Chapter 1

The Markov–Switching Vector Autoregressive Model

This first chapter is devoted to a general introduction into the Markov–switching vector autoregressive (MS-VAR) time series model. In *Section 1.2* we present the fundamental assumptions constituting this class of models. The discussion of the two components of MS-VAR processes will clarify their on time invariant vector autoregressive and Markov-chain models. Some basic stochastic properties of MS-VAR processes are presented in *Section 1.3.* Finally, MS-VAR models are compared to alternative non-normal and non-linear time series models proposed in the literature. As most non-linear models have been developed for univariate time series, this discussion is restricted to this case. However, generalizations to the vector case are also considered.

1.1 General Introduction

Reduced form vector autoregressive (VAR) models have been become a dominant research strategy in empirical macroeconomics since SIMS [1980]. In this study we will consider VAR models with changes in regime, most results will carry over to structural dynamic econometric models by treating them as restricted VAR models. When the system is subject to regime shifts, the parameters θ of the VAR process will be time-varying. But the process might be time-invariant conditional on an unobservable regime variable s_t which indicates the regime prevailing at time t. Let M denote the number of feasible regimes, so that $s_t \in \{1, \dots, M\}$. Then the con-

H.-M. Krolzig, *Markov-Switching Vector Autoregressions*

ditional probability density of the observed time series vector y_t is given by

$$p(y_t|Y_{t-1}, s_t) = \begin{cases} f(y_t|Y_{t-1}, \theta_1) & \text{if } s_t = 1 \\ \vdots & \\ f(y_t|Y_{t-1}, \theta_M) & \text{if } s_t = M, \end{cases} \tag{1.1}$$

where θ_m is the VAR parameter vector in regime $m = 1, \ldots, M$ and Y_{t-1} are the observations $\{y_{t-j}\}_{j=1}^{\infty}$.

Thus, for a given regime s_t, the time series vector y_t is generated by a vector autoregressive process of order p (VAR(p) model) such that

$$\mathsf{E}[y_t|Y_{t-1}, s_t] = \nu(s_t) + \sum_{j=1}^{p} A_j(s_t) y_{t-j},$$

where u_t is an innovation term,

$$u_t = y_t - \mathsf{E}[y_t|Y_{t-1}, s_t].$$

The innovation process u_t is a zero-mean white noise process with a variance-covariance matrix $\Sigma(s_t)$, which is assumed to be Gaussian:

$$u_t \sim \text{NID}\,(\mathbf{0}, \Sigma(s_t)).$$

If the VAR process is defined conditionally upon an unobservable regime as in equation (1.1), the description of the data generating mechanism has to be completed by assumptions regarding the regime generating process. In Markov-switching vector autoregressive (MS-VAR) models – the subject of this study – it is assumed that the regime s_t is generated by a discrete-state homogeneous Markov chain:[1]

$$\Pr(s_t|\{s_{t-j}\}_{j=1}^{\infty}, \{y_{t-j}\}_{j=1}^{\infty}) = \Pr(s_t|s_{t-1}; \rho),$$

where ρ denotes the vector of parameters of the regime generating process.

The vector autoregressive model with Markov-switching regimes is founded on at least three traditions. The first is the linear time-invariant *vector autoregressive model*, which is the framework for the analysis of the relation of the variables of the system, the dynamic propagation of innovations to the

[1] The notation $\Pr(\cdot)$ refers to a discrete probability measure, while $p(\cdot)$ denotes a probability density function.

system, and the effects of changes in regime. Secondly, the basic statistical techniques have been introduced by BAUM & PETRIE [1966] and BAUM *et al.* [1970] for *probabilistic functions of Markov chains*, while the MS-VAR model also encompasses older concepts as the *mixture of normal distributions model* attributed to PEARSON [1894] and the *hidden Markov-chain model* traced back to BLACKWELL & KOOPMANS [1975] and HELLER [1965]. Thirdly, in econometrics, the first attempt to create Markov-switching regression models were undertaken by GOLDFELD & QUANDT [1973], which remained, however, rather rudimentary. The first comprehensive approach to the statistical analysis of Markov-switching regression models has been proposed by LINDGREN [1978] which is based on the ideas of BAUM *et al.* [1970]. In time series analysis, the introduction of the Markov-switching model is due to HAMILTON [1988], [1989] on which most recent contributions (as well as this study) are founded. Finally, our consideration of MS-VAR models as a Gaussian vector autoregressive process conditioned on an exogenous regime generating process is closely related to state space models as well as the concept of doubly stochastic processes introduced by TJØSTHEIM [1986b].

The MS-VAR model belongs to a more general class of models that characterize a non-linear data generating process as piecewise linear by restricting the process to be linear in each regime, where the regime is conditioned is unobservable, and only a discrete number of regimes are feasible.[2] These models differ in their assumptions concerning the stochastic process generating the regime:

(i.) The *mixture of normal distributions* model is characterized by serially independently distributed regimes:

$$\Pr(s_t|\{s_{t-j}\}_{j=1}^{\infty}, \{y_{t-j}\}_{j=1}^{\infty}) = \Pr(s_t; \rho).$$

In contrast to MS-VAR models, the transition probabilities are independent of the history of the regime. Thus the conditional probability distribution of y_t is independent of s_{t-1},

$$\Pr(y_t|Y_{t-1}, s_{t-1}) = \Pr(y_t|Y_{t-1}),$$

---

[2] In the case of two regimes, POTTER [1990],[1993] proposed to call this class of non-linear, non-normal models the *single index generalized multivariate autoregressive* (SIGMA) model.

and the conditional mean $\mathsf{E}[y_t|Y_{t-1}, s_{t-1}]$ is given by $\mathsf{E}[y_t|Y_{t-1}]$.[3] Even so, this model can be considered as a restricted MS-VAR model where the transition matrix has rank one. Moreover, if only the intercept term will be regime-dependent, MS(M)-VAR(p) processes with Gaussian errors and *i.i.d.* switching regimes are observationally equivalent to time-invariant VAR(p) processes with non-normal errors. Hence, the modelling with this kind of model is very limited.

(ii.) In the *self-exciting threshold autoregressive* SETAR(p, d, r) model, the regime-generating process is not assumed to be exogenous but directly linked to the lagged endogenous variable y_{t-d}.[4] For a given but unknown threshold r, the 'probability' of the unobservable regime $s_t = 1$ is given by

$$\Pr(s_t = 1|\{s_{t-j}\}_{j=1}^{\infty}, \{y_{t-j}\}_{j=1}^{\infty}) = I(y_{t-d} \leq r) = \begin{cases} 1 & \text{if } y_{t-d} \leq r \\ 0 & \text{if } y_{t-d} > r, \end{cases}$$

While the presumptions of the SETAR and the MS-AR model seem to be quite different, the relation between both model alternatives is rather close. This is also illustrated in the appendix which gives an example showing that SETAR and MS-VAR models can be observationally equivalent.

(iii.) In the *smooth transition autoregressive* (STAR) model popularized by GRANGER & TERÄSVIRTA [1993], exogenous variables are mostly employed to model the weights of the regimes, but the regime switching rule can also be dependent on the history of the observed variables, *i.e.* y_{t-d}:

$$\Pr(s_t = 1|\{s_{t-j}\}_{j=1}^{\infty}, \{y_{t-j}\}_{j=1}^{\infty},) = F(y'_{t-d}\delta - r),$$

where $F(y'_{t-d}\delta - r)$ is a continuous function determining the weight of re-

[3] The likelihood function is given by

$$p(Y_T|Y_0; \theta, \bar{\xi}) = \sum_{t=1}^{T} \sum_{m=1}^{M} \bar{\xi}_m p(y_t|Y_{t-1}, \theta_m),$$

where $\theta = (\theta'_1, \ldots, \theta'_M)'$ collects the VAR parameters and $\bar{\xi}_m$ is the ergodic probability of regime m.

[4] In threshold autoregressive (TAR) processes, the indicator function is defined in a switching variable z_{t-d}, $d \geq 0$. In addition, indicator variables can be introduced and treated with error-in-variables techniques. Refer for example to COSSLETT & LEE [1985] and KAMINSKY [1993].

gime 1. For example, TERÄSVIRTA & ANDERSON [1992] use the logistic distribution function in their analysis of the U.S. business cycle.[5]

(iv.) All the previously mentioned models are special cases of an *endogenous selection Markov-switching vector autoregressive* model. In an EMS(M, d)-VAR(p) model the transition probabilities $p_{ij}(\cdot)$ are functions of the observed time series vector y_{t-d}:

$$\Pr(s_t = m | s_{t-1} = i, y_{t-d}) = p_{im}(y'_{t-d}\delta).$$

Thus the observed variables contain additional information on the conditional probability distribution of the states:

$$\Pr(s_t | \{s_{t-j}\}_{j=1}^{\infty}) \overset{a.e.}{\neq} \Pr(s_t | \{s_{t-j}\}_{j=1}^{\infty}, \{y_{t-j}\}_{j=1}^{\infty}).$$

Thus the regime generating process is no longer Markovian. In contrast to the SETAR and the STAR model, EMS-VAR models include the possibility that the threshold depends on the last regime, *e.g.* that the threshold for staying in regime 2 is different from the threshold for switching from regime 1 to regime 2 . The EMS(M, d)-VAR(p) model will be presented in *Section 10.3.* It is shown that the methods developed in this study for MS-VAR processes can easily be extended to capture EMS-VAR processes.

In this study, it will be shown that the MS-VAR model can encompass a wide spectrum of non-linear modifications of the VAR model proposed in the literature.

1.2 Markov-Switching Vector Autoregressions

1.2.1 The Vector Autoregression

Markov-switching vector autoregressions can be considered as generalizations of the basic finite order VAR model of order p. Consider the p-th order autoregression for the K-dimensional time series vector $y_t = (y_{1t}, \ldots, y_{Kt})'$, $t = 1, \ldots, T$,

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \tag{1.2}$$

[5] If $F(\cdot)$ is even, *e.g.* $F(y_{t-d} - r) = 1 - \exp\left\{-(y_{t-d} - r)^2\right\}$, a generalized *exponential autoregressive model* as proposed by OZAKI [1980] and HAGGAN & OZAKI [1981] ensues.

where $u_t \sim \mathrm{IID}(\mathbf{0}, \Sigma)$ and $y_0, \ldots, y_{1-p}$ are fixed. Denoting $A(\mathsf{L}) = \mathbf{I}_K - A_1\mathsf{L} - \ldots - A_p\mathsf{L}^p$ as the $(K \times K)$ dimensional lag polynomial, we assume that there are no roots on or inside the unit circle $|A(z)| \neq \mathbf{0}$ for $|z| \leq 1$ where L is the lag operator, so that $y_{t-j} = \mathsf{L}^j y_t$. If a normal distribution of the error is assumed, $u_t \sim \mathrm{NID}(\mathbf{0}, \Sigma)$, equation (1.2) is known as the intercept form of a stable *Gaussian* VAR(p) model. This can be reparametrized as the mean adjusted form of a VAR model:

$$y_t - \mu = A_1(y_{t-1} - \mu) + \ldots + A_p(y_{t-p} - \mu) + u_t, \tag{1.3}$$

where $\mu = (\mathbf{I}_K - \sum_{j=1}^{p} A_j)^{-1}\nu$ is the $(K \times 1)$ dimensional mean of y_t.

If the time series are subject to shifts in regime, the stable VAR model with its time invariant parameters might be inappropriate. Then, the MS–VAR model might be considered as a general regime-switching framework. The general idea behind this class of models is that the parameters of the underlying data generating process[6] of the *observed* time series vector y_t depend upon the *unobservable* regime variable s_t, which represents the probability of being in a different state of the world.

The main characteristic of the Markov-switching model is the assumption that the unobservable realization of the regime $s_t \in \{1, \ldots, M\}$ is governed by a discrete time, discrete state Markov stochastic process, which is defined by the transition probabilities

$$p_{ij} = \Pr(s_{t+1} = j | s_t = i), \quad \sum_{j=1}^{M} p_{ij} = 1 \quad \forall i, j \in \{1, \ldots, M\}. \tag{1.4}$$

More precisely, it is assumed that s_t follows an irreducible ergodic M state Markov process with the transition matrix $\mathbf{P}$. This will be discussed in *Section 1.2.4* in more detail.

In generalization of the mean-adjusted VAR(p) model in equation (1.3) we would like to consider Markov-switching vector autoregressions of order p and M regimes:

$$y_t - \mu(s_t) = A_1(s_t)\,(y_{t-1} - \mu(s_{t-1})) + \ldots + A_p(s_t)\,(y_{t-p} - \mu(s_{t-p})) + u_t, \tag{1.5}$$

where $u_t \sim \mathrm{NID}(\mathbf{0}, \Sigma(s_t))$ and $\mu(s_t), A_1(s_t), \ldots, A_p(s_t), \Sigma(s_t)$ are parameter shift functions describing the dependence of the parameters[7] $\mu, A_1, \ldots, A_p, \Sigma$ on

[6] For reasons of simplicity in notation, we do not introduce a separate notation for the theoretical representation of the stochastic process and its actual realizations.

[7] In the notation of state-space models, the varying *parameters* $\mu, \nu, A_1, \ldots, A_p, \Sigma$ become functions of the model's *hyper-parameters*.

the realized regime s_t, *e.g.*

$$\mu(s_t) = \begin{cases} \mu_1 & \text{if } s_t = 1, \\ \vdots & \\ \mu_M & \text{if } s_t = M. \end{cases} \tag{1.6}$$

In the model (1.5) there is after a change in the regime an immediate one–time jump in the process mean. Occasionally, it may be more plausible to assume that the mean smoothly approaches a new level after the transition from one state to another. In such a situation the following model with a regime-dependent intercept term $\nu(s_t)$ may be used:

$$y_t = \nu(s_t) + A_1(s_t)y_{t-1} + \ldots + A_p(s_t)y_{t-p} + u_t. \tag{1.7}$$

In contrast to the linear VAR model, the mean adjusted form (1.5) and the intercept form (1.7) of an MS(M)–VAR(p) model are not equivalent. In *Chapter 3* it will be seen that these forms imply different dynamic adjustments of the observed variables after a change in regime. While a permanent regime shift in the mean $\mu(s_t)$ causes an immediate jump of the observed time series vector onto its new level, the dynamic response to a once-and-for-all regime shift in the intercept term $\nu(s_t)$ is identical to an equivalent shock in the white noise series u_t.

In the most general specification of an MS-VAR model, all parameters of the autoregression are conditioned on the state s_t of the Markov chain. We have assumed that each regime m possesses its VAR(p) representation with parameters $\nu(m)$ (or μ_m), $\Sigma_m, A_{1m}, \ldots, A_{jm}, m = 1, \ldots, M$, such that

$$y_t = \begin{cases} \nu_1 + A_{11}y_{t-1} + \ldots + A_{p1}y_{t-p} + \Sigma_1^{1/2}u_t, & \text{if } s_t = 1 \\ \vdots & \\ \nu_M + A_{1M}y_{t-1} + \ldots + A_{pM}y_{t-p} + \Sigma_M^{1/2}u_t, & \text{if } s_t = M \end{cases}$$

where $u_t \sim \text{NID}\,(\mathbf{0}, \mathbf{I}_K)$.[8]

However for empirical applications, it might be more helpful to use a model where only some parameters are conditioned on the state of the Markov chain, while the

[8] Even at this early stage a complication arises if the mean adjusted form is considered. The conditional density of y_t depends not only on s_t but also on $s_{t-1}, \ldots, s_{t-p}$, *i.e.* M^{p+1} different conditional

other parameters are regime invariant. In *Section 1.2.2* some particular MS-VAR models will be introduced where the autoregressive parameters, the mean or the intercepts, are regime-dependent and where the error term is hetero- or homoskedastic. Estimating these particular MS-VAR models is discussed separately in *Chapter 9*.

1.2.2 Particular MS–VAR Processes

The MS-VAR model allows for a great variety of specifications. In principle, it would be possible to (i.) make all parameters regime-dependent and (ii.) to introduce separate regimes for each shifting parameter. But, this would be no practicable solution as the number of parameters of the Markov chain grows quadratic in the number of regimes and coincidently shrinks the number of observations usable for the estimation of the regime-dependent parameter. For these reasons a specific-to-general approach may be preferred for the determination of the regime generating process by restricting the shifting parameters (i.) to a part of the parameter vector and (ii.) to have identical break-points.

In empirical research, only some parameters will be conditioned on the state of the Markov chain while the other parameters will be regime invariant. In order to establish a unique notation for each model, we specify with the general MS(M) term the regime-dependent parameters:

M	Markov-switching *mean* ,
I	Markov-switching *intercept* term ,
A	Markov-switching *autoregressive parameters* ,
H	Markov-switching *heteroskedasticity* .

To achieve a distinction of VAR models with time-invariant mean and intercept term,

---

means of y_t are to be distinguished:

$$y_t = \begin{cases} \mu_1 + A_{11}(y_{t-1}-\mu_1) + \ldots + A_{p1}(y_{t-p}-\mu_1) + \Sigma_1^{1/2}u_t, & \text{if } s_t=1,\ldots,s_{t-p}=1 \\ \mu_1 + A_{11}(y_{t-1}-\mu_1) + \ldots + A_{p1}(y_{t-p}-\mu_2) + \Sigma_1^{1/2}u_t, & \text{if } s_t=1,\ldots,s_{t-p+1}=1,\ s_{t-p}=2 \\ \vdots & \\ \mu_1 + A_{11}(y_{t-1}-\mu_M) + \ldots + A_{p1}(y_{t-p}-\mu_M) + \Sigma_1^{1/2}u_t, & \text{if } s_t=1,\ s_{t-1}=M,\ldots,s_{t-p}=M \\ \vdots & \\ \mu_M + A_{1M}(y_{t-1}-\mu_1) + \ldots + A_{pM}(y_{t-p}-\mu_1) + \Sigma_M^{1/2}u_t, & \text{if } s_t=M,\ s_{t-1}=1,\ldots,s_{t-p}=1 \\ \vdots & \\ \mu_M + A_{1M}(y_{t-1}-\mu_1) + \ldots + A_{pM}(y_{t-p}-\mu_{M-1}) + \Sigma_M^{1/2}u_t, & \text{if } s_t=M \ldots s_{t-p+1}=M,\ s_{t-p}=M-1 \\ \mu_M + A_{1M}(y_{t-1}-\mu_M) + \ldots + A_{pM}(y_{t-p}-\mu_M) + \Sigma_M^{1/2}u_t, & \text{if } s_t=M,\ldots,s_{t-p}=M \end{cases}$$

Table 1.1: Special Markov Switching Vector Autoregressive Models

		MSM	MSI Specification		
		μ varying	μ invariant	ν varying	ν invariant
A_j invariant	Σ invariant	MSM–VAR	*linear* MVAR	MSI–VAR	*linear* VAR
	Σ varying	MSMH–VAR	MSH–MVAR	MSIH–VAR	MSH–VAR
A_j varying	Σ invariant	MSMA–VAR	MSA–MVAR	MSIA–VAR	MSA–VAR
	Σ varying	MSMAH–VAR	MSAH–MVAR	MSIAH–VAR	MSAH–VAR

To achieve a distinction of VAR models with time-invariant mean and intercept term, we denote the *mean* adjusted form of a vector autoregression as MVAR(p). An overview is given in *Table 1.1.* Obviously the MSI and the MSM specifications are equivalent if the order of the autoregression is zero. For this so-called hidden Markov-chain model, we prefer the notation MSI(M)-VAR(0). As it will be seen later on, the MSI(M)-VAR(0) model and MSI(M)-VAR(p) models with $p > 0$ are isomorphic concerning their statistical analysis. In *Section 10.3* we will further extend the class of models under consideration.

The MS-VAR model provides a very flexible framework which allows for heteroskedasticity, occasional shifts, reversing trends, and forecasts performed in a non-linear manner. In the following sections the focus is on models where the mean (MSM(M)–VAR(p) models) or the intercept term (MSI(M)–VAR(p) models) are subject to occasional discrete shifts; regime-dependent covariance structures of the process are considered as additional features.

1.2.3 The Regime Shift Function

At this stage it is useful to define the parameter shifts more clearly by formulating the system as a single equation by introducing "dummy" (or more precisely) indicator variables:

$$I(s_t = m) = \begin{cases} 1 \text{ if } s_t = m \\ 0 \text{ otherwise,} \end{cases}$$

where $m = 1, \ldots, M$. In the course of the following chapters it will prove helpful to collect all the information about the realization of the Markov chain in the vector ξ_t as

$$\xi_t = \begin{bmatrix} I(s_t = 1) \\ \vdots \\ I(s_t = M) \end{bmatrix}.$$

Thus, ξ_t denotes the unobserved state of the system. Since ξ_t consists of binary variables, it has some particular properties:

$$\mathsf{E}[\xi_t] = \begin{bmatrix} \Pr(s_t = 1) \\ \vdots \\ \Pr(s_t = M) \end{bmatrix} = \begin{bmatrix} \Pr(\xi_t = \iota_1) \\ \vdots \\ \Pr(\xi_t = \iota_M) \end{bmatrix},$$

where ι_m is the m-th column of the identity matrix. Thus $\mathsf{E}[\xi_t]$, or a well defined conditional expectation, represents the probability distribution of s_t. It is easily verified that $\mathbf{1}_M'\xi_t = 1$ as well as $\xi_t'\xi_t = 1$ and $\xi_t\xi_t' = \operatorname{diag}(\xi_t)$, where $\mathbf{1}_M = (1, \ldots, 1)'$ is an $(M \times 1)$ vector.

For example, we can now rewrite the mean shift function (1.6) as

$$\mu(s_t) = \sum_{m=1}^{M} \mu_m I(s_t = m).$$

In addition, we can use matrix notation to derive

$$\mu(s_t) = \mathbf{M}\xi_t,$$

where $\mathbf{M}$ is a $(K \times M)$ matrix containing the means,

$$\mathbf{M} = \begin{bmatrix} \mu_1 & \ldots & \mu_M \end{bmatrix}, \mu = \operatorname{vec}(\mathbf{M}).$$

We will occasionally use the following notation for the variance parameters:

$$\underset{(K \times MK)}{\boldsymbol{\Sigma}} = \begin{bmatrix} \Sigma_1 & \ldots & \Sigma_M \end{bmatrix}$$

$$\underset{\left(\frac{K(K+1)}{2}\times 1\right)}{\sigma_m} = \text{vech}\,(\Sigma_m), \quad \sigma = (\sigma_1', \ldots, \sigma_M')'$$

such that

$$\Sigma_t = \Sigma(s_t) = \mathbf{\Sigma}(\xi_t \otimes \mathbf{I}_K)$$

is a $(K \times K)$ matrix.

1.2.4 The Hidden Markov Chain

The description of the data-generating process is not completed by the *observational* equations (1.5) or (1.7). A model for the parameter generating process has to be formulated. If the parameters depend on a regime which is assumed to be stochastic and unobservable, a generating process for the states s_t must be postulated. Using this law, the evolution of regimes then might be inferred from the data. In the MS-VAR model the state process is an ergodic Markov chain with a finite number of states $s_t = 1, \ldots, M$ and transition probabilities p_{ij}.

It is convenient to collect the transition probabilities in the transition matrix $\mathbf{P}$,

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{11} & p_{12} & \cdots & p_{1M} \end{bmatrix}, \tag{1.8}$$

where $p_{iM} = 1 - p_{i1} - \ldots - p_{i,M-1}$ for $i = 1, \ldots, M$. To be more precise, all relevant information about the future of the *Markovian* process is included in the present state ξ_t

$$\Pr(\xi_{t+1}|\xi_t, \xi_{t-1}, \ldots; y_t, y_{t-1}, \ldots) = \Pr(\xi_{t+1}|\xi_t)$$

where the past and additional variables such as y_t reveal no relevant information beyond that of the actual state. The assumption of a *first-order* Markov process is not especially restrictive, since each Markov chain of an order greater than one can be reparametrized as a higher dimensional first-order Markov process (cf. FRIEDMANN [1994]). A comprehensive discussion of the theory of Markov chains with application to Markov-switching models is given by HAMILTON [1994b, ch. 22.2]. We will just give a brief introduction to some basic concepts related to MS-VAR models, in particular to the state-space form and the filter.

It is usually assumed that the Markov process is ergodic. A Markov chain is said to be *ergodic* if exactly one of the eigenvalues of the transition matrix $\mathbf{P}$ is unity and all other eigenvalues are inside the unit circle. Under this condition there exists a stationary or unconditional probability distribution of the regimes. The *ergodic probabilities* are denoted by $\bar{\xi} = \mathsf{E}[\xi_t]$. They are determined by the stationarity restriction $\mathbf{P}'\bar{\xi} = \bar{\xi}$ and the adding up restriction $\mathbf{1}'_M\bar{\xi} = 1$, from which it follows that

$$\bar{\xi} = \begin{bmatrix} \mathbf{I}_{M-1} - \mathbf{P}'_{1.M-1,1.M-1} & \mathbf{P}'_{1.M-1,M} \\ \mathbf{1}'_{M-1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{M-1} \\ 1 \end{bmatrix}. \tag{1.9}$$

If $\bar{\xi}$ is strictly positive, such that all regimes have a positive unconditional probability $\bar{\xi}_i > 0$, $i = 1, \ldots, M$, the process is called *irreducible*. The assumptions of ergodicity and irreducibility are essential for the theoretical properties of MS-VAR models, *e.g.* its property of being stationary. The estimation procedures, which will be introduced in *Chapter 6* and *Chapter 8* are flexible enough to capture even these degenerated cases, *e.g.* when there is a single jump ("structural break") into the absorbing state that prevails until the end of the observation period.

1.3 The Data Generating Process

After this introduction of the two components of MS-VAR models, (i.) the Gaussian VAR model as the conditional data generating process and (ii.) the Markov chain as the regime generating process, we will briefly discuss their main implications for the data generating process.

For given states ξ_t and lagged endogenous variables $Y_{t-1} = (y'_{t-1}, y'_{t-2}, \ldots, y'_1, y'_0, \ldots, y'_{1-p})'$ the conditional probability density function of y_t is denoted by $p(y_t|\xi_t, Y_{t-1})$. It is convenient to assume in (1.5) and (1.7) a normal distribution of the error term u_t, so that

$$\begin{aligned} & p(y_t|\xi_t = \iota_m, Y_{t-1}) \\ = \ & \ln(2\pi)^{-1/2} \ln|\Sigma|^{-1/2} \exp\{(y_t - \bar{y}_{mt})'\Sigma_m^{-1}(y_t - \bar{y}_{mt})\}, \end{aligned} \tag{1.10}$$

where $\bar{y}_{mt} = \mathsf{E}[y_t|\xi_t, Y_{t-1}]$ is the conditional expectation of y_t in regime m. Thus the conditional density of y_t for a given regime ξ_t is normal as in the VAR model defined in equation (1.2). Thus:

$$\begin{aligned} y_t\,|\xi_t = \iota_m, Y_{t-1} &\sim \text{NID}\left(\bar{y}_{mt}, \Sigma_m\right), \\ &\sim \text{NID}\left(\bar{y}_t'\xi_t,\ \mathbf{\Sigma}(\xi_t \otimes \mathbf{I}_K)\right), \end{aligned} \tag{1.11}$$

where the conditional means $\bar{y}_{mt}$ are summarized in the vector $\bar{y}_t$ which is *e.g.* in MSI specifications of the form

$$\bar{y}_t = \begin{bmatrix} \bar{y}_{1t} \\ \vdots \\ \bar{y}_{Mt} \end{bmatrix} = \begin{bmatrix} \nu_1 + \sum_{j=1}^p A_{1j} y_{t-j} \\ \vdots \\ \nu_M + \sum_{j=1}^p A_{Mj} y_{t-j} \end{bmatrix}.$$

Assuming that the information set available at time $t-1$ consists only of the sample observations and the pre-sample values collected in Y_{t-1} and the states of the Markov chain up to ξ_{t-1}, the conditional density of y_t is a mixture of normals[9]:

$$\begin{aligned} & p(y_t|\xi_{t-1} = \iota_i, Y_{t-1}) \\ = & \sum_{m=1}^{M} p(y_t|\xi_{t-1} = \iota_m, Y_{t-1}) \Pr(\xi_t|\xi_{t-1} = \iota_i) \\ = & \sum_{m=1}^{M} p_{im} \left(\ln(2\pi)^{-\frac{1}{2}} \ln|\Sigma_m|^{-\frac{1}{2}} \exp\{(y_t - \bar{y}_{mt})'\Sigma_m^{-1}(y_t - \bar{y}_{mt})\}\right) \end{aligned} \tag{1.12}$$

If the densities of y_t conditional on ξ_t and Y_{t-1} are collected in the vector η_t as

$$\eta_t = \begin{bmatrix} p(y_t|\xi_t = \iota_1, Y_{t-1}) \\ \vdots \\ p(y_t|\xi_t = \iota_M, Y_{t-1}) \end{bmatrix}, \tag{1.13}$$

equation (1.12) can be written as

$$p(y_t|\xi_{t-1}, Y_{t-1}) = \eta_t' \mathbf{P}' \xi_{t-1}. \tag{1.14}$$

---

[9]The reader is referred to HAMILTON [1994a] for an excellent introduction into the major concepts of Markov chains and to TITTERINGTON, SMITH & MAKOV [1985] for the statistical properties of mixtures of normals.

Since the regime is assumed to be unobservable, the relevant information set available at time $t-1$ consists only of the observed time series until time t and the unobserved regime vector ξ_t has to be replaced by the inference $\Pr(\xi_t|Y_\tau)$. These probabilities of being in regime m given an information set Y_τ are denoted $\xi_{mt|\tau}$ and collected in the vector $\hat{\xi}_{t|\tau}$ as

$$\hat{\xi}_{t|\tau} = \begin{bmatrix} \Pr(\xi_t = \iota_1|Y_\tau) \\ \vdots \\ \Pr(\xi_t = \iota_M|Y_\tau), \end{bmatrix}$$

which allows two different interpretations. First, $\hat{\xi}_{t|\tau}$ denotes the discrete conditional probability distribution of ξ_t given Y_τ. Secondly, $\hat{\xi}_{t|\tau}$ is equivalent to the conditional mean of ξ_t given Y_τ. This is due to the binarity of the elements of ξ_t, which implies that $\mathsf{E}[\xi_{mt}] = \Pr(\xi_{mt} = 1) = \Pr(s_t = m)$. Thus, the conditional probability density of y_t based upon Y_{t-1} is given by

$$\begin{aligned} p(y_t|Y_{t-1}) &= \sum_{m=1}^{M} p(y_t, \xi_{t-1} = \iota_m|Y_{t-1}) \\ &= \sum_{m=1}^{M} p(y_t|\xi_{t-1} = \iota_m, Y_{t-1}) \Pr(\xi_{t-1} = \iota_m|Y_{t-1}) \qquad (1.15) \\ &= \eta_t' \mathbf{P}' \hat{\xi}_{t-1|t-1}. \end{aligned}$$

As with the conditional probability density of a single observation y_t in (1.15) the conditional probability density of the sample can be derived analogously. The techniques of setting-up the likelihood function in practice are introduced in *Section 6.1*. Here we only sketch the basic approach.

Assuming presample values Y_0 are given, the density of the sample $Y \equiv Y_T$ for given states ξ is determined by

$$p(Y|\xi) = \prod_{t=1}^{T} p(y_t|\xi_t, Y_{t-1}). \qquad (1.16)$$

Hence, the joint probability distribution of observations and states can be calculated as

$$\begin{aligned} p(Y,\xi) &= p(Y|\xi)\ \Pr(\xi) \\ &= \prod_{t=1}^{T} p(y_t|\xi_t, Y_{t-1}) \prod_{t=2}^{T} \Pr(\xi_t|\xi_{t-1})\ \Pr(\xi_1). \end{aligned} \tag{1.17}$$

Thus, the unconditional density of Y is given by the marginal density

$$p(Y) = \int p(Y,\xi)\, d\xi, \tag{1.18}$$

where $\int f(x,\xi)d\xi := \sum_{i_1=1}^{M} \cdots \sum_{i_T=1}^{M} f(x, \xi_T = \iota_{i_T}, \ldots, \xi_1 = \iota_{i_1})$ denotes summation over all possible values of $\xi = \xi_T \otimes \xi_{T-1} \otimes \ldots \otimes \xi_1$ in equation (1.18).

Finally, it follows by the definition of the conditional density that the conditional distribution of the total regime vector ξ is given by

$$\Pr(\xi|Y) = \frac{p(Y,\xi)}{p(Y)}.$$

Thus, the desired conditional regime probabilities $\Pr(\xi_t|Y)$ can be derived by marginalization of $\Pr(\xi|Y)$. In practice these cumbrous calculations can be simplified by a recursive algorithm, a matter which is discussed in *Chapter 5*.

The regime probabilities for future periods follow from the exogenous stochastic process of ξ_t, more precisely the Markov property of regimes, $\Pr(\xi_{T+h}|\xi_T, Y) = \Pr(\xi_{T+h}|\xi_T)$,

$$\begin{aligned} \Pr(\xi_{T+h}|Y) &= \sum_{\xi_t} \Pr(\xi_{T+h}|\xi_T, Y)\Pr(\xi_T|Y) \\ &= \sum_{\xi_t} \Pr(\xi_{T+h}|\xi_T)\Pr(\xi_T|Y). \end{aligned}$$

These calculations can be summarized in the simple forecasting rule:

$$\begin{bmatrix} \Pr(s_{T+h} = 1|Y) \\ \vdots \\ \Pr(s_{T+h} = M|Y) \end{bmatrix} = [\mathbf{P}']^h \begin{bmatrix} \Pr(s_T = 1|Y) \\ \vdots \\ \Pr(s_T = M|Y) \end{bmatrix},$$

where $\mathbf{P}$ is the transition matrix as in (1.8). Forecasting MS-VAR processes is discussed in full length in *Chapter 4*.

In this section we have given just a short introduction to some basic concepts related to MS-VAR models; the following chapters will provide broader analyses of the various topics.

1.4 Features of MS-VAR Processes and Their Relation to Other Non-linear Models

The Markov switching vector autoregressive model is a very general approach for modelling time series with changes in regime. In *Chapter 3* it will be shown that MS-VAR processes with shifting means or intercepts but regime-invariant variances and autoregressive parameters can be represented as non-normal linear state space models. Furthermore, MSM-VAR and MSI-VAR models possess linear representations. These processes may be better characterized as *non-normal* than as *non-linear* time series models as the associated Wold representations coincide with those of linear models. While our primary research interest concerns the modelling of the conditional mean, we will exemplify the effects of Markovian switching regimes on the higher moments of the observed time series.

For sake of simplicity we restrict the following consideration mainly to univariate processes

$$y_t = \nu(s_t) + \sum_{j=1}^{p} \alpha_j(s_t) y_{t-j} + u_t, \quad u_t \sim \text{NID}\,(0, \sigma^2(s_t)).$$

Most of them are made for two-regimes. Thus, the process generating y_t can be re-written as

$$\begin{aligned} y_t &= [\nu_2 + (\nu_1 - \nu_2)\xi_{1t}] + \sum_{j=1}^{p}[\alpha_2 + (\alpha_1 - \alpha_2)\xi_{1t}] y_{t-j} + u_t, \\ & \quad u_t \sim \text{NID}\,(0, [\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)\xi_{1t}]). \end{aligned}$$

If the regime s_t is governed by a Markov chain, the MS(2)-AR(p) model ensues. It will be shown that even such simple MS-AR models can encompass a wide spectrum of modifications of the time-invariant normal linear time series model.

1.4.1 Non-Normality of the Distribution of the Observed Time Series

As already seen the conditional densities $p(y_t|Y_{t-1})$ are a mixture of M normals $p(y_t|\xi_t, Y_{t-1})$ with weights $p(\xi_t|Y_{t-1})$:

$$p(y_t|Y_{t-1}) = \sum_{m=1}^{M} \hat{\xi}_{mt|t-1}\varphi\left(\sigma^{-1}(y_t - \bar{y}_{mt})\right)$$

where $\varphi(\cdot)$ is a standard normal density and $\bar{y}_{mt} = \mathsf{E}[y_t|\xi_t = \iota_m, Y_{t-1}]$. Therefore the distribution of the observed time series can be multi-modal. Relying on well-known results, cf. *e.g.* TITTERINGTON *et al.* [1985, p. 162], we can notice for $M = 2$:

Example 1 *An MS(2)-AR(p) process with a homoskedastic Gaussian innovation process* $u_t \sim \text{NID}(0, \sigma^2)$ *generates bimodality of the conditional density* $p(y_t|Y_{t-1})$ *if*

$$\sigma^{-1}(\bar{y}_{1t} - \bar{y}_{2t}) > \Delta_{\bar{\xi}_1} \geq 2,$$

where the critical value $\Delta_{\bar{\xi}_1}$ *depends on the ergodic regime probability* $\bar{\xi}_1$, *e.g.* $\Delta_{0.5} = 2$ *and* $\Delta_{0.1} = \Delta_{0.9} = 3$.

In contrast to Gaussian VAR processes, MS-VAR models can produce skewness (non-zero third-order cross-moments) and leptokurtosis (fat tails) in the distribution of the observed time series. A simple model that generates leptokurtosis in the distribution of the observed time series y_t is provided by the MSH(2)-AR(0) model:

Example 2 *Let* y_t *be an MSH(2)-AR(0) process,*

$$y_t - \mu = u_t, \quad u_t \sim \text{NID}\,(0, \sigma_1^2 I(s_t = 1) + \sigma_2^2 I(s_t = 2)).$$

Then it can be shown that the excess kurtosis is given by

$$\frac{\mathsf{E}[(y_t - \mu)^4]}{\mathsf{E}[(y_t - \mu)^2]^2} - 3 = \frac{3\bar{\xi}_1\bar{\xi}_2(\sigma_1^2 - \sigma_2^2)^2}{(\bar{\xi}_1\sigma_1^2 + \bar{\xi}_2\sigma_2^2)^2}.$$

Thus, the excess kurtosis is different from zero if $\sigma_1^2 \neq \sigma_2^2$ *and* $0 < \bar{\xi}_1 < 1$.

BOX & TIAO [1968] have used such a model for the detection of outliers. In order to generate skewness and excess kurtosis it is *e.g.* sufficient to assume an MSI(2)-AR(0) model:

Example 3 *Let y_t be generated by an MSM(2)-AR(0) process:*

$$y_t - \mu = (\mu_1 - \mu)I(s_t = 1) + (\mu_2 - \mu)I(s_t = 2) + u_t, u_t \sim \text{NID}\,(0, \sigma^2),$$

so that

$$y_t - \mu = (\mu_2 - \mu) + (\mu_1 - \mu_2)\xi_{1t} + u_t.$$

Then it can be shown that the normalized third moment of y_t is given by the skewness

$$\frac{\mathsf{E}[(y_t - \mu)^3]}{\mathsf{E}[(y_t - \mu)^2]^{3/2}} = \frac{(\mu_1 - \mu_2)^3(1 - 2\bar{\xi}_1)\bar{\xi}_1(1 - \bar{\xi}_1)}{\left(\sigma^2 + (\mu_1 - \mu_2)^2\bar{\xi}_1(1 - \bar{\xi}_1)\right)^{3/2}}.$$

If the regime i with the highest conditional mean $\mu_i > \mu_j$ is less likely than the other regime, $\bar{\xi}_i < \bar{\xi}_j$, then the observed variable is more likely to be far above the mean than it is to be far below the mean.

Furthermore the normalized fourth moment of y_t is given by the excess kurtosis

$$\frac{\mathsf{E}[(y_t - \mu)^4]}{\mathsf{E}[(y_t - \mu)^2]^2} - 3 = \frac{(\mu_1 - \mu_2)^4\bar{\xi}_1(1 - \bar{\xi}_1)\left\{1 - 3\bar{\xi}_1(1 - \bar{\xi}_1)\right\}}{\left(\sigma^2 + (\mu_1 - \mu_2)^2\bar{\xi}_1\bar{\xi}_2\right)^2}.$$

Since we have that $\max_{\bar{\xi}_1 \in [0,1]}\{\bar{\xi}_1(1 - \bar{\xi}_1)\} = \frac{1}{4} < \frac{1}{3}$, the excess kurtosis is positive, i.e. the distribution of y_t has more mass in the tails than a Gaussian distribution with the same variance.

The combination of regime switching means and variances in an MSIH(2)-AR(0) process (cf. *Example 4*) is given in SOLA & TIMMERMANN [1995]. The implications for option pricing are discussed in KÄHLER & MARNET [1994b]. For an MSMH(2)-AR(4) model, the conditional variance of the one-step prediction error is given by SCHWERT [1989] and PAGAN & SCHWERT [1990].

1.4.2 Regime-dependent Variances and Conditional Heteroskedasticity

An MS(M)-AR(p) process is called *conditional heteroskedastic* if the conditional variance of the prediction error $y_t - \mathsf{E}[y_t|Y_{t-1}]$,

$$\text{Var}\,[y_t|Y_{t-1}] = \mathsf{E}\left\{(y_t - \mathsf{E}[y_t|Y_{t-1}])^2\right\}$$

is a function of the information set Y_{t-1} . Conditional heteroskedasticity can be induced by regime-dependent variances, autoregressive parameters or means.

In MS-AR models with regime-invariant autoregressive parameters, conditional heteroskedasticity implies that the conditional variance of the prediction error $y_t - \mathsf{E}[y_t|Y_{t-1}]$, is a function of the filtered regime vector $\hat{\xi}_{t-1|t-1}$. In general, an MS-AR process is called *regime-conditional heteroskedastic* if

$$\mathrm{Var}\,[y_t|\xi_{t-1}, Y_{t-1}] = \mathsf{E}\left\{(y_t - \mathsf{E}[y_t|\xi_{t-1}, Y_{t-1}])^2\right\}$$

is a function of ξ_{t-1}. Interestingly, regime-dependent variances are neither necessary nor sufficient for conditional heteroskedasticity. As stated in *Chapter 3*, a necessary and sufficient condition for conditional heteroskedasticity in MS-VAR models with regime-invariant autoregressive parameters is the serial dependence of regimes.

On the other hand, even if the white noise process u_t is homoskedastic, $\sigma^2(s_t) = \sigma^2$, the observed process y_t can be heteroskedastic. Consider the following example:

Example 4 *Let y_t be an MSI(2)-AR(0) process*

$$y_t - \mu \quad = \quad (\mu_1 - \mu)I(s_t = 1) + (\mu_2 - \mu)I(s_t = 2) + u_t,$$

with $u_t \sim$ NID $(0, \sigma^2)$ and serial correlation in the regimes according to the transition matrix $\mathbf{P}$. *Employing the ergodic regime probability $\bar{\xi}_1$, y_t can be written as*

$$y_t - \mu = (\mu_1 - \mu_2)(\xi_{1t} - \bar{\xi}_1) + u_t.$$

Thus $\mathsf{E}[y_t|Y_{t-1}] = \mu + (\mu_1 - \mu_2)(\hat{\xi}_{1t|t-1} - \bar{\xi}_1)$ *and*

$$\begin{aligned}
\mathrm{Var}\,[y_t|Y_{t-1}] &= \sigma^2 + (\mu_1 - \mu_2)^2\mathsf{E}\left[(\xi_{1t} - \bar{\xi}_1)^2|Y_{t-1}\right] \\
&= \sigma^2 + (\mu_1 - \mu_2)^2\left[\hat{\xi}_{1t|t-1}(1 - \hat{\xi}_{1t|t-1})^2 + (1 - \hat{\xi}_{1t|t-1})(-\hat{\xi}_{1t|t-1})^2\right] \\
&= \sigma^2 + (\mu_1 - \mu_2)^2\hat{\xi}_{1t|t-1}(1 - \hat{\xi}_{1t|t-1}),
\end{aligned}$$

where $\hat{\xi}_{1t|t-1} = p_{11}\hat{\xi}_{1t-1|t-1} + p_{21}(1 - \hat{\xi}_{1t-1|t-1}) = (p_{11} + p_{22} - 1)\hat{\xi}_{1t-1|t-1} + (1 - p_{22})$ *is the predicted regime probability* $\Pr(s_t = 1|Y_{t-1})$. *Thus $\{y_t\}$ is a regime-conditional heteroskedastic process.*

In contrast to ARCH models, the conditional variance in MS-VAR models (with time-invariant autoregressive parameters) is a non-linear function of past squared errors since the predicted regime probabilities generally are non-linear functions of Y_{t-1}.

Recently some approaches have been made to consider Markovian regime shifts in *variance generating processes.* The class of *autoregressive conditional heteroskedastic processes* introduced by ENGLE [1982] is used to formulate the conditional process; our assumption of an *i.i.d.* distributed error term is substituted by an ARCH process u_t, cf. *inter alia* HAMILTON & LIN [1994], HAMILTON & SUSMEL [1994], CAI [1994] and HALL & SOLA [1993b]. ARCH effects can be generated by MSA-AR processes which will be considered in the next section.

1.4.3 Regime-dependent Autoregressive Parameters: ARCH and Stochastic Unit Roots

Autoregressive conditional heteroskedasticity is known from random coefficient models. Therefore it is not very surprising that also MSA-VAR models may lead to ARCH. This effect will be considered in the following simple example.

Example 5 *Let y_t be generated by an MSA(2)-MAR(1) process with i.i.d. regimes:*

$$(y_t - \mu) = \alpha(s_t)\,(y_{t-1} - \mu) + u_t, \quad u_t \sim \text{NID}\,(0, \sigma^2).$$

Serial independence of the regimes implies $p_{11} = 1 - p_{22} = \rho$; the regime-dependent autoregressive parameters α_1, α_2 are restricted such that $\mathsf{E}[\alpha] = \alpha_1\rho + \alpha_2(1-\rho) = 0$. Thus it can be shown that

$$\begin{aligned} \mathsf{E}[y_t|Y_{t-1}] &= \mu + (\alpha_1\rho + \alpha_2(1-\rho))\,y_{t-1} = \mu, \\ \mathsf{E}[(y_t-\mu)^2|Y_{t-1}] &= \sigma^2 + (\alpha_1^2\rho + \alpha_2^2(1-\rho))\,(y_{t-1}-\mu)^2. \end{aligned}$$

Then y_t possesses an ARCH representation $y_t = \mu + e_t$ with

$$e_t^2 = \sigma^2 + \gamma e_{t-1}^2 + \varepsilon_t$$

where $\gamma = -\alpha_1\alpha_2 > 0$ and ε_t is white noise. Thus, ARCH(1) models can be interpreted as restricted MSA(2)-AR(1) models.

The theoretical foundations of MSA-VAR processes are laid in TJØSTHEIM [1986b]. Some independent theoretical results are provided by BRANDT [1986]. As pointed out by TJØSTHEIM [1986b], the dynamic properties of models with regime-dependent autoregressive parameters are quite complicated. Especially, if the process is stationary for some regimes and mildly explosive for others, the problems of stochastic unit root processes as introduced by GRANGER & SWANSON [1994] are involved.[10]

It is worth noting that the stability of each VAR sub-model and the ergodicity of the Markov chain are sufficient stability conditions; they are however not necessary to establish stability. Thus, the stability of MSA-AR models can be compatible with AR polynomials containing in some regimes roots greater than unity in absolute value and less than unity in others. Necessary and sufficient conditions for the stability of stochastic processes as the MSA-VAR model have been derived in KARLSEN [1990a], [1990b]. However in practice, their application has been found to be rather complicated (cf. HOLST *et al.* [1994]).

In this study we will concentrate our analysis on modelling shifts in the (conditional) mean and the variance of VAR processes which simplifies the analysis.

1.5 Conclusion and Outlook

In the preceding discussion of this chapter MS(M)-VAR(p) processes have been introduced as doubly stochastic processes where the conditional stochastic process is a Gaussian VAR(p) and the regime generating process is a Markov chain. As we have seen in the discussion of the relationship of the MS-VAR model to other non-linear models, the MS-VAR model can encompass many other time series models proposed in the literature or replicates at least some of their features. In the following chapter these considerations are formalized to state-space representations of MS-VAR models where the measurement equation corresponds to the conditional stochastic process and the transition equation reflects the regime generating process. In *Section 2.5* the MS-VAR model will be compared to time-varying coefficient models with smooth variations in the parameters, *i.e.* an infinite number of regimes.

[10] Models where the regime is switching between deterministic and stochastic trends are considered by MCCULLOCH & TSAY [1994a].

1.A Appendix: A Note on the Relation of SETAR to MS-AR Processes

While the presumptions of the SETAR and the MS-AR model seem to be quite different, the relation between both model alternatives is rather close. Indeed, both models can be observationally equivalent, as the following example demonstrates:

Example 6 *Consider the SETAR model*

$$y_t = \mu_2 + (\mu_1 - \mu_2)I(y_{t-d} \leq r) + u_t, \quad u_t \sim \text{NID}\,(0, \sigma^2). \tag{1.19}$$

For $d = 1$ it has been shown by CARRASCO *[1994, lemma 2.2] that (1.19) is a particular case of the Markov-switching model*

$$y_t = \mu_2 + (\mu_1 - \mu_2)I(s_t = 1) + u_t, \quad u_t \sim \text{NID}\,(0, \sigma^2),$$

which is an MSI(2)-AR(0) model. For an unknown r, define the unobserved regime variable s_t as the binary variable

$$s_t = I(y_{t-1} \leq r) = \begin{cases} 1 & \text{if } y_{t-1} \leq r \\ 2 & \text{if } y_{t-1} > r \end{cases}$$

such that

$$\begin{aligned}
\Pr(s_t = 1|s_{t-1}, Y) &= \Pr(y_{t-1} \leq r|s_{t-1}, Y) \\
&= \Pr(\mu_2 + (\mu_1 - \mu_2)I(s_{t-1} = 1) + u_{t-1} \leq r) \\
&= \Pr(u_{t-1} \leq r - \mu_2 - (\mu_1 - \mu_2)I(s_{t-1} = 1)) \\
&= \Phi\left(\frac{r - \mu_2 - (\mu_1 - \mu_2)I(s_{t-1} = 1)}{\sigma}\right) \\
&= \Pr(s_t = 1|s_{t-1}).
\end{aligned}$$

Hence s_t follows a first order Markov process where the transition matrix is defined as

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} \Phi(\frac{r-\mu_1}{\sigma}) & \Phi(\frac{\mu_1-r}{\sigma}) \\ \Phi(\frac{r-\mu_2}{\sigma}) & \Phi(\frac{\mu_2-r}{\sigma}) \end{bmatrix}.$$

If $d > 1$, the data can be considered as generated by d independent series which are each particular Markov processes. A proof can be based on the property $\Pr(s_t|\{s_{t-j}\}_{j=1}^{\infty}, Y_T) = \Pr(s_t|s_{t-2}, Y_T)$; thus s_t follows a second order Markov chain, which can be reparametrized as a higher dimensional first order Markov chain.