. 5th AFIR International Colloquium, 1994, **2**, 537–65

Krolzig, H (1997). *Markov-Switching Vector Autoregressions : Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer, Berlin

Kwiatkowski, D, Phillips, PCB, Schmidt, P & Shin, Y (1992). Testing the null hypothesis of stationary against the alternative of a unit root. *Journal of Econometrics* **54**, 159–78

Maitland, AJ (unpublished a). A review of Thomson's stochastic investment model. *Transactions of the Actuarial Society of South Africa,* **11**(2), 1996, 437–490

Maitland, AJ (unpublished b). Non-Stationarity in some South African financial and economic series. 7th International AFIR Colloquium, 1997, **7**, 631–46

Maitland, AJ (2001). An empirical approach to immunization in South Africa. *South African Actuarial Journal* **1**, 114–33

Maitland, AJ (2002). Interpolating the South African yield curve using principal components analysis: a descriptive approach. *South African Actuarial Journal* **2**, 129–45

Perron, P (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica* **57**, 1361–401

Perron, P (1991). Test consistency with varying sampling frequency. *Econometric Theory* **7**, 341–68

Phillips, PCB & Perron, P (1988). Testing for a unit root in time series regression. *Biometrika* **75**, 335–46

Ranne, A (unpublished). The Finish stochastic investment model. Transactions of the 26th International Congress of Actuaries **7**, 213–38

Sherris, M, Tedesco, L & Zehnwirth, B (1999). Investment returns and inflation models: some Australian evidence. *British Actuarial Journal* **5**(1), 237–67

Thomson, RJ (1996). Stochastic investment modelling: the case of South Africa. *British Actuarial Journal* **2**, 765–802

Thomson, RJ (2006). A typology of models used in actuarial science. *South African Actuarial Journal* **6**, 19–36

Wilkie, AD (1986). A stochastic investment model for actuarial use. *Transactions of the Faculty of Actuaries* **39**, 341–403

Wilkie, AD (1995). More on a stochastic asset model for actuarial use. *British Actuarial Journal* **1**, 777–964

Whitten, S & Thomas, G (1999). A non-linear stochastic asset model for actuarial use. *British Actuarial Journal* **5**, 919–53

APPENDIX A

UNIVARIATE MS-MODEL PARAMETER ESTIMATES

A.1   In Tables A.1 to A.12, MSIAH(M)-AR(p) refers to MS models with M states and switching in any combination of the intercept term (I), the autoregressive terms (A) and the variance of the residuals (H). Standard errors are shown in parentheses. ℓ refers to the log-likelihood value, and 'Davies' refers to the p-value as defined by equation (4).

A.2   Tables A.1 to A.3 show the model estimates for $INFL_t$ for the periods 1960:1–2006:2, 1960:1–1998:4 and 1970:1–2006:2 respectively.

Table A.1: $INFL_t$ MS-model estimates (1960:1–2006:2)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,029 9 (0,005)	0,026 4 (0,006)	0,027 4 (0,005)
c_2	0,094 4 (0,013)	0,092 0 (0,012)	–
ϕ_1	0,241 5 (0,097)	0,237 6 (0,072)	0,668 5 (0,055)
σ_1	0,032 995	0,028 097	0,0401
σ_2	–	0,036 052	–
p_{11}	0,968 2	0,958 1	–
p_{22}	0,968 8	0,966 6	–
ℓ	350,403 7	352,062 6	333,425 8
AIC	–3,723 3	–3,730 4	–
SC	–3,618 8	–3,608 6	–
Davies	0,000	0,000	–

Table A.2: $INFL_t$ MS-model estimates (1960:1–1998:4)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,032 0 (0,006)	0,020 7 (0,005)	0,031 9 (0,006)
c_2	0,097 2 (0,012)	0,084 5 (0,010)	–
ϕ_1	0,213 4 (0,084)	0,252 7 (0,080)	0,643 1 (0,062)
σ_1	0,034 518	0,026 680	0,041 3
σ_2	–	0,039 215	–
p_{11}	0,972 7	0,977 5	–
p_{22}	0,984 7	1,000 0	–
ℓ	291,673 8	295,185 7	274,924 0
AIC	–3,686 1	–3,718 5	–
SC	–3,568 3	–3,581 1	–
Davies	0,000	0,000	–

Table A.3: $INFL_t$ MS-model estimates (1970:1–2006:2)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,048 7 (0,007)	0,050 4 (0,007)	0,043 3 (0,008)
c_2	0,100 1 (0,012)	0,102 5 (0,011)	–
ϕ_1	0,214 9 (0,082)	0,202 2 (0,082)	0,556 0 (0,069)
σ_1	0,035 371	0,039 314	0,040 7
σ_2	–	0,031 462	–
p_{11}	0,982 3	0,984 0	–
p_{22}	0,985 1	0,986 0	–
ℓ	273,720 0	275,322 8	261,355 6
AIC	–3,667 4	–3,675 7	–
SC	–3,544 8	–3,532 6	–
Davies	0,000	0,000	–

A.3 Tables A.4 to A.6 show the model estimates for $SINT_t$ for the periods 1960:1–2006:2, 1960:1–1998:4 and 1970:1–2006:2 respectively.

Table A.4: $SINT_t$ MS-model estimates (1960:1–2006:2)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,007 8 (0,002)	0,007 2 (0,002)	0,004 0 (0,002)
c_2	0,031 2 (0,003)	0,023 4 (0,006)	–
ϕ_1	0,860 9 (0,020)	0,865 9 (0,037)	0,962 6 (0,019)
σ_1	0,008 222 1	0,005 074 3	0,012 4
σ_2	–	0,015 466	–
p_{11}	0,936 8	0,947 2	–
p_{22}	0,829 5	0,920 4	–
ℓ	582,855 0	604,650 1	550,236 2
AIC	–6,236 3	–6,461 1	–
SC	–6,131 8	–6,339 2	–
Davies	0,000	0,000	–

Table A.5: $SINT_t$ MS-model estimates (1960:1–1998:4)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,007 4 (0,002)	0,006 1 (0,002)	0,004 1 (0,002)
c_2	0,031 3 (0,003)	0,021 8 (0,004)	–
ϕ_1	0,861 7 (0,023)	0,879 4 (0,030)	0,968 0 (0,020)
σ_1	0,008 487 3	0,004 866 2	0,012 9
σ_2	–	0,015 909	–
p_{11}	0,929 9	0,946 6	–
p_{22}	0,842 9	0,944 8	–
ℓ	481,199 1	502,697 8	455,000 7
AIC	–6,131 6	–6,396 1	–
SC	–6,013 8	–6,258 7	–
Davies	0,000	0,000	–

Table A.6: $SINT_t$ MS-model estimates (1970:1–2006:2)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,010 9 (0,003)	0,011 8 (0,002)	0,006 6 (0,003)
c_2	0,034 6 (0,004)	0,031 5 (0,004)	–
ϕ_1	0,835 8 (0,029)	0,816 9 (0,022)	0,944 1 (0,027)
σ_1	0,009 126 4	0,005 873 9	0,013 8
σ_2	–	0,014 679	–
p_{11}	0,910 7	0,918 5	–
p_{22}	0,829 2	0,911 1	–
ℓ	440,839 7	450,952 1	418,699 4
AIC	−5,956 7	−6,081 5	–
SC	−5,834 1	−5,938 5	–
Davies	0,000	0,000	–

A.4   Tables A.7 to A.9 show the model estimates for $LINT_t$ for the periods 1960:1–2006:2, 1960:1–1998:4 and 1970:1–2006:2 respectively.

Table A.7: $LINT_t$ MS-model estimates (1960:1–2006:2)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,003 1 (0,001)	0,001 7 (0,004)	0,002 3 (0,001)
c_2	0,016 8 (0,002)	0,006 6 (0,005)	–
$\phi_{1,1}$	0,959 7 (0,012)	0,980 3 (0,099)	0,980 2 (0,013)
$\phi_{1,2}$	–	0,950 3 (0,039)	
σ_1	0,005 118 9	0,002 409 4	–
σ_2	–	0,008 200 8	0,006 6
p_{11}	0,934 3	0,966 8	–
p_{22}	0,428 8	0,986 1	–
ℓ	678,034 5	705,533 5	–
AIC	−7,265 2	−7,540 9	667,186 3
SC	−7,160 8	−7,401 6	–
Davies	0,000	0,000	–

Table A.8: $LINT_t$ MS-model estimates (1960:1–1998:4)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,003 0 (0,001)	0,001 6 (0,001)	0,002 3 (0,001)
c_2	0,017 1 (0,002)	0,027 1 (0,008)	–
$\phi_{1,1}$	0,964 9 (0,011)	0,983 1 (0,015)	0,984 6 (0,013)
$\phi_{1,2}$	–	0,817 4 (0,057)	–
σ_1	0,004 961 2	0,002 51	0,006 5
σ_2	–	0,008 190 3	–
p_{11}	0,938 7	0,973 2	–
p_{22}	0,420 7	0,984 4	–
ℓ	572,267 1	605,621 4	561,765 7
AIC	−7,306 7	−7,711 2	–
SC	−7,188 9	−7,554 2	–
Davies	0,000	0,000	–

Table A.9: $LINT_t$ MS-model estimates (1970:1–2006:2)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	0,016 7 (0,005)	0,012 9 (0,004)	0,004 4 (0,002)
c_2	0,028 4 (0,007)	0,021 4 (0,006)	–
ϕ_1	0,813 4 (0,054)	0,852 3 (0,041)	0,964 6 (0,020)
σ_1	0,006 360 6	0,004 081 8	0,007 4
σ_2	–	0,008 318 9	–
p_{11}	0,951 8	0,973 0	–
p_{22}	0,953 2	0,982 7	–
ℓ	513,379 1	524,408 8	511,078 5
AIC	−6,950 4	−7,087 8	–
SC	−6,827 8	−6,944 7	–
Davies	0,203 4	0,000 0	–

A.5   Tables A.10 to A.12 show the model estimates for $XSEQ_t$ for the periods 1960:1–2006:2, 1960:1–1998:4 and 1970:1–2006:2 respectively.

Table A.10: $XSEQ_t$ MS-model estimates (1960:1–2006:2)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	−0,660 6 (0,202)	−0,033 3 (0,093)	0,070 7 (0,034)
c_2	0,147 3 (0,045)	0,140 4 (0,041)	–
σ_1	0,402 25	0,620 57	0,468 0
σ_2	–	0,305 81	–
p_{11}	0,336 1	0,825 9	–
p_{22}	0,931 1	0,885 3	–
ℓ	−118,751 3	−112,934 5	−121,516 6
AIC	1,337 9	1,285 8	–
SC	1,424 9	1,390 2	–
Davies	0,136 8	0,003 4	–

Table A.11: $XSEQ_t$ MS-model estimates (1960:1–1998:4)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	−0,669 3 (0,209)	−0,035 0 (0,107)	0,060 1 (0,038)
c_2	0,141 2 (0,048)	0,117 2 (0,041)	–
σ_1	0,405 82	0,657 36	0,477 3
σ_2	–	0,306 53	–
p_{11}	0,329 4	0,838 4	–
p_{22}	0,929 0	0,900 4	–
ℓ	−102,235 1	−95,875 9	−104,784 9
AIC	1,383 7	1,314 5	–
SC	1,481 9	1,432 3	–
Davies	0,164 6	0,002 5	–

Table A.12: $XSEQ_t$ MS-model estimates (1970:1–2006:2)

	MSI(2)–AR(1)	MSIH(2)–AR(1)	AR(1)
c_1	−0,308 5 (0,507)	−0,039 6 (0,123)	0,058 6 (0,041)
c_2	0,161 5 (0,099)	0,121 0 (0,054)	–
σ_1	0,454 70	0,659 33	0,496 2
σ_2	–	0,337 45	–
p_{11}	0,339 8	0,776 4	–
p_{22}	0,815 8	0,863 0	–
ℓ	−103,943 0	−99,564 1	−104,348
AIC	1,492 4	1,446 1	–
SC	1,594 5	1,568 7	–
Davies	0,847 1	0,088 4	–

APPENDIX B

MAXIMUM-LIKELIHOOD ESTIMATION

B.1   Suppose we have a sample of size T from a vector-valued autoregressive process $y_t \in \Re^n$, and suppose that the parameters, θ, of this process switch between a finite number of alternatives based on the unobserved states, s_t. If the probability of switching from one state to another can be expressed in the form of a Markov chain, then Hamilton (1990:51) shows that the maximum-likelihood estimates for the transition probabilities satisfy :

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} \Pr\{ s_t = j, s_{t-1} = i \mid Y_T; \lambda \}}{\sum_{t=2}^{T} \Pr\{ s_{t-1} = i \mid Y_T; \lambda \}}. \tag{B1}$$

B.2   Using the notation introduced in section 5 of this paper, it follows that:

$$\hat{p}_{ij}^* = \frac{\sum_{t=2}^{T} \Pr\{ S_t^* = j, S_{t-1}^* = i \mid Y_T; \lambda \}}{\sum_{t=2}^{T} \Pr\{ S_{t-1}^* = i \mid Y_T; \lambda \}}; \tag{B2}$$

which, from equation (13), can be written as:

$$\hat{p}_{ij}^* = \frac{\sum_{t=2}^{T} \Pr\left(s_t^1 = j_1, .., s_t^N = j_N, s_{t-1}^1 = i_1, .., s_{t-1}^N = i_N \mid Y_T^1, .., Y_T^N; \lambda \right)}{\sum_{t=2}^{T} \Pr\left(s_{t-1}^1 = i_1, .., s_{t-1}^N = i_N \mid Y_T^1, .., Y_T^N; \lambda \right)}. \tag{B3}$$

B.3   If we now assume that the joint smoothed inference in equation (B3) is independent for each of the variables, then:

$$\Pr\left(s_t^1 = j_1, .., s_t^N = j_N, s_{t-1}^1 = i_1, .., s_{t-1}^N = i_N \mid Y_t^1, .., Y_t^N; \lambda \right) = \prod_{n=1}^{N} \Pr\left(s_t^n = j_{n1}, s_{t-1}^n = i_n \mid Y_t^n; \lambda^n \right);$$

and: $$\Pr\left(s_t^1 = j_1, .., s_t^N = j_N \mid Y_t^1, .., Y_t^N; \lambda \right) = \prod_{n=1}^{N} \Pr\left(s_t^n = j_{n1} \mid Y_t^n; \lambda^n \right). \tag{B4}$$

B.4   This assumption is not restrictive since it is exactly the assumption we made in fitting the univariate MS models in our bottom-up identification strategy. Substituting into equation (B3) gives the desired result:

$$\hat{p}_{ij}^{*} = \frac{\sum_{t=2}^{T}\prod_{n=1}^{N}\Pr\left\{s_t^n = j_n, s_{t-1}^n = i_n \mid Y_T^n; \lambda^n\right\}}{\sum_{t=2}^{T}\prod_{n=1}^{N}\Pr\left\{s_{t-1}^n = i_n \mid Y_T^n; \lambda^n\right\}}. \tag{B5}$$

B.5 Generalising from equation (2), let $\varepsilon_t^{n,j}$ denote the error term associated with regime $S_t^{*}= j$ for the nth variable at time t, and collect these into a vector, $\varepsilon^{n,j}$, spanning the length of the time series. Let $\xi_{t|T}^{j}$ denote the smoothed regime probability, $\Pr\left(S_t^{*} = j \mid Y_T; \lambda\right)$, and collect these into a vector, ξ_T^{j}, spanning the length of the time series. Then, under the assumption of heteroscedasticity in either of the two variables under consideration, Krolzig (1997:108) shows that the maximum-likelihood estimate of the covariance in state j between variables $n=p$ and $n=q$ is:

$$Cov_j\left(y^p : y^q\right) = \left(\sum_t \varepsilon_t^{p,j} \cdot \varepsilon_t^{q,j} \cdot \xi_{t|T}^{j}\right)/T_j, \quad where \ \ T_j = \sum_t \xi_{t|T}^{j}. \tag{B6}$$

B.6 This is the usual cross-product of error terms, weighted by the smoothed regime probability at each point in time and divided by the probability-weighted time that the series have been in regime j.

B.7 Krolzig (*op. cit.*) also shows that the maximum-likelihood estimate of the covariance under homoscedasticity is the cross-product of error terms, weighted by the smoothed regime probability at each time and then summed across all regimes before dividing by T.

$$Cov_j\left(y^p : y^q\right) = \frac{1}{T}\sum_{j=1}^{M^{\text{Total}}} T_j Cov_j\left(y^p : y^q\right) \tag{B7}$$

B.8 From the univariate analysis of the series, it may transpire that certain variables are heteroscedastic while others are homoscedastic. Let $S_t^{*} = u(k)$ denote the kth union of states with constant variance for variables $n = p$ and $n = q$. Then the maximum likelihood estimate of the covariance for the kth union of states is:

$$Cov_{u(k)}\left(y^p : y^q\right) = \sum_{u(k)} T_j \cdot Cov_j\left(y^p : y^q\right)/\sum_{u(k)} T_j. \tag{B8}$$

(The summation is taken across all states that are homoscedastic.)

B.9 If the correlations are assumed to be constant across all states, an estimate of the correlation between variables $n = p$ and $n = q$ can be obtained as follows:

$$Corr\left(y^p : y^q\right) = \sum_{j=1}^{M^{\text{Total}}} T_j \cdot \left(Cov_j\left(y^p : y^q\right)/\sqrt{\sigma_j^p \cdot \sigma_j^q}\right)/T. \tag{B9}$$