Moments of Markov Switching Models

By

Allan Timmermann

DISCUSSION PAPER 323

May 1999

FINANCIAL MARKETS GROUP

AN ESRC RESEARCH CENTRE

LONDON SCHOOL OF ECONOMICS

ISSN 0956-8549-323

Moments of Markov Switching Models*

Allan Timmermann
UCSD and London School of Economics†

July 27, 1999

Abstract

This paper derives the moments for a range of Markov switching models. We characterize in detail the patterns of volatility, skewness and kurtosis that these models can produce as a function of the transition probabilities and parameters of the underlying state densities entering the switching process. The autocovariance of the level and squares of time series generated by Markov switching processes is also derived and we use these results to shed light on the relationship between volatility clustering, regime switches and structural breaks in time series models.

JEL Code: C1.

Key Words: Markov Switching, Higher Order Moments, Mixtures of Normals, Volatility Clustering.

* Two anonymous referees and an associate editor provided many useful suggestions for improvements of the paper. Discussions with Jim Hamilton and Martin Sola also were very helpful.

† Financial Markets Group, Houghton Street, London WC2A 2AE, England.

1. Introduction

Markov switching models have become increasingly popular in economic studies of industrial production, interest rates, stock prices and unemployment rates. However, so far no study has characterized in any detail the moments that these models can generate. This is an important omission since Markov switching models are often adopted by reseachers wishing to account for specific features of economic time series such as the asymmetry of economic activity over the business cycle (Hamilton (1989), Neftci (1984)) or the fat tails, volatility clustering and mean reversion in stock prices (Cecchetti, Lam and Mark (1990), Pagan and Schwert (1990), Turner, Startz and Nelson (1989)) and interest rates (Gray (1996), Hamilton (1988)). These features translate into the higher order moments and serial correlation of the data generating process, so a characterization of the moments and autocorrelation function generated by Markov switching will allow researchers to better understand when to make use of this class of models. The contribution of this paper is to characterize the moments and serial correlation of the level and the squared values of Markov switching processes.

Markov switching models belong to a general class of mixture distributions. Econometricians' initial interest in this class of distributions was based on their ability to flexibly approximate general classes of density functions and generate a wider range of values for the skewness and kurtosis than is obtainable through use of a single distribution. Along these lines Granger and Orr (1972) and Clark (1973) considered time-independent mixtures of normal distributions as a means of modeling non-normally distributed data. These initial models, however, did not capture the time-dependence in the conditional variance found in many economic time series, as evidenced by the vast literature on ARCH models that started with Engle (1982).

By allowing the mixing probabilities to display time-dependence, Markov switching models can be seen as a natural generalisation of the original time-independent mixture of normals model and we show that this feature enables them to generate a wide range of coefficients of skewness, kurtosis and serial correlation even when based on a very small number of underlying states.

Regime switches in economic time series can be parsimomiously represented by Markov switching models by letting the mean, variance and possibly the dynamics of the series depend on the realization of a finite number of discrete states. In increasing order of generality we consider in this paper three types of models, each of which has been adopted in applied econometric studies. The basic Markov switching model is

$$y_t = \mu_{s_t} + \sigma_{s_t} \varepsilon_t; \qquad (1)$$

where $S_t = 1; 2; \ldots; k$ denotes the unobserved state indicator which follows an ergodic k-state Markov process and ε_t is a zero-mean random variable which is identically and independently distributed over time. The number of states, k, is assumed to be finite. This model has been used in empirical work by Engel and Hamilton (1990). The second Markov switching model allows for state-independent autoregressive dynamics:

$$y_t = \mu_{s_t} + \sum_{j=1}^{q} \phi_j \left(y_{t-j} - \mu_{s_{t-j}} \right) + \sigma_{s_t} \varepsilon_t. \tag{2}$$

Hamilton (1989) used this type of autoregressive model to analyze growth in US GDP. Finally we also consider a model with state-dependent dynamics in the autoregressive part:

$$y_t = \mu_{s_t} + \sum_{j=1}^{q} \phi_{j s_{t-j}} \left(y_{t-j} - \mu_{s_{t-j}} \right) + \sigma_{s_t} \varepsilon_t. \tag{3}$$

For all models the errors are assumed to be independently distributed with respect to all past and future realizations of the state variable, i.e. $F(\varepsilon_t \mid S_{t+i}) = F(\varepsilon_t)$ for all values of i, where F(.) denotes the cumulative density function of ε_t.[1] The stochastic transition probability matrix P that determines the evolution in S_t is given by

$$\begin{aligned} Prob(S_{t+j} = j \mid S_t = i, S_{t-1} = k, ...,) &= Prob(S_{t+1} = j \mid S_t = i) = p_{ij}, \\ 0 \leq p_{ij} \leq 1, & \quad \sum_{j=1}^{k} p_{ij} = 1 \text{ for all } i, \end{aligned} \tag{4}$$

so that the states follow a homogenous Markov chain. In practice, if the process is not irreducible and not all states are visited with non-zero probability in the steady state, then the moment analysis can simply be conducted on the subset of states occurring with non-zero stationary probability.[2] The higher is p_{ii}, the longer the process is expected to remain in state i. For this reason we shall refer to p_{ii} as measuring the 'persistence' of the mixing of the underlying state densities.

We find in this paper that these persistence parameters are very important in determining the higher order moments of the Markov switching process. Furthermore, once autoregressive parameters are introduced into the process as in the second and third switching models, this gives rise to cross-product terms that enhance the set of third and fourth order moments and the patterns in

[1] Herein lies a key difference to ARCH models which is another type of time-dependent mixture process. While Markov switching models mix a finite number of states with different mean and volatility parameters based on an exogenous state process, ARCH models mix distributions with volatility parameters drawn from an infinite set of states driven by lagged innovations to the series.

[2] An analysis of the unconditional moments starting from the steady state need not assume that the Markov process is irreducible since, if either a single state or a block of states is absorbing, all other states will have zero steady state probabilities.

serial correlation and volatility dynamics that these models can generate. Even low-order autoregressive Markov switching processes with a small number of states provide the basis for very flexible econometric models. Our results prove useful to understanding the literature on structural breaks and volatility clustering since the models that have been adopted in this literature are closely related to Markov switching models with infrequent switches between states.

The paper is organized as follows. Section 2 provides results on the moments of the basic Markov switching process without autoregressive terms. Section 3 extends the results to the second model with state-independent, autoregressive terms, while Section 4 analyses the moments of the most general model with state-dependence both in the mean, variance and autoregressive coefficients. Section 5 reports the autocovariance structure generated by the three Markov switching models and Section 6 discusses their relation to structural break and ARCH models.

2. The Basic Markov Switching Model

Let $\underset{\sim}{\pi} = (\pi_1, ..., \pi_k)'$ be the k-vector of steady state (ergodic) probabilities that solve the system of equations $P' \underset{\sim}{\pi} = \underset{\sim}{\pi}$. These probabilities can be computed as the eigenvector (scaled so that its elements sum to one) associated with the unit eigenvalue of P'. Assuming that the state process started an infinite number of periods back in time or that it is initialized by a random variable drawn from the stationary distribution, $\underset{\sim}{\pi}$ is the vector of unconditional probabilities applying to the k states. The following proposition provides the moments of the basic Markov switching model

Proposition 1

Suppose the stationary Markov switching process (1), (4) started from its steady state characterized by the set of unconditional probabilities $(\underset{\sim}{\pi})$. Then the centered moments of the process are given by

$$E[(y_t - \mu)^n] = \sum_{i=1}^{k} \pi_i \sum_{j=0}^{n} {}_nC_j \ \sigma_i^j \ E[\varepsilon_t^j](\mu_i - \mu)^{n-j};$$

where ${}_nC_j = \frac{n!}{(n-j)!j!}$. When ε_t is t-distributed with τ degrees of freedom, the centered moments are

$$E[(y_t - \mu)^n] = \sum_{i=1}^{k} \pi_i \sum_{j=0}^{n} {}_nC_j \ \sigma_i^j \ a_j(\mu_i - \mu)^{n-j}$$

where

$$a_j = \frac{\tau^{j/2} \beta\left(\frac{j+1}{2}, \frac{\tau - j}{2}\right)}{\beta\left(\frac{1}{2}, \frac{\tau}{2}\right)}; \quad \text{if } \tau > j \text{ and } j \text{ is even}$$

$$a_j = 0; \quad \text{otherwise}$$

and $\beta(.)$ is the Beta function. When ε_t follows a normal distribution we have

$$E[(y_t - \mu)^n] = \sum_{i=1}^{k} \pi_i \sum_{j=0}^{n} {}_nC_j \, \sigma_i^j \, b_j (\mu_i - \mu)^{n-j};$$

where

$$b_j = \prod_{h=1}^{j/2} (2h - 1); \quad \text{provided } j \text{ is even and}$$
$$b_j = 0; \quad \text{otherwise.}$$

A proof of Proposition 1 is given in the appendix. Since researchers are often particularly interested in the variance, skewness and kurtosis of their data we characterize these moments more explicitly for the empirically popular mixture of normals model with two states:

Corollary 1

Suppose that there are two states and that the increments are Gaussian. Then the unconditional variance (σ^2), coefficient of skewness ($\sqrt{b_1}$) and coefficient of excess kurtosis (b_2) of the basic Markov switching process (1), (4) are given by:

$$\begin{aligned}
\sigma^2 &= (1-\pi_1)\sigma_2^2 + \pi_1\sigma_1^2 + (1-\pi_1)\pi_1(\mu_2-\mu_1)^2; \\
\sqrt{b_1} &\equiv \frac{E[(y_t-\mu)^3]}{(E[(y_t-\mu)^2])^{3/2}} = \frac{\pi_1(1-\pi_1)(\mu_1-\mu_2)\left\{3(\sigma_1^2-\sigma_2^2)+(1-2\pi_1)(\mu_2-\mu_1^2)^2\right\}}{((1-\pi_1)\sigma_2^2+\pi_1\sigma_1^2+(1-\pi_1)\pi_1(\mu_2-\mu_1)^2)^{3/2}} \\
b_2 &\equiv \frac{E[(y_t-\mu)^4] - \left(3E[(y_t-\mu)^2]\right)^2}{(E[(y_t-\mu)^2])^2} = \frac{a}{b} \\
\text{where} \quad a &= 3\pi_1(1-\pi_1)(\sigma_2^2-\sigma_1^2)^2 + 6(\mu_2-\mu_1)^2\pi_1(1-\pi_1)(2\pi_1-1)(\sigma_2^2-\sigma_1^2) \\
&\quad +\pi_1(1-\pi_1)(\mu_2-\mu_1)^4(1-6\pi_1(1-\pi_1)): \\
b &= \left((1-\pi_1)\sigma_2^2+\pi_1\sigma_1^2+(1-\pi_1)\pi_1(\mu_2-\mu_1)^2\right)^2
\end{aligned}$$

For a proof, see the appendix. It follows from Corollary 1 that, when the innovations are drawn from a Gaussian density, a necessary condition for the Markov switching process to generate skewness is that the means in the states differ ($\mu_1 \neq \mu_2$) and that differences in the variances of the states alone are insufficient to generate skewness.[3] This also suggests that Markov switching models fitted to high-frequency financial data whose means are often very small in all states may have trouble replicating successfully the skewness found in these data.

[3] This is similar to the result in Bollerslev (1986) that the skewness of standard GARCH models without leverage effects is zero.

It is useful to evaluate these expressions through some numerical examples. To establish a benchmark for some relevant parameter values, we report in Table 1 the maximum likelihood estimates from a simple two-state Markov switching model fitted to monthly excess returns on the stock price index in four countries. In three out of four markets the switching model identifies an 'outlier' state in which excess returns are very volatile with a large negative mean and a 'normal' state with low volatility and small positive mean returns. The difference between the volatility parameter in the two states can be very large and is around three times higher in state 2 than in state 1 for the UK and US markets. Likewise the mean return differentials across the two states are very large in three of the four markets although the mean parameter is imprecisely estimated in the state with high volatility. In Germany the mean return parameters in the two states are roughly of the same magnitude but of opposite signs. Based on this evidence we consider in our numerical examples a variety of combinations of differences in the relative size of the mean and volatility parameters in two or more states.

Figure 1 presents a plot of the coefficients of skewness and excess kurtosis of a two-state Markov switching process with parameters $\underset{\sim}{\mu}_s = (1\ -1)$; $\underset{\sim}{\sigma}_s = (1\ 1)$. Thus the skewness and kurtosis in Figure 1 is entirely driven by differences in the mean parameters of the model, while the variances are identical in the two states. To construct the figure, these parameters were kept fixed and the probabilities of staying in the two states were varied over the grids [0.01, 0.99]. High, positive values of both skewness and excess kurtosis are obtained when the probability of staying in the second state (p_{22}) is around 0.9 and the probability of staying in the first state (p_{11}) is not too high. Likewise, a large negative skewness and a large positive excess kurtosis is obtained when p_{11} is around 0.9 and p_{22} is small. A large, negative excess kurtosis is obtained along the line where the two transition probabilities are identical so that the process spends the same time in the two states. This has the effect of moving some of the tail probability mass towards the centre and hence lowers the kurtosis.

To investigate the impact on the skewness and kurtosis of switching between two states with different means and volatilities, we used a second parameter configuration, setting $\underset{\sim}{\mu}_s = (1\ -3)$; $\underset{\sim}{\sigma}_s = (2\ 4)$. This parameter configuration is close to what was found for three of the four models in Table 1. The plot, which is presented in Figure 2, is very different from Figure 1. Now a very large excess kurtosis is produced when the probability of staying in the low volatility state (p_{11}) is high and the probability of staying in the high volatility state (p_{22}) is low. This case mixes large, but rare, outliers with a low dispersion distribution, thus generating the extreme kurtosis. Since the high volatility state also has a low mean, a high value of p_{11} and a low value of p_{22} generate large negative skewness because of the presence in the distribution of large negative outliers.[4]

Researchers using Markov switching models sometimes identify states with similar means but very different volatilities. To shed light on the sort of moments this situation gives rise to, Figure 3 plots the excess kurtosis against combinations of variance parameters in two states that vary between 0.1 and 10. The mean parameters are identical in the two states and P_{11} = 0.97, P_{22} = 0.75, so state 1 is highly persistent while state 2 is not, matching the estimates in Table 1. Since the mean parameters in the two states are the same, the skewness equals zero throughout. However, a very substantial excess kurtosis is generated when σ_1^2 is very small and σ_2^2 is very high. In this case the process spends most of the time in the low volatility regime but occasionally shifts to a high volatility state, thus increasing the kurtosis.

These findings are also representative of mixture processes with more than two states. Suppose that the asssumptions of Proposition 1 hold and that the increments are normally distributed. We refer to such a process as MSI. Then it follows from the proposition that the higher order moments have the convenient representation

$$E[(y_t - \mu)^n] = \underset{\sim}{\pi}' \, M_n' \, \underset{\sim n}{\xi} ; \tag{5}$$

where $\underset{\sim n}{\xi}$ is the n + 1-vector $(1 \; {}_nC_1 \; ... \; {}_nC_j \; ... \; {}_nC_{n-1} \; 1)'$, and

$$M_n = \begin{bmatrix} (\mu_1 - \mu)^n & . & . & (\mu_k - \mu)^n \\ 0 & . & . & 0 \\ \sigma_1^2(\mu_1 - \mu)^{n-2} & . & . & \sigma_k^2(\mu_k - \mu)^{n-2} \\ 0 & . & . & 0 \\ 3\sigma_1^4(\mu_1 - \mu)^{n-4} & . & . & 3\sigma_k^4(\mu_k - \mu)^{n-4} \\ . & . & . & . \\ b_n\sigma_1^n & . & . & b_n\sigma_k^n \end{bmatrix}$$

is the (n+1)xk matrix of moments $E[\sigma_{s_t}^j \epsilon_t^{2j} (\mu_{s_t} - \mu)^{n-j}$ ...

symmetric with respect to the underlying state densities. To illustrate this point, Figure 4 plots the coe±cients of skewness and kurtosis as a function of p_{11} and p_{22} for the 3-state case with $p_{13} = p_{23} = 0{:}1$; $p_{31} = p_{32} = p_{33} = 1{=}3$; $\underset{\sim s}{¹} = (1\ {-3}\ 0)$, $\underset{\sim s}{¾} = (2\ 4\ 3)$. Hence a third 'intermediate' state is mixed with the two states from Figure 2. Naturally, the same high values of skewness and kurtosis are not reached in Figures 2 and 4 since p_{11} and p_{22} are now constrained to be less than 0.90, but the shapes of the two ¯gures in this range of p_{11}, p_{22}-values are very similar.

3. The Simple Autoregressive Markov Switching Model

Before dealing with the general case of equation (2) that involves q lags of $(y_t - ¹_{s_t})$, we ¯rst consider the more tractable ¯rst-order autoregressive model with normally distributed increments. This case demonstrates the e®ect of additional terms entering the expression for the moments that researchers pay most attention to. Notice from (2) that when q = 1, the Markov switching process becomes

$$y_t = ¹_{s_t} + Á_1(y_{t-1} - ¹_{s_{t-1}}) + ¾_{s_t}²_t; \tag{6}$$

which we shall refer to as MSII. Upon substituting backward we get

$$y_t - ¹_{s_t} = \sum_{i=0}^{\infty} Á_1^i ¾_{s_{t-i}}²_{t-i}{:} \tag{7}$$

From the assumption that $E[²_{t-i} \mid S_t] = E[²_{t-i}] = 0$ it still holds that $E[y_{t-1} - ¹_{s_{t-1}} \mid S_{t-1}] = 0$. Thus the ¯rst moment of y_t is unchanged and $E[y_t] = \underset{\sim}{¼}' E[\underset{\sim t}{y} | \underset{\sim t}{S}] = \underset{\sim}{¼}' \underset{\sim s}{¹}$, where $E[\underset{\sim t}{y} | \underset{\sim t}{S}]$ is the k-vector whose i'th element consists of $E[y_t \mid S_t = i]$.

To state concisely the moments for this process, let $\underset{\sim}{¶}$ be a k-vector of ones, while I_k is the k-dimensional identity matrix, $\odot$ is the element-by-element multiplication operator and B is the $(k \times k)$ matrix of transition probabilities for the 'time-reversed' Markov chain that moves back in time:

$$\text{Prob}\,(S_t = j \mid S_{t+1} = i) = b_{ij}; \quad 0 \,\square\, b_{ij} \,\square\, 1; \quad \sum_{j=1}^{k} b_{ij} = 1{:} \tag{8}$$

Since $\text{Prob}(S_t = j \cap S_{t+1} = i) = \text{Prob}(S_{t+1} = i \mid S_t = j)\text{Prob}(S_t = j) =$
$\text{Prob}(S_t = j \mid S_{t+1} = i)\ \text{Prob}(S_{t+1} = i)$, the 'backward' transition probability matrix B is related to the 'forward' transition probabilities as follows:[5]

$$b_{ij} = p_{ji}\left(\frac{¼_j}{¼_i}\right); \tag{9}$$

[5] If $b_{ij} = p_{ij}$ for all i and j, then the process is said to be time-reversible. Since the diagonal elements of B and P are identical, the probability of remaining in a given state is always the same regardless of whether time moves forward or backward. Furthermore, in the case with two states it is easy to verify from the de¯nitions of $¼_1$ and $¼_2$ that the Markov chain will be time reversible, although this does not hold generally for processes with several states.

assuming $\pi_i > 0$, and $b_{ij} = 0$ if j belongs to a group of absorbing states while i does not. Again, if this condition is not satisfied, the analysis can be performed on the sub-set of states for which the steady-state probabilities exceed zero. Using this notation, Proposition 2 presents expressions for the second to fourth centered moments of the first-order autoregressive Markov switching model:[6]

Proposition 2

Suppose y_t follows the first-order autoregressive Gaussian Markov switching process

$$y_t = \mu_{s_t} + \phi_1(y_{t-1} - \mu_{s_{t-1}}) + \sigma_{s_t}\varepsilon_t; \quad |\phi_1| < 1; \quad \varepsilon_t \sim IIN(0;1);$$

and assume that the process started from its steady state distribution. Then the variance, skewness and kurtosis of y_t are given by

$$\begin{aligned}
E\left[(y_t - \mu)^2\right] &= \underset{\sim}{\pi}'\left((\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) + \frac{\underset{\sim s}{\sigma}^2}{1-\phi_1^2}\right); \\
E\left[(y_t - \mu)^3\right] &= \underset{\sim}{\pi}'\left((\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota})\right) \\
&\quad + 3\phi_1^2\,\underset{\sim}{\pi}'\left((B \cdot (I_k - \phi_1^2 B)^{-1}\underset{\sim s}{\sigma}^2) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota})\right) \\
&\quad + 3\,\underset{\sim}{\pi}'\left((\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot \underset{\sim s}{\sigma}^2\right); \\
E\left[(y_t - \mu)^4\right] &= \underset{\sim}{\pi}'\left((\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota})\right) \\
&\quad + 6\,\underset{\sim}{\pi}'\left((\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot \underset{\sim s}{\sigma}^2\right) \\
&\quad + \underset{\sim}{\pi}'(I_k - \phi_1^4 B)^{-1}\left(3\,\underset{\sim s}{\sigma}^4 + 6\phi_1^2(B \cdot (I_k - \phi_1^2 B)^{-1}\underset{\sim s}{\sigma}^2) \odot \underset{\sim s}{\sigma}^2\right) \\
&\quad + 6\phi_1^2\,\underset{\sim}{\pi}'\left((B(I_k - \phi_1^2 B)^{-1}\underset{\sim s}{\sigma}^2) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu\underset{\sim}{\iota})\right):
\end{aligned}$$

Comparing Proposition 2 to the results for the case with normal increments stated in Proposition 1, we see that the unconditional variance is increased by the persistence in the y_t process. This is no different from the usual autoregressive model without a Markov switching effect. Turning next to the skewness, a new term reflecting the expectation of the cross-product term $(\mu_{s_t} - \mu)(y_{t-1} - \mu_{s_{t-1}})^2$ enters into this expression. This term will be negative when low mean states tend to follow high

[6] The proposition requires that $(I_k - \phi_1 B)$ is invertible: Since $|\phi_1| < 1$ this will automatically be satisfied for all transition probability matrices since B has a single eigenvalue equal to unity and its remaining eigenvalues are less than one.

variance states and will otherwise be positive. To demonstrate this effect, Figure 5 plots skewness for the same parameter values as in Figure 2 and with $\phi_1 = 0.9$. A comparison of Figures 2 and 5 shows that introducing an autoregressive term into the model produces a very different profile of the skewness. Large negative skew is still obtained when p_{11} and p_{22} are both high so that the process does not switch very often and the negative term $E[(\mu_{s_t} - \mu)\sigma^2_{s_t}]$ dominates. For lower values of p_{11} and p_{22} the process switches more often and the positive term $E[(\mu_{s_t} - \mu)(y_{t-1} - \mu_{s_{t-1}})^2]$ flattens the skewness profile compared to Figure 2.

Three components change in the expression for the kurtosis once an autoregressive lag is introduced in the Markov switching process. First, the contribution to kurtosis from $\sigma^4_{s_t}\varepsilon^4_t$ is scaled by a factor $(I_k - \phi_1^4 B)^{-1}$. Furthermore, terms reflecting the expectation of $\left(y_{t-1} - \mu_{s_{t-1}}\right)^2$ times $(\mu_{s_t} - \mu)^2$ or $\sigma^2_{s_t}$ also contribute to kurtosis. Hence if high volatility states are followed by states with either high volatility or a mean parameter far away from the unconditional mean, this will tend to create fat tails and increase kurtosis. The highest kurtosis now occurs when both p_{11} and p_{22} are high but not too close to one so that rare jumps resulting from mean shifts also add to the tail mass.

Proposition 2 states results for a first-order autoregressive process whose moments are analytically tractable to derive. Similar analytical expressions for higher order autoregressive Markov switching processes contain the same types of components and are not very insightful to derive as they quickly become intractable. However, we still need to have a method for obtaining the moments of specific higher order processes that researchers have in mind. We demonstrate how to do this in the case of the second moment of such processes and note that the skewness and kurtosis can also be derived using similar techniques. In the general case with an arbitrary, but finite, number of autoregressive lags, the second moment of the Markov process can be derived in one of two ways. First, from (2) we have

$$\begin{aligned} E\left[(y_t - \mu)^2\right] &= E\left[(\mu_{s_t} - \mu)^2\right] + \sum_{j=1}^{q} \phi_j^2 E\left[(y_{t-j} - \mu_{s_{t-j}})^2\right] + E\left[\sigma^2_{s_t}\varepsilon^2_t\right] \\ &\quad + 2\sum_{i>j}^{q}\sum_{j=1}^{q} \phi_i\phi_j E\left[(y_{t-i} - \mu_{s_{t-i}})(y_{t-j} - \mu_{s_{t-j}})\right]. \end{aligned} \tag{10}$$

Applying the steady-state probabilities to the first and third terms we get

$$\begin{aligned} E\left[(y_t - \mu)^2\right] &= \underset{\sim}{\pi}'\left((\underset{\sim}{\mu}_s - \mu\underset{\sim}{\iota}) \odot (\underset{\sim}{\mu}_s - \mu\underset{\sim}{\iota})\right) + \sum_{j=1}^{q} \phi_j^2 E\left[(y_{t-j} - \mu_{s_{t-j}})^2\right] + \underset{\sim}{\pi}'\underset{\sim}{\sigma}^2_s \\ &\quad + 2\sum_{i>j}^{q}\sum_{j=1}^{q} \phi_i\phi_j E\left[(y_{t-i} - \mu_{s_{t-i}})(y_{t-j} - \mu_{s_{t-j}})\right]. \end{aligned} \tag{11}$$

This equation involves variances and covariances of the process relative to the state-specific means. To derive an expression for these terms, we exploit the associated Yule-Walker equations. Multiplying (2) through by $(y_{t-1} - \mu_{s_{t-1}}), \ldots, (y_{t-q} - \mu_{s_{t-q}})$, we get a system of equations

$$E\left[(y_t - \mu_{s_t})(y_{t-m} - \mu_{s_{t-m}})\right] = \sum_{i=1}^{q} \phi_i E\left[(y_{t-i} - \mu_{s_{t-i}})(y_{t-m} - \mu_{s_{t-m}})\right], \quad m = 1, \ldots, q. \tag{12}$$

An extra term $\underset{\sim}{\pi}'\underset{\sim}{\sigma}^2_s$ appears on the right hand side of (12) when $m = 0$. This gives $(q+1)$ equations that can used to jointly determine $E\left[(y_t - \mu_{s_t})^2\right]$;..., $E\left[(y_t - \mu_{s_t})(y_t - \mu_{s_{t-q}})\right]$. These terms can then be substituted into (11) to get the unconditional variance of the process around μ.

An alternative, and as far as we know, new method for obtaining second moments of Markov switching processes combines the companion form of the autoregressive model with the technique of expanding the state space of a Markov process proposed by Cox and Miller (1965). This is very convenient from a computational point of view and does not require solving a set of $q + 1$ equations in the autocovariances. We can always write (2) as

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ \vdots \\ y_{t-q+1} \end{bmatrix} = \begin{bmatrix} \mu_{s_t} \\ \mu_{s_{t-1}} \\ \mu_{s_{t-2}} \\ \vdots \\ \vdots \\ \mu_{s_{t-q+1}} \end{bmatrix} + \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \cdots & \phi_{q-1} & \phi_q \\ 1 & 0 & & & & \\ & 1 & & & 1 & \\ & & 1 & & & \\ & 0 & & 1 & & \\ & & & & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} - \mu_{s_{t-1}} \\ y_{t-2} - \mu_{s_{t-2}} \\ y_{t-3} - \mu_{s_{t-3}} \\ \vdots \\ \vdots \\ y_{t-q} - \mu_{s_{t-q}} \end{bmatrix} \tag{13}$$

$$+ \begin{bmatrix} \sigma_{s_t} & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & & & & \\ & & 0 & & 0 & \\ & & & 0 & & \\ & 0 & & & 0 & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \\ \vdots \\ \varepsilon_{t-q+1} \end{bmatrix};$$

or, in matrix form,

$$\underset{\sim t}{z} = \underset{\sim s_t^*}{\mu} + \Psi\left(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{\mu}\right) + \Sigma_{s_t^*} \underset{\sim t}{\varepsilon}; \tag{14}$$

where $\underset{\sim s_t^*}{\mu}$ now lies in the $k^* = k \cdot q$ dimensional state space formed as the Cartesian product $S_t \times S_{t-1} \times S_{t-q+1}$ of the original q state spaces.

Taking expectations of $\underset{\sim t}{z}$ around the unconditional mean vector we get the $q \times q$ covariance matrix

$$
\begin{aligned}
E\left[(\underset{\sim t}{z} - \underset{\sim}{¹¶})(\underset{\sim t}{z} - \underset{\sim}{¹¶})'\right] &= E\left[(\underset{\sim s_t^*}{¹} - \underset{\sim}{¹¶})(\underset{\sim s_t^*}{¹} - \underset{\sim}{¹¶})'\right] \\
&+ ª\, E\left[(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{¹})(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{¹})'\right] ª' \\
&+ E\left[\sum\nolimits_{s_t^*} \underset{\sim t \sim t}{{}^{2}\;{}^{2\prime}} \sum\nolimits_{s_t^*}'\right]; \qquad (15)
\end{aligned}
$$

where we used that $\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{¹} = \sum_{i=0}^{\infty} ª^i \sum_{s_{t-i-1}^*} \underset{\sim t-i-1}{{}^{2}}$ is uncorrelated with $(¹_{s_t^*} - ¹¶)$ and $\sum_{s_t^*} \underset{\sim t}{{}^{2}}$. Notice also from (14) that

$$
E\left[(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})'\right] = ª\, E\left[(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{¹})(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{¹})'\right] ª' + E\left[\sum\nolimits_{s_t^*} \underset{\sim t \sim t}{{}^{2}\;{}^{2\prime}} \sum\nolimits_{s_t^*}'\right]; \qquad (16)
$$

where $E\left[\sum_{s_t^*} \underset{\sim t \sim t}{{}^{2}\;{}^{2\prime}} \sum_{s_t^*}'\right]$ is a $q \times q$ matrix whose ¯rst element is $E\left[¾^2_{s_t}\right]$ with all other elements being zero. We refer to this matrix as V. It then follows that, provided $\underset{\sim t}{z}$ is a stationary process so $E\left[(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})'\right] = E\left[(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{¹})(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{¹})'\right]$, the q^2-vector of second moments centered around the state means is

$$
\begin{aligned}
\text{vec}\left(E\left[(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})'\right]\right) &= (ª \quad ª) \cdot \text{vec}\left(E\left[(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})'\right]\right) \\
&+ \text{vec}\left(E\left[\sum\nolimits_{s_t^*} \underset{\sim t \sim t}{{}^{2}\;{}^{2\prime}} \sum\nolimits_{s_t^*}'\right]\right); \qquad (17)
\end{aligned}
$$

so that

$$
\text{vec}\left(E\left[(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})(\underset{\sim t}{z} - \underset{\sim s_t^*}{¹})'\right]\right) = \left(I_{q^2} - (ª \quad ª)\right)^{-1} \text{vec}(V): \qquad (18)
$$

Substituting this back into (15) we get the q^2-vector of unconditional second moments:

$$
\begin{aligned}
\text{vec}\left(E\left[(\underset{\sim t}{z} - \underset{\sim}{¹¶})(\underset{\sim t}{z} - \underset{\sim}{¹¶})'\right]\right) &= \text{vec}\left(E\left[(\underset{\sim s_t^*}{¹} - \underset{\sim}{¹¶})(\underset{\sim s_t^*}{¹} - \underset{\sim}{¹¶})'\right]\right) \\
&+ \left(I_{q^2} + (ª \quad ª)\left(I_{q^2} - (ª \quad ª)\right)^{-1}\right) \text{vec}(V): \qquad (19)
\end{aligned}
$$

This expression is very convenient for computation of the variance of higher order autoregressive Markov switching models. It also allows calculation of autocovariances from the cross-product terms. For example, $E\left[(y_t - ¹)(y_{t-1} - ¹)\right]$ will be the second element of the vector in (19).

4. State-dependent Autoregressive Dynamics

Again we initially consider the ¯rst-order autoregressive model and then demonstrate how to generalize the results to the less tractable, but similar, general case. The underlying model (which we shall refer to as MSIII) is

$$y_t = {}^1{}_{s_t} + Á_{1s_{t-1}}(y_{t-1} - {}^1{}_{s_{t-1}}) + ¾_{s_t}{}^2{}_t; \quad \left|Á_{s_{t-1}}\right| < 1; \quad {}^2{}_t \sim IIN(0;1): \tag{20}$$

Backwards substitution gives

$$y_t - {}^1{}_{s_t} = \sum_{l=1}^{\infty}\left(\prod_{j=1}^{l} Á_{1s_{t-j}} ¾_{s_{t-j}}\right) {}^2{}_{t-l} + ¾_{s_t}{}^2{}_t; \tag{21}$$

so that $E[y_t - {}^1{}_{s_t}] = 0$ and $E[y_t] = \underset{\sim}{¼}'\,\underset{\sim}{^1}_s = {}^1$. To state parsimoniously the moments of this process, it is convenient to de¯ne the $(k \times k)$ diagonal matrix of state-speci¯c autoregressive coe±cients

$$© = \begin{bmatrix} Á_{11} & 0 & & \\ 0 & Á_{12} & : & \\ & : & : & \\ & & & Á_{1k} \end{bmatrix};$$

where $Á_{1r}$ is the ¯rst order autoregressive coe±cient in state r. Proposition 3 provides the second to fourth moments of this process:

Proposition 3

Consider the ¯rst-order Gaussian Markov switching process with a state-dependent autoregressive term

$$y_t = {}^1{}_{s_t} + Á_{1s_{t-1}}\left(y_{t-1} - {}^1{}_{s_{t-1}}\right) + ¾_{s_t}{}^2{}_t; \qquad {}^2{}_t \sim IIN(0;1); \qquad \left|Á_{s_{t-1}}\right| < 1;$$

and suppose that the process started from its steady state. The variance, skewness and kurtosis of y_t are given by

$$\begin{aligned}
E\left[(y_t - {}^1)^2\right] &= \underset{\sim}{¼}'\left((\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶}) \odot (\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶}) + ©^2(I_k - B©^2)^{-1}\,\underset{\sim_s}{¾^2} + \underset{\sim_s}{¾^2}\right)\\
E\left[(y_t - {}^1)^3\right] &= \underset{\sim}{¼}'\left((\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶}) \odot (\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶}) \odot (\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶})\right)\\
&\quad + 3\,\underset{\sim}{¼}'\left((B©^2(I_k - B©^2)^{-1}\,\underset{\sim_s}{¾^2}) \odot (\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶})\right) + 3\,\underset{\sim}{¼}'\left((\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶}) \odot \underset{\sim_s}{¾^2}\right)\\
E\left[(y_t - {}^1)^4\right] &= \underset{\sim}{¼}'\left((\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶}) \odot (\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶}) \odot (\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶}) \odot (\underset{\sim_s}{^1} - {}^1\,\underset{\sim}{¶})\right)
\end{aligned}$$

$$
\begin{aligned}
&+6\,\boldsymbol{\pi}'\left(\left(B\Phi^2(I_k - B\Phi^2)^{-1}\boldsymbol{\sigma}_s^2\right)\odot(\boldsymbol{\mu}_s - \mu\boldsymbol{\iota})\odot(\boldsymbol{\mu}_s - \mu\boldsymbol{\iota})\right)\\
&+6\,\boldsymbol{\pi}'\left((\boldsymbol{\mu}_s - \mu\boldsymbol{\iota})\odot(\boldsymbol{\mu}_s - \mu\boldsymbol{\iota})\odot\boldsymbol{\sigma}_s^2\right)\\
&+\boldsymbol{\pi}'\,\Phi^4(I_k - B\Phi^4)^{-1}\left(3\,\boldsymbol{\sigma}_s^4 + 6\left(B\Phi^2(I_k - B\Phi^2)^{-1}\boldsymbol{\sigma}_s^2\right)\odot\boldsymbol{\sigma}_s^2\right)\\
&+6\,\boldsymbol{\pi}'\left(\left(B\Phi^2(I_k - B\Phi^2)^{-1}\boldsymbol{\sigma}_s^2\right)\odot\boldsymbol{\sigma}_s^2\right) + 3\,\boldsymbol{\pi}'\boldsymbol{\sigma}_s^4 .
\end{aligned}
$$

To demonstrate the effect on the coefficients of skewness and kurtosis of having different autoregressive parameters in the different states, Figure 6 keeps the parameters from Figure 5 but sets $\phi_{11} = 0.99$ and $\phi_{12} = 0.81$, so that, compared with the choice of ϕ_1 in Figure 5, the serial correlation is now 0.09 higher in state 1 and 0.09 lower in state 2. The skewness is now positive for high values of p_{11} and p_{22} and is otherwise negative.[7] The kurtosis still peaks when p_{11} and p_{22} are large but is now flat on a larger part of the grid. Letting the autoregressive coefficients differ between the two states can significantly change theeisaneignteresantrgy or, in matrix form,

$$\underset{\sim t}{z} = \underset{\sim s_t^*}{\mu} + \Phi_{s_{t-1}^*}\left(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{\mu}\right) + \Sigma_{s_t^*}\underset{\sim t}{\varepsilon}. \tag{23}$$

To analyze this case we first derive an expression for $E[(\underset{\sim t}{z} - \underset{\sim s_t^*}{\mu})(\underset{\sim t}{z} - \underset{\sim s_t^*}{\mu})']$. Let $\mathrm{vec}\left(E[(\underset{\sim t}{z} - \underset{\sim}{\mu}_{s_t^*})^2 \mid S_t^* = i]\right)$ be short-hand notation for the q^2-vector of expectations of $(\underset{\sim t}{z} - \underset{\sim}{\mu}_{s_t^*})(\underset{\sim t}{z} - \underset{\sim}{\mu}_{s_t^*})'$ conditional on $S_t^* = i$; let $V_i \equiv \mathrm{vec}\left(E[\Sigma_{s_t^*}\underset{\sim t}{\varepsilon}\underset{\sim t}{\varepsilon}'\Sigma_{s_t^*}' \mid S_t^* = i]\right)$ be a q^2 vector of conditional expectations of the residual term and let Φ_i be the qxq matrix $\Phi_{s_{t-1}^*}$ conditional on $S_{t-1}^* = i$. As in Section 3, the dimension of the expanded state space is $k^* = kq$. Stacking the moment conditions resulting from (23), we get the k^*q^2-vector

$$\begin{bmatrix} \mathrm{vec}\left(E[(\underset{\sim t}{z} - \underset{\sim}{\mu}_{s_t^*})^2 \mid S_t^* = 1]\right) \\ \mathrm{vec}\left(E[(\underset{\sim t}{z} - \underset{\sim}{\mu}_{s_t^*})^2 \mid S_t^* = 2]\right) \\ \vdots \\ \mathrm{vec}\left(E[(\underset{\sim t}{z} - \underset{\sim}{\mu}_{s_t^*})^2 \mid S_t^* = k^*]\right) \end{bmatrix}$$

$$= \left(B^* \otimes I_{q^2}\right)\begin{bmatrix} (\Phi_1 \otimes \Phi_1)\,\mathrm{vec}\left(E[(\underset{\sim t-1}{z} - \underset{\sim}{\mu}_{s_{t-1}^*})^2 \mid S_{t-1}^* = 1]\right) \\ (\Phi_2 \otimes \Phi_2)\,\mathrm{vec}\left(E[(\underset{\sim t-1}{z} - \underset{\sim}{\mu}_{s_{t-1}^*})^2 \mid S_{t-1}^* = 2]\right) \\ \vdots \\ (\Phi_{k^*} \otimes \Phi_{k^*})\,\mathrm{vec}\left(E[(\underset{\sim t-1}{z} - \underset{\sim}{\mu}_{s_{t-1}^*})^2 \mid S_{t-1}^* = k^*]\right) \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_{k^*} \end{bmatrix} \tag{24}$$

$$= \left(B^* \otimes I_{q^2}\right)\widehat{\Phi}\begin{bmatrix} \mathrm{vec}\left(E\left[(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{\mu})^2 \mid S_{t-1}^* = 1\right]\right) \\ \mathrm{vec}\left(E\left[(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{\mu})^2 \mid S_{t-1}^* = 2\right]\right) \\ \vdots \\ \mathrm{vec}\left(E\left[(\underset{\sim t-1}{z} - \underset{\sim s_{t-1}^*}{\mu})^2 \mid S_{t-1}^* = k^*\right]\right) \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_{k^*} \end{bmatrix};$$

where B^* is the $(k^* \times k^*)$ matrix of backward transition probabilities while $\widehat{\Phi}$ is the $(k^*q^2) \times (k^*q^2)$ block-diagonal matrix formed as

$$\widehat{\Phi} = \begin{bmatrix} \Phi_1 \otimes \Phi_1 & & & 0 \\ & \Phi_2 \otimes \Phi_2 & & \\ 0 & & \ddots & \\ & & & \Phi_{k^*} \otimes \Phi_{k^*} \end{bmatrix}.$$

Under stationarity of the y_t process we thus have

$$\begin{bmatrix} \text{vec}\left(E[(\underset{\sim t}{z} - \underset{\sim}{\text{¶}}\ {}^{1}{}_{s_t^*})^2 \mid S_t^* = 1]\right) \\ \text{vec}\left(E[(\underset{\sim t}{z} - \underset{\sim}{\text{¶}}\ {}^{1}{}_{s_t^*})^2 \mid S_t^* = 2]\right) \\ ::: \\ \text{vec}\left(E[(\underset{\sim t}{z} - \underset{\sim}{\text{¶}}\ {}^{1}{}_{s_t^*})^2 \mid S_t^* = k^*]\right) \end{bmatrix} = \left(I_m - \left(B^* \quad I_{q^2}\right)\widehat{\text{a}}\right)^{-1} \begin{bmatrix} V_1 \\ V_2 \\ ::: \\ V_{k^*} \end{bmatrix}; \qquad (25)$$

where $m = k^*q^2$. The q^2-vector of unconditional second moments centered around the state means can now be extracted from (25) by pre-multiplying by a $(q^2 \times q^2k^*)$ matrix ¤ given by

$$\text{¤} = \left((\underset{\sim}{\text{¼}}^*)' \quad I_{q^2}\right) = \begin{bmatrix} \text{¼}_1 & 0 & ::: & \text{¼}_2 & 0 & 0 & ::: & \text{¼}_{k^*} & ::: & 0 \\ 0 & \text{¼}_1 & 0 & ::: & \text{¼}_2 & 0 & ::: & 0 & ::: & 0 \\ ::: & ::: & ::: & ::: & ::: & ::: & ::: & ::: & ::: & ::: \\ 0 & ::: & \text{¼}_1 & 0 & ::: & \text{¼}_2 & ::: & 0 & ::: & \text{¼}_{k^*} \end{bmatrix};$$

so that

$$E\left[(\underset{\sim t}{z} - \underset{\sim s_t^*}{{}^{1}})(\underset{\sim t}{z} - \underset{\sim s_t^*}{{}^{1}})'\right] = \text{¤}\left(I_m - \left(B^* \quad I_{q^2}\right)\widehat{\text{a}}\right)^{-1}(V_1\ V_2 ::: V_{k^*})': \qquad (26)$$

Finally the unconditional variance of $\underset{\sim t}{z}$ follows from (23) as

$$\begin{aligned} \text{vec}\left(E[(\underset{\sim t}{z} - {}^{1}\underset{\sim}{\text{¶}})(\underset{\sim t}{z} - {}^{1}\underset{\sim}{\text{¶}})']\right) &= \text{vec}\left(E[(\underset{\sim s_t^*}{{}^{1}} - {}^{1}\underset{\sim}{\text{¶}})(\underset{\sim s_t^*}{{}^{1}} - {}^{1}\underset{\sim}{\text{¶}})']\right) \\ &\quad + \text{¤} \cdot \widehat{\text{a}}\left(I_m - \left(B^* \quad I_{q^2}\right)\widehat{\text{a}}\right)^{-1} \begin{bmatrix} V_1 \\ V_2 \\ : \\ V_{k^*} \end{bmatrix} + \text{vec}(V); \qquad (27) \end{aligned}$$

where $V = E[\sum_{s_t^*} \underset{\sim t \sim t}{{}^{2}\,{}^{2\prime}} \sum'_{s_t^*}]$. Again these expressions are easy to implement, do not require solving a set of Yule-Walker equations, and should prove useful to researchers wanting to charac

Proposition 4

The autocovariance functions of the stationary Gaussian Markov switching processes MSI, MSII and MSIII starting from their steady state are given by

$$E\left[(y_{t+n} - \mu)(y_t - \mu)\right] =$$

$$\underset{\sim}{\pi}' \left((B^n(\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota})) \odot (\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota}) \right) \qquad \text{(MSI)}$$

$$\underset{\sim}{\pi}' \left((B^n(\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota})) \odot (\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota}) \right) + \phi_1^n \underset{\sim}{\pi}' (I_k - \phi_1^2 B)^{-1} \underset{\sim s}{\sigma}^2 \qquad \text{(MSII)}$$

$$\underset{\sim}{\pi}' \left((B^n(\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota})) \odot (\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota}) \right) + \underset{\sim}{\pi}' \, \Gamma_n (I_k - B\Phi^2)^{-1} \underset{\sim s}{\sigma}^2 \qquad \text{(MSIII)}$$

where Γ_n is the $(k \times k)$ diagonal matrix

$$\Gamma_n = \begin{bmatrix} \gamma_{1,n} & 0 & & \\ 0 & \gamma_{2,n} & \vdots & \\ & \vdots & \vdots & \\ & & & \gamma_{k,n} \end{bmatrix}$$

with diagonal elements $\gamma_{r,n} = E\left[\left(\prod_{i=0}^{n-1} \phi_{1s_{t+i}}\right) \mid S_t = r\right]$, so that $\Gamma_1 = \Phi$.

Notice that we can compute the elements of Γ_n as

$$\gamma_{r,n} = \phi_{1r} \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} \ldots \sum_{j_{n-1}=1}^{k} P_{rj_1} P_{j_1 j_2} \ldots P_{j_{n-2} j_{n-1}} \phi_{j_1} \phi_{j_s} \ldots \phi_{j_{n-1}}$$

which is the product of the probabilities and the associated autoregressive coefficients on paths emanating from state r. For higher order autocovariances, the expression for Γ_n will be quite complex since it reflects the entire set of possible paths followed by the process over the n-period horizon.

Again the basic Markov switching model with two states provides some intuition for the result since it allows us to considerably simplify the autocovariance function:

$$E[(y_t - \mu)(y_{t-n} - \mu)] = \pi_1(1 - \pi_1)(\mu_1 - \mu_2)^2 \text{vec}(P^n)' v_1; \qquad (28)$$

where $v_1 = ((1 - \pi_1); -(1 - \pi_1); -\pi_1; \pi_1)'$. Of particular interest is the first-order autocovariance for the levels of the process which is given by $(\mu_1 - \mu_2)^2 \pi_1 (1 - \pi_1)(p_{11} + p_{22} - 1)$. This expression

MSI: $y_t = \mu_{s_t} + \sigma_{s_t} \varepsilon_t; \quad \varepsilon_t \sim IIN(0;1);$

MSII: $y_t = \mu_{s_t} + \phi_1 (y_{t-1} - \mu_{s_{t-1}}) + \sigma_{s_t} \varepsilon_t; \quad \varepsilon_t \sim IIN(0;1); \quad |\phi_1| < 1;$

MSIII: $y_t = \mu_{s_t} + \phi_{1s_{t-1}} (y_{t-1} - \mu_{s_{t-1}}) + \sigma_{s_t} \varepsilon_t; \quad \varepsilon_t \sim IIN(0;1); \quad |\phi_{s_{t-1}}| < 1.$

has an intuitive interpretation since the autocovariance will be positive provided the presence of the process in the two states is persistent ($p_{11} + p_{22} > 1$), otherwise a negative autocovariance will result. Without a Markov switching effect ($p_{11} = 1 - p_{22}$) there will be no first order autocorrelation in the process.

Many studies find significant serial correlation in the squared values of economic time series. The success of ARCH models in empirical work can be explained by the fact that these models can generate such autocorrelation patterns. We show in the following Proposition that Markov switching models can also give rise to autocorrelation in the squares of a time series:

Proposition 5

The autocovariance function of the squared values of the stationary Markov switching process starting from its steady state is given by

$$
\begin{aligned}
& \mathrm{E}\left[(y_t^2 - \mathrm{E}[y_t^2])(y_{t-n}^2 - \mathrm{E}[y_{t-n}^2])\right] \\
= \; & \underset{\sim}{\pi}' \left((B^n(\underset{\sim}{\sigma}_s^2 + \underset{\sim}{\mu}_s^2)) \odot (\underset{\sim}{\mu}_s^2 + \underset{\sim}{\sigma}_s^2) \right) - (\mathrm{E}[y_t^2])^2
\end{aligned}
\qquad \text{(MSI)}
$$

$$
\begin{aligned}
= \; & \underset{\sim}{\pi}' \left((B^n (I_k - \phi_1^2 B)^{-1} \underset{\sim}{\sigma}_s^2) \odot \underset{\sim}{\mu}_s^2 \right) + \underset{\sim}{\pi}' \left((B^n \underset{\sim}{\mu}_s^2) \odot \underset{\sim}{\mu}_s^2 \right) \\
& + \phi_1^{2n} \underset{\sim}{\pi}' (I_k - \phi_1^4 B)^{-1} \left(3 \underset{\sim}{\sigma}_s^4 + 6 \phi_1^2 \left((B (I_k - \phi_1^2 B)^{-1} \underset{\sim}{\sigma}_s^2) \odot \underset{\sim}{\sigma}_s^2 \right) \right) \\
& + \phi_1^{2n} \underset{\sim}{\pi}' \left(((I_k - \phi_1^2 B)^{-1} \underset{\sim}{\sigma}_s^2) \odot \underset{\sim}{\mu}_s^2 \right) \\
& + \underset{\sim}{\pi}' \left(\sum_{i=1}^{n} \phi_1^{2(n-i)} \left(B^i (I_k - \phi_1^2 B)^{-1} \underset{\sim}{\sigma}_s^2 + B^i \underset{\sim}{\mu}_s^2 \right) \odot \underset{\sim}{\sigma}_s^2 \right) \\
& + 4 \phi_1^n \underset{\sim}{\pi}' \left(B^n (((I_k - \phi_1^2 B)^{-1} \underset{\sim}{\sigma}_s^2) \odot \underset{\sim}{\mu}_s)) \odot \underset{\sim}{\mu}_s \right) - \left(\mathrm{E}[y_t^2] \right)^2
\end{aligned}
\qquad \text{(MSII)}
$$

$$
\begin{aligned}
= \; & \underset{\sim}{\pi}' \left((B^n (I_k - B\Phi^2)^{-1} \underset{\sim}{\sigma}_s^2) \odot \underset{\sim}{\mu}_s^2 \right) + \underset{\sim}{\pi}' \left((B^n \underset{\sim}{\mu}_s^2) \odot \underset{\sim}{\mu}_s^2 \right) \\
& + \underset{\sim}{\pi}' \, \Upsilon_{2,n} (I_k - B\Phi^4)^{-1} \left(3 \underset{\sim}{\sigma}_s^4 + 6 (B\Phi^2 (I_k - B\Phi^2)^{-1} \underset{\sim}{\sigma}_s^2) \odot \underset{\sim}{\sigma}_s^2 \right) \\
& + \underset{\sim}{\pi}' \left((\Upsilon_{2,n} (I_k - B\Phi^2)^{-1} \underset{\sim}{\sigma}_s^2) \odot \underset{\sim}{\mu}_s^2 \right) + \underset{\sim}{\pi}' \left(\sum_{i=1}^{n} (\Upsilon_i (I_k - B\Phi^2)^{-1} \underset{\sim}{\sigma}_s^2 + \Upsilon_i \underset{\sim}{\mu}_s^2) \right) \\
& + 4 \underset{\sim}{\pi}' \left(\Gamma_n (((I_k - B\Phi^2)^{-1} \underset{\sim}{\sigma}_s^2) \odot \underset{\sim}{\mu}_s \right) - \left(\mathrm{E}[y_t^2] \right)^2 ;
\end{aligned}
\qquad \text{(MSIII)}
$$

where $\Upsilon_{2,n}$ is a diagonal $k \times k$ matrix whose r'th diagonal element is given by $\Upsilon_{2,n}[r;r] = E\left[\left(\prod_{i=0}^{n-1} \phi^2_{1s_{t+i}}\right) \mid S_t = r\right]$ Υ_i is a diagonal matrix with elements $\Upsilon_i[r;r] = E\left[\left(\prod_{j=i}^{n-1} \phi^2_{1s_{t+j}}\right) \sigma^2_{s_{t+i}} \mid S_t = r\right]$ and Γ_n is a diagonal matrix with $\Gamma_n[r;r] = E\left[\left(\prod_{i=0}^{n-1} \phi_{1s_{t+i}}\right) \mu_{s_{t+n}} \mid S_t = r\right]$.

Proofs of Propositions 4 and 5 are provided in the Appendix. To compute the autocorrelation, the autocovariances need to be scaled by the variance of y_t^2, $E[(y_t^2 - E[y_t^2])^2]$. For the three models, this can be shown to be given by

$$\begin{aligned} E[(y_t^2 - E[y_t^2])^2] &= E[y_t^4] - (E[y_t^2])^2 \\ &= \underset{\sim}{\pi}'\underset{\sim s}{\mu}^4 + 6\underset{\sim}{\pi}'(\underset{\sim s}{\mu} \odot \underset{\sim s}{\mu} \odot \underset{\sim s}{\sigma}^2) + 3\underset{\sim}{\pi}'\underset{\sim s}{\sigma}^4 - \left(\mathrm{var}(y_t) + E[y_t]^2\right)^2 \end{aligned} \qquad \text{(MSI)}$$

$$\begin{aligned} = {} & \underset{\sim}{\pi}'\underset{\sim s}{\mu}^4 + 6\phi_1^2\underset{\sim}{\pi}'\left((B(I_k - \phi_1^2 B)^{-1}\underset{\sim s}{\sigma}^2) \odot \underset{\sim s}{\mu} \odot \underset{\sim s}{\mu}\right) + 6\underset{\sim}{\pi}'(\underset{\sim s}{\mu}^2 \odot \underset{\sim s}{\sigma}^2) \\ & + \underset{\sim}{\pi}'(I_k - \phi_1^4 B)^{-1}\left(3\underset{\sim s}{\sigma}^4 + 6\phi_1^2(B(I_k - \phi_1^2 B)^{-1}\underset{\sim s}{\sigma}^2) \odot \underset{\sim s}{\sigma}^2\right) - \left(\mathrm{var}(y_t) + E[y_t]^2\right)^2 \end{aligned} \qquad \text{(MSII)}$$

$$\begin{aligned} = {} & \underset{\sim}{\pi}'\underset{\sim s}{\mu}^4 + 6\underset{\sim}{\pi}'\left((B\Phi^2(I_k - B\Phi^2)^{-1}\underset{\sim s}{\sigma}^2) \odot (\underset{\sim s}{\mu}^2 + \underset{\sim s}{\sigma}^2)\right) + 6\underset{\sim}{\pi}'(\underset{\sim s}{\mu}^2 \odot \underset{\sim s}{\sigma}^2) \\ & + \underset{\sim}{\pi}'\Phi^4(I_k - B\Phi^4)\left(3\underset{\sim s}{\sigma}^4 + 6(B\Phi^2(I_k - B\Phi^2)^{-1}\underset{\sim s}{\sigma}^2) \odot \underset{\sim s}{\sigma}^2\right) \\ & + 3\underset{\sim}{\pi}'\underset{\sim s}{\sigma}^4 - \left(\mathrm{var}(y_t) + E[y_t]^2\right)^2 \end{aligned} \qquad \text{(MSIII)}$$

Intuition is provided by the autocovariance of y_t^2 in the basic two-state model which is given by

$$E\left[(y_t^2 - E[y_t^2])(y_{t-n}^2 - E[y_{t-n}^2])\right] = \pi_1(1-\pi_1)(\mu_2^2 - \mu_1^2 + \sigma_2^2 - \sigma_1^2)^2 \mathrm{vec}(P^n)'v_1. \qquad (30)$$

Hence the first-order autocovariance is $E[(y_t^2 - E[y_t^2])(y_{t-1}^2 - E[y_{t-1}^2])] = \pi_1(1-\pi_1)(\mu_2^2 - \mu_1^2 + \sigma_2^2 - \sigma_1^2)^2$ $(p_{11} + p_{22} - 1)$. Again this expression has the intuitive interpretation that there will be positive autocorrelation in y_t^2 if the state mixing process is persistent ($p_{11} + p_{22} > 1$), otherwise the squared values of the process will be negatively autocorrelated. If $p_{11} = 1 - p_{22}$, there will be no first-order serial correlation in the squared values of the series. If there is no switching between the two regimes, i.e. $\pi_1(1-\pi_1) = 0$, there will be no autocorrelation in the squared values of the series.

Figure 7 plots the first order autocorrelation in the squared levels of a Markov switching process for the case where $\underset{\sim s}{\mu} = (1;1)$ and $\underset{\sim s}{\sigma} = (2;4)$. Since the two mean parameters are identical across states, there is no serial correlation in the level of the series. Reflecting the different volatility parameters, quite high serial correlation in the squares of the series is obtained, however, when the persistence of the two states is high.

6. Discussion of Results

Our results on the moments of the regime switching models show that the Markovian dependency in the mixture probabilities significantly expands the scope for asymmetry and fat tails that can be generated by time-independent mixture models. This may help to explain the relative success that these models have had in applications to economic time series that clearly display such features.

Our theoretical results are also closely related to the literature on structural breaks and persistence. Structural breaks are usually thought of as one-off events that introduce non-stationarities in the data. In the context of a Markov switching model a structural break can be modeled as follows. Consider the following partitioned matrix of transition probabilities:

$$P = \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix};$$

where, say, P_1 is a $k_1 \times k_1$ matrix, P_4 is a $k_2 \times k_2$ matrix and $k_1 + k_2 = k$. If some element of P_2 is nonzero while the elements of P_3 all equal zero, then once a state ordered $k_1 + 1$ or higher is reached then the process can never return to a state ordered k_1 or lower, so this event represents a break from a switching model that mixes one block of states to another model mixing a different block of states. If the values of P_2 and P_3 are close to, but not all equal to, zero then there exists a stationary 'hyper model' that includes both blocks of states, but it will be very difficult in any finite sample to distinguish between a structural break and infrequently occuring regime switches.

The parameter values for which the Markov switching models seem best capable of generating volatility clustering and autocorrelation, i.e. those values representing infrequent mixing of regimes with quite different mean and variance parameters, are also those for which the switching models most closely resemble structural break models. Thus our findings are clearly related to the literature on structural breaks and persistence in the first and second moments of time series. Perron (1989) considers a single exogenous shift in the level or slope of the trend function of a time series while Perron (1990) investigates the case of an exogenous break in the mean level. This latter case closely corresponds to a Markov switching model with $k = 2$, $\sigma_1 = \sigma_2$, and $\mu_1 \neq \mu_2$. In both cases such a break in the time series is found to lead to higher rates of non-rejection of the null of a unit root since it introduces persistence in the series centered around the sample mean. As the size of the mean shift gets larger, the estimated autoregressive parameter in a model without serial correlation but with a single shift in the mean gets closer to one.[9]

We carried out an experiment to show that these studies are closely related to Proposition 4. Setting $\sigma_1^2 = \sigma_2^2 = 1$, $p_{11} = p_{22} = 0.99$, $\mu_1 = 1$, and varying μ_2 we obtained the following population moments for the first-order autocorrelation of the process:

First order autocorrelation

value of μ_2	MSI	MSII ($\phi_1 = 0.9$)
2	0.196	0.904
3	0.495	0.916
4	0.688	0.930

Even when there is no autocorrelation in any of the states (as for MSI), a shift in the mean can still produce very substantial persistence around the unconditional mean, and the effect is an increasing function of the size of the shift. As the shift gets large, and consistent with Perron (1990), the autocorrelation coefficient for this case with rare shift gets closer to one.

The proposition that persistence in the squared level of a Markov switching process can be the result of rare, large shifts in the unconditional variance is also related to earlier empirical findings. Lamoureux and Lastrapes (1990) study second moments and find that the existence of deterministic structural shifts in the unconditional variance, when not accounted for, will increase the persistence of squared residuals.[10] They show that the resulting upward bias in GARCH estimates of persistence of variance can be quite substantial. For a number of individual stock return series once a structural break in the intercept term in the conditional variance equation is accounted for, the estimated persistence declines from an average of 0.98 to an average of 0.82. Indeed, Lamoureux and Lastrapes propose to use Markov switching models as a way of handling misspecification problems due to occasional shifts in the conditional variance. Likewise, Diebold (1986) suggests that the very high persistence in the conditional variance observed in GARCH models may reflect a failure to include dummies for shifts in the intercept in the variance equation caused by exogenous shocks such as monetary policy regime changes.[11]

[9] In related work Hendry and Neale (1991) use Monte Carlo methods to quantify the effect that the introduction of shifts in the intercept of an autoregressive process has on the loss in the power of standard unit root tests.

[10] Yet another possibility is that a misspecified mean leads to overrejection of the null of no conditional heteroskedasticiy as recently reported by Lumsdaine and Ng (1997).

[11] In principle ARCH and Markov switching effects could be combined to produce a highly flexible, nonlinear mixture

Appendix

This appendix contains the proofs of the propositions and the corollary.

Proof of Proposition 1

From the law of iterated expectations we have

$$\begin{aligned} E\left[(y_t - \mu)^n\right] &= E\left[E[(y_t - \mu)^n \mid S_t]\right] = \sum_{i=1}^{k} \pi_i E\left[(\mu_i + \sigma_i \epsilon_t - \mu)^n\right] && \text{(A1)} \\ &= \sum_{i=1}^{k} \pi_i \sum_{j=0}^{n} {}_nC_j \; \sigma_i^j \; (\mu_i - \mu)^{n-j} E[\varepsilon_t^j] \end{aligned}$$

where we used Newton's binomial formula and the assumption that the steady-state probabilities apply. The expressions for the cases where ϵ_t follows a t-distribution or a normal distribution are based on the moment-generating functions for these distributions. For example, from the moment generating function of the normal distribution we have that $E[\epsilon_t^j] = 0$, if j is odd and

$$E[\epsilon_t^j] = \prod_{h=1}^{j/2} (2h - 1) \equiv b_j;$$

otherwise. Substituting this expression into (A1) we get the result.

Proof of Corollary 1

First consider the variance. From the law of iterated expectations we have

$$\begin{aligned} E\left[(y_t - \mu)^2\right] &= E\left[E[(y_t - \mu)^2 \mid S_t]\right] \\ &= \pi_1 E\left[(\mu_1 + \sigma_1 \epsilon_t - \mu)^2\right] + (1 - \pi_1) E\left[(\mu_2 + \sigma_2 \epsilon_t - \mu)^2\right] \\ &= (1 - \pi_1)\sigma_2^2 + \pi_1 \sigma_1^2 + (1 - \pi_1)(\mu_2 - \mu)^2 + \pi_1(\mu_1 - \mu)^2 && \text{(A2)} \\ &= (1 - \pi_1)\sigma_2^2 + \pi_1 \sigma_1^2 + \pi_1(1 - \pi_1)(\mu_2 - \mu_1)^2: \end{aligned}$$

We next compute the skewness and kurtosis of this model:

$$\begin{aligned} E\left[(y_t - \mu)^3\right] &= E\left[E[(y_t - \mu)^3 \mid S_t]\right] \\ &= (1 - \pi_1) E\left[(\mu_2 + \sigma_2 \epsilon_t - \mu)^3\right] + \pi_1 E\left[(\mu_1 + \sigma_1 \epsilon_t - \mu)^3\right] \\ &= \pi_1(1 - \pi_1)(\mu_1 - \mu_2)\left\{3(\sigma_1^2 - \sigma_2^2) + (1 - 2\pi_1)(\mu_2 - \mu_1)^2\right\}; && \text{(A3)} \end{aligned}$$

model as proposed by Hamilton and Susmel (1994). Our results can explain why such an extension may be necessary because the Markov switching model only appears to be able to generate limited persistence in the squared values of a time series.

and hence the coefficient of skewness is given by

$$\frac{E[(y_t-\mu)^3]}{(E[(y_t-\mu)^2])^{3/2}}=\frac{\pi_1(1-\pi_1)(\mu_1-\mu_2)\left\{3(\sigma_1^2-\sigma_2^2)+(1-2\pi_1)(\mu_2-\mu_1)^2\right\}}{((1-\pi_1)\sigma_2^2+\pi_1\sigma_1^2+\pi_1(1-\pi_1)(\mu_2-\mu_1)^2)^{3/2}}. \qquad (A4)$$

To compute the coefficient of kurtosis we proceed as follows

$$\begin{aligned} E\left[(y_t-\mu)^4\right] &= E\left[E[(y_t-\mu)^4 \mid S_t]\right] \\ &= (1-\pi_1)E\left[(\mu_2+\sigma_2\varepsilon_t-\mu)^4\right]+\pi_1 E\left[(\mu_1+\sigma_1\varepsilon_t-\mu)^4\right] \\ &= (1-\pi_1)\left\{3\sigma_2^4+(\mu_2-\mu)^4+6\sigma_2^2(\mu_2-\mu)^2\right\} \qquad (A5)\\ &\quad +\pi_1\left\{3\sigma_1^4+(\mu_1-\mu)^4+6\sigma_1^2(\mu_1-\mu)^2\right\}. \end{aligned}$$

Simplifying the expression by using that $\mu=\pi_1\mu_1+(1-\pi_1)\mu_2$, we get the coefficient of excess kurtosis

$$\frac{E[(y_t-\mu)^4]-3(E[(y_t-\mu)^2])^2}{(E[(y_t-\mu)^2])^2}\equiv\frac{a}{b}$$

$$\begin{aligned} a &= 3\pi_1(1-\pi_1)(\sigma_2^2-\sigma_1^2)^2+6(\mu_2-\mu_1)^2\pi_1(1-\pi_1)(2\pi_1-1)(\sigma_2^2-\sigma_1^2) \\ &\quad +\pi_1(1-\pi_1)(\mu_2-\mu_1)^4(1-6\pi_1(1-\pi_1)) \\ b &= \left((1-\pi_1)\sigma_2^2+\pi_1\sigma_1^2+\pi_1(1-\pi_1)(\mu_2-\mu_1)^2\right)^2. \qquad (A6) \end{aligned}$$

Proof of Proposition 2

Squaring y_t around its unconditional mean and taking expectations we have

$$\begin{aligned} E\left[(y_t-\mu)^2\right] &= E\left[(\mu_{s_t}-\mu)^2+\phi_1^2(y_{t-1}-\mu_{s_{t-1}})^2+\sigma_{s_t}^2\varepsilon_t^2\right] \\ &\quad +2\phi_1 Cov(\mu_{s_t}-\mu; y_{t-1}-\mu_{s_{t-1}})+2Cov(\mu_{s_t}-\mu;\sigma_{s_t}\varepsilon_t) \qquad (A7)\\ &\quad +2\phi_1 Cov(y_{t-1}-\mu_{s_{t-1}};\sigma_{s_t}\varepsilon_t). \end{aligned}$$

From (7), the assumption that ε_t is iid and the independence at all leads and lags between ε_t and S_t, it follows that all three covariance terms are equal to zero so that

$$\begin{aligned} E\left[(y_t-\mu)^2\right] &= \underset{\sim}{\pi}' E\left[(\underset{\sim t}{y}-\mu\underset{\sim}{\iota})\odot(\underset{\sim t}{y}-\mu\underset{\sim}{\iota})\,|\underset{\sim t}{S}\right] \\ &= \underset{\sim}{\pi}'(\underset{\sim s}{\mu}-\mu\underset{\sim}{\iota})\odot(\underset{\sim s}{\mu}-\mu\underset{\sim}{\iota})+\underset{\sim}{\pi}'\phi_1^2\left(\sum_{i=0}^{\infty}\phi_1^{2i}P^i\underset{\sim s}{\sigma^2}\right)+\underset{\sim}{\pi}'\underset{\sim s}{\sigma^2} \end{aligned}$$

$$= \underset{\sim}{¼}' \left((\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) + Á_1^2(1 - Á_1^2)^{-1} \underset{\sim s}{¾^2} + \underset{\sim s}{¾^2} \right) \quad \text{(A8)}$$

$$= \underset{\sim}{¼}' \left((\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) + \frac{\underset{\sim s}{¾^2}}{1 - Á_1^2} \right);$$

where $\underset{\sim}{¶}$ is a k-vector of ones, $\odot$ is the element-by-element multiplication operator and $E[(\underset{\sim t}{y} - ¹ \underset{\sim}{¶}) \odot (\underset{\sim t}{y} - ¹ \underset{\sim}{¶}) | \underset{\sim t}{S}]$ stacks the vector $E[(y_t - ¹) \odot (y_t - ¹) \mid S_t = i]$ for $i = 1; :::; k$. The third equality uses the property of the steady state probabilities that $\underset{\sim}{¼}' P^i = \underset{\sim}{¼}'$.

The third centered moment can be derived using the result that

$$E[(y_t - ¹)^3] = E[(¹_{s_t} - ¹)^3] + 3Á_1^2 E[(¹_{s_t} - ¹)(y_{t-1} - ¹_{s_{t-1}})^2] + 3E[(¹_{s_t} - ¹)¾^2_{s_t}{}^{2 2}_{t}]; \quad \text{(A9)}$$

which can be veri¯ed by expanding equation (6) around ¹. To derive an expression for the second term notice that

$$E[(y_t - ¹_{s_t})^2 | S_t = i] = Á_1^2 \sum_{j=1}^{k} E[(y_{t-1} - ¹_{s_{t-1}})^2 \mid S_{t-1} = j \cap S_t = i] \bullet$$

$$\text{Prob}(S_{t-1} = j \mid S_t = i) + ¾_i^2;$$

$$= Á_1^2 \sum_{j=1}^{k} E[(y_{t-1} - ¹_{s_{t-1}})^2 \mid S_{t-1} = j] \cdot b_{ij} + ¾_i^2; \quad \text{(A10)}$$

since the expectation of $(y_{t-1} - ¹_{s_{t-1}})^2$ conditional on S_{t-1} is the same as its expectation conditional on S_{t-1} and S_t. Stacking the equations resulting from setting $i = 1; :::; k$, we have

$$E[(\underset{\sim t}{y} - \underset{\sim s_t}{¹})^2 |\underset{\sim t}{S}] = Á_1^2 \cdot B \cdot E[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{¹})^2 |\underset{\sim t-1}{S}] + \underset{\sim s}{¾^2}; \quad \text{(A11)}$$

so that, under stationarity of the process,

$$E\left[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{¹})^2 |\underset{\sim t-1}{S}\right] = \left(I_k - Á_1^2 B\right)^{-1} \cdot \underset{\sim s}{¾^2}; \quad \text{(A12)}$$

where I_k is the k-dimensional identity matrix. Using this in (A11) and noting that $E\left[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{¹})^2 |\underset{\sim t}{S}\right] = B(I_k - Á_1^2 B)^{-1} \cdot \underset{\sim s}{¾^2}$, we get

$$\begin{aligned} E[(y_t - ¹)^3] = {} & \underset{\sim}{¼}' \left((\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \right) \\ & + 3Á_1^2 \underset{\sim}{¼}' \left((B(I_k - Á_1^2 B)^{-1} \underset{\sim s}{¾^2}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \right) \quad \text{(A13)} \\ & + 3 \underset{\sim}{¼}' \left((\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \odot \underset{\sim s}{¾^2} \right) \end{aligned}$$

The fourth centered moment is given by

$$
\begin{aligned}
E[(y_t - \mu)^4] &= E[(\mu_{s_t} - \mu)^4] + 6\phi_1^2 E[(\mu_{s_t} - \mu)^2 (y_{t-1} - \mu_{s_{t-1}})^2] + 6E[(\mu_{s_t} - \mu)^2 \sigma_{s_t}^2 \varepsilon_t^2] \\
&\quad + \phi_1^4 E[(y_{t-1} - \mu_{s_{t-1}})^4] + 6\phi_1^2 E[(y_{t-1} - \mu_{s_{t-1}})^2 \sigma_{s_t}^2 \varepsilon_t^2] + E[\sigma_{s_t}^4 \varepsilon_t^4]. \qquad \text{(A14)}
\end{aligned}
$$

Most of these terms are simple to evaluate, but notice that

$$
E[(\underset{\sim t}{y} - \underset{\sim s_t}{\mu})^4 | \underset{\sim t}{S}] = \phi_1^4 B \cdot E[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{\mu})^4 | \underset{\sim t-1}{S}] + 3 \underset{\sim s}{\sigma^4} + 6\phi_1^2 E[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{\mu})^2 \underset{\sim s_t}{\sigma^2} | \underset{\sim t}{S}].
$$

From (6) we also have

$$
\begin{aligned}
E[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{\mu})^2 \underset{\sim s_t}{\sigma^2} | \underset{\sim t}{S}] &= \left(B \cdot E[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{\mu})^2 | \underset{\sim t-1}{S}]\right) \odot \underset{\sim s}{\sigma^2} \\
&= \left(B \cdot (I_k - \phi_1^2 B)^{-1} \underset{\sim s}{\sigma^2}\right) \odot \underset{\sim s}{\sigma^2}.
\end{aligned}
$$

Thus, provided the process is stationary, we get

$$
E[(\underset{\sim t}{y} - \underset{\sim s_t}{\mu})^4 | \underset{\sim t}{S}] = (I_k - \phi_1^4 B)^{-1} \left(3 \underset{\sim s}{\sigma^4} + 6\phi_1^2 \left(B \cdot (I_k - \phi_1^2 B)^{-1} \underset{\sim s}{\sigma^2}\right) \odot \underset{\sim s}{\sigma^2}\right). \qquad \text{(A15)}
$$

Using this together with the equation $E[(y_t - \mu)^4] = \underset{\sim}{\pi}' E[(\underset{\sim t}{y} - \mu \underset{\sim}{\iota})^4 | \underset{\sim t}{S}]$, (A14) becomes

$$
\begin{aligned}
E[(y_t - \mu)^4] &= \underset{\sim}{\pi}' \left((\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \underset{\sim}{\mu}) \odot (\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota})\right) \\
&\quad + 6\phi_1^2 \underset{\sim}{\pi}' \left((B(I_k - \phi_1^2 B)^{-1} \underset{\sim s}{\sigma^2}) \odot (\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota})\right) \\
&\quad + 6 \underset{\sim}{\pi}' \left((\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota}) \odot (\underset{\sim s}{\mu} - \mu \underset{\sim}{\iota}) \odot \underset{\sim s}{\sigma^2}\right) \\
&\quad + \phi_1^4 \underset{\sim}{\pi}' (I_k - \phi_1^4 B)^{-1} \left(3 \underset{\sim s}{\sigma^4} + 6\phi_1^2 (B \cdot (I_k - \phi_1^2 B)^{-1} \underset{\sim s}{\sigma^2}) \odot \underset{\sim s}{\sigma^2}\right) \qquad \text{(A16)} \\
&\quad + 6\phi_1^2 \underset{\sim}{\pi}' \left((B \cdot (I_k - \phi_1^2 B)^{-1} \underset{\sim s}{\sigma^2}) \odot \underset{\sim s}{\sigma^2}\right) + 3 \underset{\sim}{\pi}' \underset{\sim s}{\sigma^4}.
\end{aligned}
$$

Collecting terms we get the expression in Proposition 2.

Proof of Proposition 3

The unconditional variance of the process underlying MSIII can be evaluated from the expression

$$
\begin{aligned}
E[(y_t - \mu)^2] &= E\left[(\mu_{s_t} - \mu)^2 + \phi_{1 s_{t-1}}^2 (y_{t-1} - \mu_{s_{t-1}})^2 + \sigma_{s_t}^2 \varepsilon_t^2\right] + 2\mathrm{Cov}(\mu_{s_t} - \mu; \sigma_{s_t} \varepsilon_t) \\
&\quad + 2\mathrm{Cov}\left(\mu_{s_t} - \mu; \phi_{1 s_{t-1}} (y_{t-1} - \mu_{s_{t-1}})\right) \qquad \text{(A17)} \\
&\quad + 2\mathrm{Cov}\left(\phi_{1 s_{t-1}} (y_{t-1} - \mu_{s_{t-1}}); \sigma_{s_t} \varepsilon_t\right)
\end{aligned}
$$

Again the covariance terms are zero so only the second term has changed from (A7). To evaluate this term we condition on S_t to get the set of equations

$$E\left[(\underset{\sim t}{y} - \underset{\sim s_t}{¹})^2 | \underset{\sim t}{S}\right] = B \cdot ©^2 E\left[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{¹})^2 | \underset{\sim t-1}{S}\right] + \underset{\sim s}{¾^2}; \qquad (A18)$$

where © is the $k \times k$ diagonal matrix de¯ned just before Proposition 3. Under stationarity of y_t we have

$$E\left[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{¹})^2 | \underset{\sim t-1}{S}\right] = (I_k - B©^2)^{-1} \underset{\sim s}{¾^2}; \qquad (A19)$$

from which the variance of y_t easily follows from (A17):

$$E[(y_t - ¹)^2] = \underset{\sim}{¼'} \left((\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \right) + \underset{\sim}{¼'} \left(©^2 (I_k - B©^2)^{-1} \underset{\sim s}{¾^2} + \underset{\sim s}{¾^2} \right): \qquad (A20)$$

To derive the third moment, notice that the only new term relative to (A9) is $3E[(¹_{s_t} - ¹)Á^2_{1 s_{t-1}}(y_{t-1} - ¹_{s_{t-1}})^2]$. Conditioning on S_t and applying the steady state probabilities, we get

$$E[(¹_{s_t} - ¹)Á^2_{1 s_{t-1}}(y_{t-1} - ¹_{s_{t-1}})^2] = \underset{\sim}{¼'} \left((B \cdot ©^2 (I_k - B©^2)^{-1} \underset{\sim s}{¾^2}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \right): \qquad (A21)$$

Inserting this in (A9) and using the similar results for MSII, the skewness result follows immediately. The fourth moment of MSIII can be evaluated from the expression

$$\begin{aligned} E[(y_t - ¹)^4] &= E[(¹_{s_t} - ¹)^4] + 6E\left[(¹_{s_t} - ¹)^2 Á^2_{1 s_{t-1}} (y_{t-1} - ¹_{s_{t-1}})^2\right] \\ &\quad + 6E[(¹_{s_t} - ¹)^2 ¾^2_{s_t} 2^2_t] + E[Á^4_{1 s_{t-1}} (y_{t-1} - ¹_{s_{t-1}})^4] \\ &\quad + 6E[Á^2_{1 s_{t-1}} (y_{t-1} - ¹_{s_{t-1}})^2 ¾^2_{s_t} 2^2_t] + E[¾^4_{s_t} 2^4_t]: \end{aligned} \qquad (A22)$$

The second, fourth and ¯fth terms in this expression have changed compared with (A14). We derive these expressions as follows:

$$E[(¹_{s_t} - ¹)^2 Á^2_{s_{t-1}} (y_{t-1} - ¹_{s_{t-1}})^2] = \underset{\sim}{¼'} \left((B©^2 (I_k - B©^2)^{-1} \underset{\sim s}{¾^2}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \odot (\underset{\sim s}{¹} - ¹ \underset{\sim}{¶}) \right); \qquad (A23)$$

$$E\left[Á^2_{1 s_{t-1}} (y_{t-1} - ¹_{s_{t-1}})^2 ¾^2_{s_t} 2^2_t\right] = \underset{\sim}{¼'} \left((B©^2 (I_k - B©^2)^{-1} \underset{\sim s}{¾^2}) \odot \underset{\sim s}{¾^2} \right): \qquad (A24)$$

Finally, conditioning on S_t we get the system of equations

$$\begin{aligned} E\left[(\underset{\sim t}{y} - \underset{\sim s_t}{¹})^4 | \underset{\sim t}{S}\right] &= B©^4 \cdot E\left[(\underset{\sim t-1}{y} - \underset{\sim s_{t-1}}{¹})^4 | \underset{\sim t-1}{S}\right] + 3 \underset{\sim s}{¾^4} \\ &\quad + 6 \left(B©^2 (I_k - B©^2)^{-1} \underset{\sim s}{¾^2} \right) \odot \underset{\sim s}{¾^2}: \end{aligned} \qquad (A25)$$

Hence

$$\mathrm{E}\left[(\underset{\sim t}{\mathrm{y}} - \underset{\sim s_t}{¹})^4 | \underset{\sim t}{\mathrm{S}}\right] = (\mathrm{I}_k - \mathrm{B}©^4)^{-1}\left(3\,\underset{\sim s}{¾^4} + 6(\mathrm{B}©^2(\mathrm{I}_k - \mathrm{B}©^2)^{-1}\underset{\sim s}{¾^2}) \odot \underset{\sim s}{¾^2}\right); \quad \text{(A26)}$$

and the fourth term entering (A22) is $\underset{\sim}{¼'}\ ©^4$ times this expression. Inserting these terms in (A22), we get the kurtosis expression stated in Proposition 3.

Proof of Proposition 4

We ¯rst compute $\mathrm{E}[(\mathrm{y}_t - ¹)(\mathrm{y}_{t-1} - ¹)]$ for MSI and then demonstrate how to generalize the result. Notice that

$$\begin{aligned}\mathrm{E}[(\mathrm{y}_t - ¹)(\mathrm{y}_{t-1} - ¹)] &= \mathrm{E}\left[((¹_{s_t} - ¹) + ¾_{s_t}²_t)\,((¹_{s_t-1} - ¹) + ¾_{s_t-1}²_{t-1})\right] \\ &= \mathrm{E}[(¹_{s_t} - ¹)(¹_{s_t-1} - ¹)] = \underset{\sim}{¼'}\left((\mathrm{B}(\underset{\sim s}{¹} - ¹\,\underset{\sim}{¶})) \odot (\underset{\sim s}{¹} - ¹\,\underset{\sim}{¶})\right) : \text{(A27)}\end{aligned}$$

When we consider the n'th order autocovariance of y_t, the only part of the calculation that changes is the transition probabilities, i.e. instead of using B we use B^n.

To compute the autocovariances of MSII, notice that

$$\begin{aligned}\mathrm{E}[(\mathrm{y}_t - ¹)(\mathrm{y}_{t-1} - ¹)] &= \mathrm{E}\left[\left((¹_{s_t} - ¹) + Á_1(\mathrm{y}_{t-1} - ¹_{s_{t-1}}) + ¾_{s_t}²_t\right) : \right. \\ &\qquad \left.\left((\mathrm{y}_{t-1} - ¹_{s_{t-1}}) + (¹_{s_{t-1}} - ¹)\right)\right] \quad \text{(A28)}\end{aligned}$$

Obviously $\mathrm{E}[¾_{s_t}²_t(\mathrm{y}_{t-1} - ¹_{s_{t-1}})] = \mathrm{E}[¾_{s_t}²_t(¹_{s_{t-1}} - ¹)] = 0$, and, from the independence of $²_{t-i}$ and S_t,

$$\mathrm{E}\left[((¹_{s_t} - ¹)(\mathrm{y}_{t-1} - ¹_{s_{t-1}}))\right] = \mathrm{E}\left[(¹_{s_{t-1}} - ¹)Á_1(\mathrm{y}_{t-1} - ¹_{s_{t-1}})\right] = 0: \quad \text{(A29)}$$

This leaves us with the terms $\mathrm{E}[(¹_{s_t} - ¹)(¹_{s_{t-1}} - ¹) + Á_1(\mathrm{y}_{t-1} - ¹_{s_{t-1}})^2]$, and hence

$$\mathrm{E}[(\mathrm{y}_t - ¹)(\mathrm{y}_{t-1} - ¹)] = \underset{\sim}{¼'}\left((\mathrm{B}(\underset{\sim s}{¹} - ¹\,\underset{\sim}{¶})) \odot (\underset{\sim s}{¹} - ¹\,\underset{\sim}{¶})\right) + Á_1\,\underset{\sim}{¼'}\,(\mathrm{I}_k - Á_1^2\mathrm{B})^{-1}\,\underset{\sim s}{¾^2} : \quad \text{(A30)}$$

This can easily be generalized to obtain the n'th order autocovariance by noting that

$$\mathrm{y}_{t+n} - ¹ = ¹_{s_{t+n}} - ¹ + Á_1^n(\mathrm{y}_t - ¹_{s_t}) + \sum_{i=1}^{n} Á_1^{n-i}¾_{s_{t+i}}²_{t+i}; \quad \text{(A31)}$$

so that the autocovariance becomes

$$\mathrm{E}\left[(\mathrm{y}_{t+n} - ¹)(\mathrm{y}_t - ¹)\right] = \mathrm{E}[(¹_{s_{t+n}} - ¹)(¹_{s_t} - ¹) + Á_1^n(\mathrm{y}_t - ¹_{s_t}) \quad {}^{\mathrm{n}}_{1} + +$$

.52 1.44 TD /F1 10.08 T 12.7)

To obtain the ¯rst order autocovariance of the MSIII process, we need to evaluate $E[(¹_{s_t} - ¹)(¹_{s_{t-1}} - ¹) + Á_{1s_{t-1}}(y_{t-1} - ¹_{s_{t-1}})^2]$:

$$E\left[(y_t - ¹)(y_{t-1} - ¹)\right] = \underset{\sim}{¼}' \left((B(\underset{\sim s}{¹} - ¹\ \underset{\sim}{¶})) \odot (\underset{\sim s}{¹} - ¹\ \underset{\sim}{¶}) \right) + \underset{\sim}{¼}'\ ©(I_k - B©^2)^{-1}\ \underset{\sim s}{¾^2} : \qquad \text{(A33)}$$

Second and higher order autocorrelations do not follow as easily from this expression as (A32) follows from (A31), however, due to the state dependence in Á. To evaluate the n'th order autocovariance, we need to compute

$$E\left[(y_{t+n} - ¹)(y_t - ¹)\right] = E\left[(¹_{s_{t+n}} - ¹)(¹_{s_t} - ¹)\right] + E\left[\left(\prod_{i=0}^{n-1} Á_{1s_{t+i}}\right)(y_t - ¹_{s_t})^2\right] : \qquad \text{(A34)}$$

This expression can be written as stated in Proposition 4.

Proof of Proposition 5

Again we ¯rst derive the result for the simple MS model (MSI). Note that for this process

$$y_t^2 = \left(¹_{s_t}^2 + 2¹_{s_t}¾_{s_t}²_t + ¾_{s_t}^2 ²_t^2\right); \qquad \text{(A35)}$$

$$y_{t-1}^2 = (y_{t-1} - ¹_{s_{t-1}})^2 + 2¹_{s_{t-1}}(y_{t-1} - ¹_{s_{t-1}}) + ¹_{s_{t-1}}^2; \qquad \text{(A36)}$$

so their expected cross-product is

$$\begin{aligned} E[y_t^2 y_{t-1}^2] &= E\left[¹_{s_t}^2(y_{t-1} - ¹_{s_{t-1}})^2 + ¹_{s_t}^2 ¹_{s_{t-1}}^2 + ¾_{s_t}^2 ²_t^2 (y_{t-1} - ¹_{s_{t-1}})^2 + ¾_{s_t}^2 ²_t^2 ¹_{s_{t-1}}^2\right] \\ &= \underset{\sim}{¼}'\left((B\ \underset{\sim s}{¾^2}) \odot \underset{\sim s}{¹^2}\right) + \underset{\sim}{¼}'\left((B\ \underset{\sim s}{¹^2}) \odot \underset{\sim s}{¹^2}\right) + \underset{\sim}{¼}'\left((B\ \underset{\sim s}{¾^2}) \odot \underset{\sim s}{¾^2}\right) \\ &\quad + \underset{\sim}{¼}'\left((B\ \underset{\sim s}{¹^2}) \odot \underset{\sim s}{¾^2}\right); \end{aligned} \qquad \text{(A37)}$$

where we used that, for the simple Markov switching model, $E[(\underset{\sim t}{y} - \underset{\sim s}{¹})^2 \,|\, \underset{\sim t}{S}] = \underset{\sim s}{¾^2}$. The expression in Proposition 5 follows by collecting terms. Once again, the general expression for $E[(y_t^2 - E[y_t^2])(y_{t-n}^2 - E[y_{t-n}^2])]$ can be derived by substituting B^n for B.

For MSII, the ¯rst order autocovariance of y_t can be based on

$$\begin{aligned} y_t^2 - E[y_t^2] &= ¹_{s_t}^2 + Á_1^2(y_{t-1} - ¹_{s_{t-1}})^2 + ¾_{s_t}^2 ²_t^2 + 2¹_{s_t}Á_1(y_{t-1} - ¹_{s_{t-1}}) \\ &\quad + 2¹_{s_t}¾_{s_t}²_t + 2Á_1(y_{t-1} - ¹_{s_{t-1}})¾_{s_t}²_t - E[y_t^2]; \end{aligned} \qquad \text{(A38)}$$

and

$$y_{t-1}^2 - E[y_{t-1}^2] = (y_{t-1} - ¹_{s_{t-1}})^2 + 2¹_{s_{t-1}}(y_{t-1} - ¹_{s_{t-1}}) + ¹_{s_{t-1}}^2 - E[y_{t-1}^2]: \qquad \text{(A39)}$$

Using that the expected value of cross-product terms in which ε_t enter linearly is zero we get

$$
\begin{aligned}
& E\left[(y_t^2 - E[y_t^2])(y_{t-1}^2 - E[y_{t-1}^2])\right] = \\
= \; & E\Big[\mu_{s_t}^2 (y_{t-1} - \mu_{s_{t-1}})^2 + \mu_{s_t}^2 \mu_{s_{t-1}}^2 + \phi_1^2 (y_{t-1} - \mu_{s_{t-1}})^4 + \phi_1^2 (y_{t-1} - \mu_{s_{t-1}})^2 \mu_{s_{t-1}}^2 \\
& + \sigma_{s_t}^2 \varepsilon_t^2 (y_{t-1} - \mu_{s_{t-1}})^2 + \sigma_{s_t}^2 \varepsilon_t^2 \mu_{s_{t-1}}^2 + 4\phi_1 \mu_{s_t} \mu_{s_{t-1}} (y_{t-1} - \mu_{s_{t-1}})^2\Big] - (E[y_t^2])^2. \qquad \text{(A40)}
\end{aligned}
$$

A similar expression holds for MSIII, substituting $\phi_{s_{t-1}}$ in place of ϕ_1. To get the n'th order autocovariance we use that, for MSII,

$$
\begin{aligned}
y_{t+n}^2 = \; & \mu_{s_{t+n}}^2 + \phi_1^{2n} (y_t - \mu_{s_t})^2 + \left(\sum_{i=1}^{n} \phi_1^{(n-i)} \sigma_{s_{t+i}} \varepsilon_{t+i} \right)^2 \\
& + 2\mu_{s_{t+n}} \phi_1^n (y_t - \mu_{s_t}) + f(\varepsilon_{t+n}, ..., \varepsilon_{t+1}), \qquad \text{(A41)}
\end{aligned}
$$

where $f(\cdot)$ is a linear function of its arguments and hence will be uncorrelated with terms dated period t or earlier. For MSIII, the similar expression is more complex:

$$
\begin{aligned}
y_{t+n}^2 = \; & \mu_{s_{t+n}}^2 + \left(\prod_{i=0}^{n-1} \phi_{1 s_{t+i}}^2 \right) (y_t - \mu_{s_t})^2 + \left(\sum_{i=1}^{n} \left(\prod_{j=i}^{n-1} \phi_{1 s_{t+j}} \right) \sigma_{s_{t+i}} \varepsilon_{t+i} \right)^2 \\
& + 2\mu_{s_{t+n}} \left(\prod_{i=0}^{n-1} \phi_{1 s_{t+i}} \right) (y_t - \mu_{s_t}) + g(\varepsilon_{t+n}, ..., \varepsilon_{t+1}), \qquad \text{(A42)}
\end{aligned}
$$

where again $g(\cdot)$ is a linear function of its arguments and we define $\prod_{i=n}^{n-1} \phi_{1 s_{t+i-1}} \equiv 1$. Using these expressions, equation (A40) can be extended to obtain the n-period autocovariances:

$$
\begin{aligned}
& E\left[(y_{t+n}^2 - E[y_{t+n}^2])(y_t^2 - E[y_t^2])\right] \\
= \; & E\Big[\mu_{s_{t+n}}^2 (y_t - \mu_{s_t})^2 + \mu_{s_{t+n}}^2 \mu_{s_t}^2 + \phi_1^{2n} (y_t - \mu_{s_t})^4 + \phi_1^{2n} (y_t - \mu_{s_t})^2 \mu_{s_t}^2 \qquad \text{(A43)} \\
& + \left((y_t - \mu_{s_t})^2 + \mu_{s_t}^2\right) \left(\sum_{i=1}^{n} \phi_1^{2(n-i)} \sigma_{s_{t+i}}^2 \right) + 4\phi_1^n \mu_{s_{t+n}} \mu_{s_t} (y_t - \mu_{s_t})^2 \Big] - (E[y_t^2])^2 \qquad \text{(MSII)} \\
= \; & E\Bigg[\mu_{s_{t+n}}^2 (y_t - \mu_{s_t})^2 + \mu_{s_{t+n}}^2 \mu_{s_t}^2 + (y_t - \mu_{s_t})^4 \prod_{i=0}^{n-1} \phi_{1 s_{t+i}}^2 + (y_t - \mu_{s_t})^2 \mu_{s_t}^2 \prod_{i=0}^{n-1} \phi_{1 s_{t+i}}^2 \\
& + \left((y_t - \mu_{s_t})^2 + \mu_{s_t}^2\right) \sum_{i=1}^{n} \left(\prod_{j=i}^{n-1} \phi_{1 s_{t+j}}^2 \right) \sigma_{s_{t+i}}^2 + 4\mu_{s_{t+n}} \mu_{s_t} (y_t - \mu_{s_t})^2 \prod_{i=0}^{n-1} \phi_{1 s_{t+i}} \Bigg] - (E[y_t^2])^2 \qquad \text{(MSIII)}
\end{aligned}
$$

which can be written as stated in Proposition 5.

References

Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31, 307-27.

Cecchetti, S.G., Lam, P. , and Mark, N.C. (1990) Mean Reversion in Equilibrium Asset Prices. American Economic Review 80, 398-418.

Clark, P.C. (1973) A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Econometrica 41, 135-155.

Cox, D.R. and Miller, H.D. (1965) The Theory of Stochastic Processes. Chapman and Hall.

Diebold, F. (1986) Comment on Modelling the Persistence of Conditional Variance. Econometric Reviews 5, 51-56.

Engel, C. and J.D. Hamilton (1990) Long Swings in the Dollar: Are They in the Data and Do Markets Know It?. American Economic Review 89, 689-713.

Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. In°ation. Econometrica 50, 987-1008.

Granger, C.W.J., and Orr, D. (1972) In¯nite Variance and Research Strategy in Time Series Analysis. Journal of American Statistical Association 67, 275-285.

Gray, S.F. (1996) Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process. Journal of Financial Economics 42, 27-62.

Hamilton, J.D. (1988) Rational Expectations Econometric Analysis of Changes in Regime. An Investigation of the Term Structure of Interest Rates. Journal of Economic Dynamics and Control 12, 385-423.

Hamilton, J.D. (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica 57, 357 - 84.

Hamilton, J.D. and Susmel, R. (1994) Autoregressive Conditional Heteroskedasticity and Changes in Regime. Journal of Econometrics 64, 307-333.

Hendry, D. and Neale, A.J. (1991) A Monte Carlo Study of the Effects of Structural Breaks on Tests for Unit Roots. Pages 95-119 in P. Huckl and A. Westlund (Eds.) Economic Structural Change. Analysis and Forecasting. Springer Verlag, Berlin.

Lamoureux, C.G., and Lastrapes, W.D. (1990) Persistence in Variance, Structural Change and the GARCH Model. Journal of Business and Economic Statistics 8, 225-234.

Lumsdaine, R.L. and Ng, S. (1997) Testing for ARCH in the Presence of a Possibly Misspecified Conditional Mean. Working paper, Princeton University and Boston College.

Neftci, S.N. (1984) Are Economic Time Series Asymmetric over the Business Cycle? Journal of Political Economy 92, 307-328.

Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59, 347-370.

Pagan, A. and Schwert, G.W. (1990) Alternative Models for Conditional Stock Volatility. Journal of Econometrics 45, 267-90.

Perron, P. (1989) The Great Crash, the Oil Price Shock and the Unit Root Hypothesis. Econometrica 57, 1361-1401.

Perron, P. (1990) Testing for a Unit Root in a Time Series with a Changing Mean. Journal of Business and Economic Statistics 8, 153-162.

Turner, C.M., Startz, R., and Nelson, C.R. (1989) A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market. Journal of Financial Economics 25, 3-22.