9 Markov Switching Models

On-Line Ex. 9A. We use the same monthly US excess returns data as in Example 9.1 in the main text (the sample is 1986:01 – 2016:12) to compare the fit and the estimates of two simple MSIH(2,0)-type models. The first is the same first-order Markov switching model already analyzed (*p*-values are in parenthesis):

$$x_{t+1}^{US} = \begin{cases} \underset{(0.000)}{1.096} + \underset{(0.000)}{2.649} z_{t+1}^{US} & \text{if } S_{t+1}^{US} = 1 \text{ (bull)} \\ \underset{(0.751)}{0.206} + \underset{(0.000)}{5.662} z_{t+1}^{US} & \text{if } S_{t+1}^{US} = 2 \text{ (bear)} \end{cases} \qquad \hat{P}^{US} = \begin{bmatrix} 0.966 & 0.034 \\ 0.035 & 0.965 \end{bmatrix}$$

This model features persistent states and implies a maximized log-likelihood of -1051.8 and information criteria of 5.687, 5.712, and 5.751 for AIC, Hannan-Quinn, and BIC, respectively. Next, we re-estimate the model imposing the restriction that $\Pr\left(S_{t+1} = j \mid S_t = i\right) = \Pr\left(S_{t+1} = j \mid S_t = j\right) = p_j$ or, equivalently, that we a simple IID switching model may be adequate. We find:

$$x_{t+1}^{US} = \begin{cases} \underset{(0.000)}{1.451} + \underset{(0.000)}{3.449} z_{t+1}^{US} & \text{if } S_{t+1}^{US} = 1 \text{ (bull)} \\ \underset{(0.239)}{-3.224} + \underset{(0.000)}{6.294} z_{t+1}^{US} & \text{if } S_{t+1}^{US} = 2 \text{ (bear)} \end{cases} \qquad \hat{P}^{US} = \begin{bmatrix} 0.829 & 0.171 \\ 0.829 & 0.171 \end{bmatrix}$$

The estimated transition matrix is radically different and the second bear regime has lost any persistence. This model achieves a maximized log-likelihood of -1065.7, considerably inferior to a MC with memory, and information criteria of 5.757, 5.777, and 5.809 for AIC, Hannan-Quinn, and BIC, respectively: these all exceed the criteria reported for the persistent MS model which is then to be preferred.

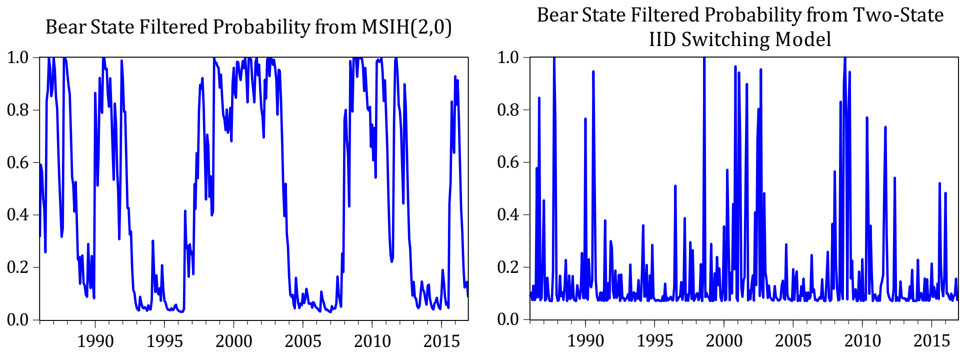

Figure 9.A1 –Bear State Probabilities from Two MS Models

The two plots compare the (filtered) probabilities of a bear market state deriving from two models to reveal that imposing such "hard" restrictions on the transition matrix as those implied by an IID mixture may severely change the economic meaning of the estimates obtained. Because a simple switching framework characterizes the bear regime as rather episodical and non-persistent (see the right plot), while a Markov switching framework reveals that the US stock market would visit this state often (almost 50% of the sample), the opposite derives from the simple switching case. This difference may have important effects in applied work and practical applications of the model.

On-Line Supp. 9B. (**Misspecification Tests Applied to MS Models**) Once a restricted set of (or more simply, one) MS models has been estimated, either the need of further improvements could arise as the result of a few diagnostic checks or the best model will be chosen based on the success of such checks. Although the EM algorithm naturally delivers estimates of the parameters $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\xi}}_{1|0}$ besides the smoothed sequence of state probabilities $\{\hat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^{T}$ and would therefore lead to define the (smoothed) residuals as $\hat{\boldsymbol{\varepsilon}}_{t|T} = \boldsymbol{y}_t - \mathbf{X}_t \hat{\boldsymbol{A}} \hat{\xi}_{t|T}$, t = 1, 2, ..., T, these are not well suited to the use in diagnostic checks as they are full-sample statistics and hence they structurally overestimate the explanatory power of a MS model. On the contrary, the **one-step ahead prediction errors**,

$$\hat{\boldsymbol{\varepsilon}}_{t|t-1} = \boldsymbol{y}_t - \mathbf{X}_t \hat{\boldsymbol{A}} \hat{\boldsymbol{P}}' \hat{\boldsymbol{\xi}}_{t-1|t-1} = \boldsymbol{y}_t - \mathbf{X}_t \hat{\boldsymbol{A}} \hat{\boldsymbol{\xi}}_{t|t-1}, \qquad (9B.1)$$

are limited information statistics (being based on filtered probabilities) and uncorrelated with the information set $\Im_{t-1}$ because $E[\boldsymbol{y}_t | \Im_{t-1}] = \mathbf{X}_t \hat{\boldsymbol{A}} \mathbf{P}' \hat{\boldsymbol{\xi}}_{t-1|t-1}$ and therefore the prediction errors in (9.B1) form a **martingale difference sequence**, i.e.:

$$E[\hat{\boldsymbol{\varepsilon}}_{t|t-1} | \Im_{t-1}] = E[\boldsymbol{y}_t | \Im_{t-1}] - \mathbf{X}_t \hat{\boldsymbol{A}} \hat{\boldsymbol{P}}' \hat{\boldsymbol{\xi}}_{t-1|t-1} = \boldsymbol{0}. \qquad (9.B2)$$

In the case of heteroskedastic MS models, such prediction errors shall need to be appropriately standardized:

$$\hat{\boldsymbol{z}}_{t|t-1} = \sum_{k=1}^{K} (\boldsymbol{e}'_k \hat{\boldsymbol{P}}' \hat{\boldsymbol{\xi}}_{t-1|t-1}) \hat{\boldsymbol{\varepsilon}}_{t|t-1} = \sum_{k=1}^{K} \hat{\xi}_{t|t-1}^{k} (\boldsymbol{y}_t - \mathbf{X}_t \hat{\boldsymbol{A}} \hat{\boldsymbol{P}}' \hat{\boldsymbol{\xi}}_{t-1|t-1}). \quad (9.B3)$$

Therefore standard tests of the hypothesis in (9.B2) or $E[\hat{\boldsymbol{z}}_{t|t-1} | \Im_{t-1}] = \boldsymbol{0}$ (such as Portmanteau tests of no serial correlation) could be used in order to detect any deviation from the martingale structure.[1] Here, $E[\hat{\boldsymbol{z}}_{t|t-1} | \Im_{t-1}] = \boldsymbol{0}$ means that none of the information contained in $\Im_{t-1}$ can help to forecast subsequent prediction errors, so that (9.B2) implies the possibility of testing restrictions such as $E[\hat{\boldsymbol{z}}_{t|t-1}\hat{\boldsymbol{z}}'_{t-h|t-h-1} | \Im_{t-1}] = \mathbf{O}$, $E[\hat{\boldsymbol{z}}_{t|t-1} g(\hat{\boldsymbol{z}}'_{t-h|t-h-1}) | \Im_{t-1}] = \mathbf{O}$ or $E[\hat{\boldsymbol{z}}_{t|t-1} q(\Im_{t-1})] = \mathbf{O}$ $\forall h \geq 1$ where $g(\cdot)$ is any smooth function from $\mathbb{R}^N$ to $\mathbb{R}^N$ and $q(\cdot)$ is any function that extracts information from $\Im_{t-1}$. Importantly, even though the MS model may have assumed conditional normality of the errors, there is no presumption that the one-step ahead forecast errors be normally distributed, as they are themselves mixtures of normal densities.

Finally, common sense suggests that correct specification of a MS model should give smoothed probability distributions $\{\hat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^{T}$ that consistently signal switching among states with only limited periods in which the associated distribution is flatly spread out over the entire support and uncertainty dominates. **Regime Classification Measures** (RCMs) have been popularized as a way to assess whether the number of regimes K is adequate. In simple two-regime frameworks, the early work by Hamilton (1988) offered a rather intuitive regime classification measure,

$$RCM_1 = 100\frac{K^2}{T}\sum_{t=1}^{T}\prod_{k=1}^{K}\hat{\xi}_{t|T}^{k}, \tag{9.B4}$$

i.e., the sample average of the products of the smoothed state probabilities. Clearly, when a MS model offers precise indications on the nature of the regime at each time t, the implication is that for at least one value of $k = 1,...,K$, $\hat{\xi}_{t|T}^{k} \simeq 1$ so that $\sum_{k=1}^{K}\hat{\xi}_{t|T}^{k} \simeq 0$ because most other smoothed probabilities will be zero. Therefore a good MS model will imply $RCM_1 = 0$.[2] However, when applied to models with

---

[1] With the caveat that that the one-step ahead prediction errors do not possess a Gaussian density and hence the approximate validity of Portmanteau standard tests can only be guessed.

[2] On the opposite, the worst possible MS model will have

$K > 2$, RCM_1 has one obvious disadvantage: a model can imply an enormous degree of uncertainty on the current regime, but still imply $\sum_{k=1}^{K} \hat{\xi}_{t|T}^{k} \simeq 0$ for most values of t. For instance, when $K = 3$, it is easy to see that if $\hat{\xi}_{t|T}^{1} = 1/2$, $\hat{\xi}_{t|T}^{2} = 1/2$, and $\hat{\xi}_{t|T}^{3} = 0$ $\forall t$, then $RCM_1 = 0$ even though this remains a rather uninformative switching model to use in practice. As a result, it is rather common to witness that as K exceeds 2, almost all switching models (good and bad) will automatically imply values of RCM_1 that are very close to 0. Guidolin (2009) proposes a number of alternative measures that may shield against this type of problems, for instance

$$RCM_2 = 100\left[1 - \frac{K^{2K}}{(K-1)^2}\frac{1}{T}\sum_{t=1}^{T}\prod_{k=1}^{K}\left(\hat{\xi}_{t|T}^{k} - \frac{1}{K}\right)^2\right] \qquad (9.B5)$$

We re-examine Example 9.4 in the textbook to ask whether the MS model previously selected passes a few misspecification tests. We analyze the residuals and standardized residuals from the best fitting, after penalizing for the size of the parameter vector to be estimated, MSIH(3) regression model that has emerged from our earlier work. The various panels of Table 9.B1 analyze the sample ACFs and the associated Ljung-Box statistics (up to order 12) for the level of the residuals, the square and absolute value of the standardized residuals, and the cross-sample ACF between the standardized residuals and the two regressors.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	P-value
		1	0.103	0.103	17.051	0.000
		2	-0.026	-0.037	18.124	0.000
		3	0.045	0.052	21.384	0.000
		4	-0.007	-0.019	21.467	0.000
		5	-0.005	0.001	21.510	0.001
		6	-0.010	-0.013	21.684	0.001
		7	-0.016	-0.012	22.092	0.002
		8	-0.012	-0.010	22.344	0.004
		9	-0.035	-0.033	24.347	0.004
		10	-0.022	-0.015	25.162	0.005
		11	-0.014	-0.012	25.462	0.008
		12	-0.020	-0.016	26.098	0.010

Table 9.B1, panel (a) – SACF of MSIH(3) One-Week Prediction Errors

Panel (a) shows that there is some statistically significant

$\hat{\xi}_{t|T}^{1} = \dots = \hat{\xi}_{t|T}^{K} = 1/K$ so that $\sum_{k=1}^{K}\hat{\xi}_{t|T}^{k} = 1/K^2$ and $RCM_1 = 100$. Therefore $RCM_1 \in [0,100]$ and lower values are to be preferred to higher ones.

autocorrelation left in the model residuals, even though this is limited to a first-order pattern.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	P-value
		1	-0.008	-0.008	0.0948	0.758
		2	0.009	0.009	0.2152	0.898
		3	0.043	0.044	3.2687	0.352
		4	-0.002	-0.001	3.2757	0.513
		5	-0.002	-0.003	3.2827	0.656
		6	0.041	0.039	5.9888	0.424
		7	0.045	0.046	9.2264	0.237
		8	0.033	0.033	10.990	0.202
		9	0.035	0.032	12.976	0.164
		10	0.030	0.027	14.464	0.153
		11	-0.021	-0.023	15.167	0.175
		12	0.046	0.041	18.558	0.100

Table 9.B1, panel (b) – SACF of MSIH(3) Squared Standardized One-Week Prediction Errors

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	P-value
		1	-0.013	-0.013	0.2535	0.615
		2	-0.013	-0.013	0.5112	0.774
		3	0.022	0.022	1.3307	0.722
		4	0.025	0.025	2.3506	0.672
		5	0.003	0.004	2.3626	0.797
		6	0.045	0.045	5.6214	0.467
		7	0.041	0.041	8.2892	0.308
		8	0.025	0.026	9.2656	0.320
		9	0.046	0.046	12.721	0.176
		10	0.046	0.045	16.210	0.094
		11	-0.010	-0.011	16.375	0.128
		12	0.046	0.042	19.773	0.071

Table 9.B1, panel (c) – SACF of MSIH(3) Absolute Standardized One-Week Prediction Errors

In panels (b) and (c), we have instead evidence that any heteroskedasticity patterns are well captured by the three state model. This is consistent with $E[\hat{\mathbf{z}}_{t|t-1} g(\hat{\mathbf{z}}'_{t-h|t-h-1}) | \Im_{t-1}] = \mathbf{O}$.

RESID_MS3,VXO(-i)	RESID_MS3,VXO(+i)	i	lag	lead
		0	0.0649	0.0649
		1	-0.0069	0.0260
		2	-0.0052	0.0164
		3	-0.0082	0.0148
		4	0.0013	0.0013
		5	-0.0050	0.0066
		6	0.0049	-0.0093
		7	-0.0038	-0.0058
		8	0.0120	-0.0176
		9	-0.0236	-0.0098
		10	-0.0275	0.0129
		11	-0.0471	0.0033
		12	-0.0541	-0.0033

RESID_MS3,TERM(-i)	RESID_MS3,TERM(+i)	i	lag	lead
		0	-0.0773	-0.0773
		1	-0.0007	-0.0926
		2	0.0029	-0.0836
		3	0.0029	-0.0896
		4	0.0216	-0.0805
		5	0.0253	-0.0768
		6	0.0314	-0.0684
		7	0.0292	-0.0574
		8	0.0168	-0.0516
		9	0.0099	-0.0452
		10	0.0049	-0.0406
		11	0.0061	-0.0337
		12	0.0028	-0.0279

Table 9.B1, panels (d)-(e) – Sample Cross-ACF of MSIH(3) One-Week Prediction Errors

In panels (d)-(e) we have evidence that both regressors at time *t* are correlated with one-week prediction errors between time *t*-1 and *t*, which is normal. However, lagged regressors fail to forecast prediction errors which is consistent with $E[\hat{\mathbf{z}}_{t|t-1} q(\Im_{t-1})] = \mathbf{O}$. On the contrary, we are not worried about the fact that lagged prediction errors appear to precisely predict the subsequent values of the regressors (in particular the term spread), even though this may represent evidence in favor of adopting a fully multivariate strategy based on the estimation of MSVARH models, in which also excess bond returns predict subsequent VXO and term spread values and this is taken into account.

Finally, using a tool that has been introduced in Chapter 5, we have also applied to the one-step ahead prediction errors the Brock, Dechert, Scheinkman and LeBaron's (1996)portmanteau test of independence, as $E[\hat{\mathbf{z}}_{t|t-1} q(\Im_{t-1})] = \mathbf{O}$ and $E[\hat{\mathbf{z}}_{t|t-1} g(\hat{\mathbf{z}}'_{t-h|t-h-1}) | \Im_{t-1}]$ $= \mathbf{O}$ also imply independence. We select the BDS parameter δ selected to be one standard deviation of the residuals and a maximum *m*=6. Because we apply the tests to the standardized prediction errors, we compute *p*-values using a bootstrap with 20,000 repetitions. We find that for all values of *m* between 2 and 6, the null hypothesis of IIDness is never formally rejected, with the smallest *p*-value of 0.056 for *m*=6.

Of course, the BDS finding may depend on the fact that some first-order serial correlation had been left in the residuals. Therefore we proceed to re-estimate the MSIH(3) regression with a time-invariant AR(1) term added:

$$x_{t+1}^{10Y} = \begin{cases} \underset{(0.000)}{-0.376} & \text{if } S_{t+1}=1 \\ \underset{(0.002)}{-0.712} & \text{if } S_{t+1}=2 \\ \underset{(0.062)}{-0.651} & \text{if } S_{t+1}=3 \end{cases} + \underset{(0.000)}{0.212} x_t^{10Y} + \underset{(0.000)}{0.113} spread_t + \underset{(0.155)}{0.008} VXO_t +$$

$$+ \begin{cases} \underset{(0.000)}{0.782} & \text{if } S_{t+1}=1 \\ \underset{(0.000)}{1.111} & \text{if } S_{t+1}=2 \\ \underset{(0.000)}{2.069} & \text{if } S_{t+1}=3 \end{cases} \times \varepsilon_{t+1}^{10Y} \qquad \hat{\boldsymbol{P}} = \begin{bmatrix} 0.977 & 0.016 & 0.007 \\ 0.022 & 0.974 & 0.004 \\ 0.000 & 0.069 & 0.931 \end{bmatrix}$$

The AR(1) term is now highly significant, but it seems to take away some of the accuracy in estimation of the VXO coefficient and takes us to the usual problem: under the efficient market hypothesis and many asset pricing models, past excess returns and term spread slopes should not forecast subsequent ones (but empirically they do), while past variance should forecast excess returns, and empirically they seem to! Let's now whether the resulting one-week ahead prediction errors appear now to be serially uncorrelated and independent.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	P-value
		1	0.002	0.002	0.0078	0.930
		2	-0.031	-0.031	1.5683	0.457
		3	0.074	0.074	10.445	0.015
		4	-0.020	-0.022	11.100	0.025
		5	-0.001	0.004	11.102	0.049
		6	-0.010	-0.017	11.256	0.081
		7	0.001	0.005	11.259	0.128
		8	-0.011	-0.013	11.473	0.176
		9	-0.030	-0.028	12.971	0.164
		10	-0.017	-0.018	13.436	0.200
		11	-0.019	-0.019	14.031	0.231
		12	-0.042	-0.039	16.856	0.155

Table 9.B2 – SACF of MSIH(3)-AR(1) One-Week Ahead Prediction Errors

In Table 9.B2, we see that the improvement is visible, even though a third-order lag which may possibly be attributed to sampling variation (i.e., bad luck). The same BDS test applied before, based on a bootstrap with 20,000 repetitions, yields no rejections of the null hypothesis of IIDness, with the smallest p-value of 0.093 for m=3.
Finally, we perform RCM calculations for the three-state model just estimated and compare it with the other models estimated before. Table 9.B3 reports the results. Visibly, RCM_1 drops to almost zero for all models with 3 regimes, irrespective of their actual regime

classification quality that is instead revealed by RCM_2. Moreover, it appears that the last MS regression estimated, which also included an autoregressive term, does offer the best possible regime classification quality and this may be taken as a positive indication of strong model classification, even though an RCM_2 of 39.6 is less than impressive in absolute terms.

Model	K	Switching Regressors	Non-Switching Regressors	RCM1	RCM2
MSI(2)-Regress	2	Const, Term spread, VXO	__	43.45	56.82
MSIH(2)-Regress	2	Const, Term spread, VXO	__	37.68	49.57
MSIH(2)-Regress	2	Const	Term spread, VXO	37.58	49.42
MSIH(3)-Regress	3	Const, Term spread, VXO	__	1.19	66.79
MSIH(3)-Regress	3	Const	Term spread, VXO	1.68	57.09
MSIH(3)-Regress	3	Const, Excess(-1), Term spread, VXO	__	3.84	87.09
MSIH(3)-Regress	3	Const, Excess(-1)	Term spread, VXO	2.23	58.42
MSIH(3)-Regress	3	Const	Excess(-1), Term spread, VXO	**0.72**	**39.58**

Table 9.B3 – Regime Classification Measures for a Variety of MSI(AR)H(K) Regression Models

-Line Supp.9C. (**The Risk-Return Trade-Off**) Despite its key role in many applications, estimating and understanding the dynamics over time of the market risk premium has proven difficult. The market risk premium can be defined as the mean of market returns in excess of some risk-free rate, say $E[R_{t+1} - R^f]$. For instance, even though classical finance theory suggests estimating the risk premium based on the theoretical relationship between mean returns and the contemporaneous variance of returns, for a long time empirical research has failed to document a significantly positive relationship between average returns and the filtered/predicted levels of market volatility (see, e.g., Glosten, Jagannathan, and Runkle, 1993). In fact, a number of researchers have instead unveiled a negative relationship between volatility and market prices, the so-called **volatility feedback effect**. As already discussed in Chapter 5 (where it was called leverage effect), this feedback refers to the intuitive idea that an exogenous change in the level of market volatility initially generates additional return volatility as stock prices adjust in response to new information about future discounted expected returns.

Because the aggregate stock market portfolio remains one of the most natural starting points to an understanding of asset pricing phenomena, it is surprising that there is still a good deal of controversy around the issue of how to measure risk at the market level. Recent empirical studies have documented two puzzling results.

First, there is evidence of a weak, or even negative, relation between conditional mean returns and the conditional volatility of returns. Second, they document significant time variation in this relation. For instance, in a modified GARCH-in mean framework using post-World War II monthly data, Glosten, Jagannathan, and Runkle (1993) find that the estimated coefficient on volatility in a return/volatility regression is negative: a higher conditional volatility would depress the conditional risk premium, not the opposite. Or, equivalently, negative news that depress the risk premium, would increase conditional variance.

More recently, Lettau and Ludvigson (2001) have provided evidence suggesting the failure to find a positive relationship between excess returns and market volatility may result from not controlling for shifts in investment opportunities, i.e., regimes. However, within applications of MS models to financial economics, this idea dates back at least to a seminal paper by Turner, Startz and Nelson (1989, henceforth TSN). TSN introduce a model of the aggregate market portfolio (the Standard and Poor's index) in which *excess* returns are drawn from a mixture of two normal densities because market portfolio returns are assumed to switch between two states. The regimes are characterized by the variances of their densities as a high-variance state and a low-variance state. The state itself is assumed to be generated by a first-order Markov process,

$$x_t = \mu_t + \varepsilon_t \quad \varepsilon_t \; NID(0,\sigma_{S_t}^2), \qquad (9.C1)$$

where $\sigma_1^2 \geq \sigma_0^2$ and the conditional mean $\mu_t \equiv E[x_t \mid \Im_{t-1}]$ is specified below. Of course this is an odd MSIH(2) model, in the sense that variance shifts with regimes in the usual way but the intercept varies according to some function that also involves the Markov chain S_t.

TSN develop two models based on the heteroskedastic structure discussed above. Each incorporates a different assumption about the agents' information sets. In the first model, economic agents know (because they observe it) the realization of the Markov state process, even though the econometrician does not observe it. There are two risk premia in this specification. The first is the difference between the mean of the distribution in the low-variance state and the riskless return. Agents require an increase in return over the riskless rate to hold an asset with a random return. The second premium is the added return necessary to compensate for increased risk in the high-variance state:

$$E[x_t \mid S_t] = \begin{cases} \mu_0 & \text{if } S_t = 0 \\ \mu_1 & \text{if } S_t = 1 \end{cases}. \qquad (9.C2)$$

The parameter estimates from their model suggest that whereas the first risk premium is positive, the second is negative, $\hat{\mu}_0 > 0$ and $\hat{\mu}_1 < 0$. Monthly data on S&P 500 index returns for 1946-1987 reveal that the two regimes identified by $\sigma_1^2 \geq \sigma_0^2$ and $\hat{\mu}_1 \neq \hat{\mu}_0$ are highly persistent, with average durations of 3 months for the high variance regime and of 43 months for the low variance one. Estimates of this simple MSIH model, in which agents are assumed to know the state, do not support a risk premium that increases with risk, which is puzzling: parameter estimates indicate that agents require an increase in average annual returns over T-bills of approximately 10% to hold the risky asset in *low*-variance periods. The estimates also suggest, however, that the premium declines as the level of risk increases, that is, $\hat{\mu}_1 < \hat{\mu}_0$. Further, not only is $\hat{\mu}_1$ significantly less than $\hat{\mu}_0$, it is also significantly negative. Therefore TSN reject the hypothesis of a risk premium increasing in the variance. As we have seen in Example 9.6, this occurs also with reference to more recent data on the S&P 500.

As already hinted at, misspecification is a likely explanation for TSN's result. If agents are uncertain about the state, so that they are basing their decisions on forecasts of the regime in the following period, estimates assuming they know the state with certainty will be inconsistent. Accordingly, in their second model TSN assume that neither economic agents nor the econometrician observe the states. In each period, agents form probabilities of each possible state in the following period conditional on current and past excess returns, and use these probabilities in making their portfolio choices. Each period, investors update their prior beliefs about that period's state with current information using Bayes' rule, as in Section 5. The parameter of interest is then the increase in expected return necessary to compensate the agents for a given percentage increase in the prior probability of the high-variance state. Agents' portfolio choice may be specified as a simple function of this probability:

$$\mu_t = \alpha + \lambda \Pr(S_t = 1 \mid \Im_{t-1}) \qquad (9.C3)$$

where the constant, α, represents the agents' required excess return for holding an asset in the low-variance state. Note that this is an intuitive and yet ad-hoc model: there is no reason for μ_t to depend linearly on the filtered probability of a high-variance state, $\Pr(S_t = 1 | \Im_{t-1})$. Yet, this simple model means that agents require an increase in the excess return in period *t* when faced with an increase in their prior probability that the high-variance state will prevail in that period, and this intuition is sufficiently sound for the model to represent a starting point. In fact, TSN generalize slightly this model to

$$\mu_t = (1 - S_t)\alpha_0 + S_t \alpha_1 + \lambda \Pr(S_t = 1 | \Im_{t-1}). \qquad (9.C4)$$

TSN are able to sign all the parameters in this simple empirical model. The stock price at time *t* should reflect all available information. This requires that the price at *t* should fall below its value at $t-1$ if some new unfavorable information about fundamentals, such as an increase in variance, arrives between $t-1$ and *t*. This fall is necessary to ensure that the return from time *t* to $t+1$ is expected to be higher than usual so as to compensate stockholders for the added risk. According to this scenario, the return between $t-1$ and *t* will be negative on average for those periods in which adverse information is newly acquired, and positive on average when favorable information is acquired. This means that the coefficient λ attached to $\Pr(S_t = 1 | \Im_{t-1})$ represents the effect when agents anticipate as of time $t-1$ that the return of time *t* will be drawn from the high-variance distribution. According to standard mean-variance theory, foreknowledge of a high-variance should be compensated by a higher expected return. The predicted variance in this model is simply

$$\begin{aligned} E[\sigma_t^2 | \Im_{t-1}] &= [1 - \Pr(S_t = 1 | \Im_{t-1})]\sigma_0^2 + \Pr(S_t = 1 | \Im_{t-1})\sigma_1^2 + \\ &+ [1 - \Pr(S_t = 1 | \Im_{t-1})]\Pr(S_t = 1 | \Im_{t-1})(\alpha_1 - \alpha_0)^2. \end{aligned} \qquad (9.C5)$$

Thus when $\Pr(S_t = 1 | \Im_{t-1}) \in (0,1/2)$ is high, because

$$\frac{\partial E[\sigma_t^2 | \Im_{t-1}]}{\partial \Pr(S_t = 1 | \Im_{t-1})} = (\sigma_1^2 - \sigma_0^2) + [1 - 2\Pr(S_t = 1 | \Im_{t-1})](\alpha_1 - \alpha_0)^2 \qquad (9C6)$$

is positive when $\Pr(S_t = 1 | \Im_{t-1}) < 0.5$, the expected excess return should be positive so that the parameter λ is positive. On the other hand, it could be that today's high-variance state, $S_t = 1$, was not anticipated in the previous period. In this case $\Pr(S_t = 1 | \Im_{t-1})$ is small so that the average return between $t-1$ and *t* is dominated by α_1. During a period in which agents are surprised by the event

$S_t = 1$, the stock price must fall below what would have been seen had $S_t = 0$ occurred instead. This will make the return between $t-1$ and t lower and will show up as a negative value for α_1. Similar reasoning suggests that if the variance unexpectedly decreases, the return between $t-1$ and t will turn out to be higher than usual, suggesting that α_0 should be positive.

TSN also manage to establish the sign of a linear combination of the parameters. The risk premium in t is given by the expected value of the excess return conditional on the current information set. Thus, the risk premium is

$$\mu_t = [1 - \Pr(S_t = 1 | \Im_{t-1})]\alpha_0 + (\alpha_1 + \lambda)\Pr(S_t = 1 | \Im_{t-1}). \qquad (9.C7)$$

If agents are risk-averse, this equation should always be positive and increase with $\Pr(S_t = 1 | \Im_{t-1})$. The expectation will always be positive as long as $\alpha_0 \geq 0$ and $\alpha_1 + \lambda \geq 0$. Finally, if both of these conditions hold with inequality and $\alpha_1 + \lambda > \alpha_0$ then

$$\frac{\partial E[r_t | \Im_{t-1}]}{\partial \Pr(S_t = 1 | \Im_{t-1})} = \alpha_1 + \lambda - \alpha_0 > 0, \qquad (9.C8)$$

i.e., the risk premium will increase with agents' prior probability of the high-variance state.

When estimated on S&P 500 monthly data, this model yields parameter estimates that are largely consistent with asset pricing theory. The estimates ($\hat{\alpha}_0 = 0.70\%$, $\hat{\alpha}_1 = -3.36\%$ and $\hat{\lambda} = 2.88$) provide support for a risk premium rising as the anticipated level of risk rises. If the agents are certain that next period's return will be drawn from the low-variance density, agents anticipate a monthly return of 5% percent. Likewise, if agents are certain next period's return will be drawn from the high-variance density, then agents will require a monthly return of 180% annually. These estimates suggest that agents perceive stocks to be a very risky asset during high-variance periods. The unconditional probability of the high-variance state is however only 0.0352. This means that in spite of that 180% spike in expectation during high-variance regimes, the risk premium will average approximately 9% on an annual basis. This number is close to the average excess return observed in the data, 7.5%. However, one problem remains: because $\hat{\alpha}_1 + \hat{\lambda} - \hat{\alpha}_0 = -1.18 < 0$, the risk premium does not increase with anticipated variance; the variance of the linear combination is large in relation to the point estimate, the t-statistic is -0.21, so that the model provides no evidence for a risk premium changing proportionally or inversely with variance. This result is consistent with evidence as early as French, Schwert, and Stambaugh's (1987) who also find little evidence of a relation between the risk premium and volatility.

---

n-Line Ex. 9D. (**Non-Normalities under MS Mixture: Implied Conditional Moments**) Some insights may be gained from considering a simple univariate MSIH(2) model written as

$$y_{t+1} = S_{t+1}\mu_1 + (1-S_{t+1})\mu_0 + [S_{t+1}\sigma_1 + (1-S_{t+1})\sigma_0]z_{t+1} \quad z_{t+1} \sim N(0,1) \quad (9.D1)$$

in which $S_{t+1} = 0,1$ is *unobservable* at all points in time. You can easily see that in this special case, $S_{t+1}\mu_1 + (1-S_{t+1})\mu_0$ reproduces the regime-dependence in $\mu_{S_{t+1}}$; the same applies to $S_{t+1}\sigma_1 + (1-S_{t+1})\sigma_0$.

Let's start by checking moments for the benchmark, single-regime case in which $K=1$. Because these will be important below, we compute both unconditional and conditional moments. When $K=1$, it is as if $S_t = 0$ always, which means there is only one regime and μ and σ may lose the pedix that refers to the regime. Therefore, when we perform calculations for time t conditional moments and for unconditional moments, respectively, we have:

$$E_t[y_{t+1}] = E_t[\mu + \sigma z_{t+1}] = \mu + \sigma E_t[z_{t+1}] = \mu$$

$$E[y_{t+1}] = E[\mu + \sigma z_{t+1}] = \mu + \sigma E[z_{t+1}] = \mu$$

$$Var_t[y_{t+1}] = Var_t[\mu + \sigma z_{t+1}] = \sigma^2 Var_t[z_{t+1}] = \sigma^2$$

$$Var[y_{t+1}] = Var[\mu + \sigma z_{t+1}] = \sigma^2 Var[z_{t+1}] = \sigma^2$$

$$Skewness_t[y_{t+1}] = \frac{E_t[(y_{t+1}-E_t[y_{t+1}])^3]}{(Var_t[y_{t+1}])^{1.5}} = \frac{E_t[(\mu+\sigma z_{t+1}-\mu)^3]}{\sigma^3}$$
$$= \frac{\sigma^3 E_t[z_{t+1}^3]}{\sigma^3} = 0 \text{ (as } z_{t+1} \sim N(0,1))$$
$$Skewness[y_{t+1}] = \frac{E[(\mu+\sigma z_{t+1}-\mu)^3]}{\sigma^3} = \frac{\sigma^3 E[z_{t+1}^3]}{\sigma^3} = 0$$
$$Exkurt_t[y_{t+1}] = \frac{E_t[(y_{t+1}-E_t[y_{t+1}])^4]}{(Var_t[y_{t+1}])^2} - 3 = \frac{E_t[(\mu+\sigma z_{t+1}-\mu)^4]}{\sigma^4} - 3 \quad (9.D2)$$
$$= \frac{\sigma^4 E_t[z_{t+1}^4]}{\sigma^4} - 3 = 0 \text{ (as } z_{t+1} \sim N(0,1))$$
$$Exkurt[y_{t+1}] = \frac{E_t[(\mu+\sigma z_{t+1}-\mu)^4]}{\sigma^4} - 3 = \frac{\sigma^4 E[z_{t+1}^4]}{\sigma^4} - 3 = 0$$

Because $z_{t+1} \sim N(0,1)$, $y_{t+1} = \mu + \sigma z_{t+1}$, and σ is constant, we have that y_{t+1} has a normal conditional and unconditional distribution. Things are a tad more involved when $K = 2$. In this case, when you apply the conditioning, you will also need to condition with respect to the current state, S_t:

$$E[y_{t+1}|S_t] = E[S_{t+1}\mu_1 + (1-S_{t+1})\mu_0 + (S_{t+1}\sigma_1 + (1-S_{t+1})\sigma_0)z_{t+1}|S_t]$$
$$= E[S_{t+1}|S_t]\mu_1 + E[(1-S_{t+1})|S_t]\mu_0 + E[S_{t+1}|S_t]E[\sigma_1 z_{t+1}|S_t] +$$
$$+E[(1-S_{t+1})|S_t]E[\sigma_0 z_{t+1}|S_t]$$
$$= \Pr(S_{t+1}=1|S_t)\mu_1 + (1-\Pr(S_{t+1}=1|S_t))\mu_0$$
$$E[y_{t+1}] = E[S_{t+1}\mu_1 + (1-S_{t+1})\mu_0 + (S_{t+1}\sigma_1 + (1-S_{t+1})\sigma_0)z_{t+1}]$$
$$= E[S_{t+1}]\mu_1 + E[(1-S_{t+1})]\mu_0 + E[S_{t+1}]E[\sigma_1 z_{t+1}] + E[(1-S_{t+1})]E[\sigma_0 z_{t+1}]$$
$$= \bar{\xi}_1\mu_1 + (1-\bar{\xi}_1)\mu_0$$
(9.D3)

where $\bar{\xi}_1$ is the unconditional probability of regime 1, and $(1-\bar{\xi}_1)$ is the unconditional probability of regime 2.[3] Insofar as

---

[3] The reason for $E[S_{t+1}\sigma_j z_{t+1}|S_t] = E[S_{t+1}|S_t]E[\sigma_j z_{t+1}|S_t] = 0 \;\; j=0,1$ is that given S_t, S_{t+1} is independent of any other random variable indexed at time $t+1$, and in particular S_{t+1} is independent of z_{t+1} (just think of the way we have simulated returns from MS in Section 2 of Chapter 9 in the textbook). Moreover, $E[\sigma_1 z_{t+1}|S_t] = \sigma_1 E[z_{t+1}|S_t] = \sigma_1 E[z_{t+1}] = 0$ because $E[z_{t+1}] = 0$ by construction. The same applies to

$\bar{\xi}_1 \neq \Pr(S_{t+1}|S_t)$ and $(1-\bar{\xi}_1) \neq 1-\Pr(S_{t+1}|S_t)$,

clearly $E[R_{t+1}|S_t] \neq E[R_{t+1}]$ as the first moment will be a time-varying one. As for variances:

$$
\begin{aligned}
Var_t[y_{t+1}] &= \Pr(S_{t+1}=1|S_t)E[(\mu_1+\sigma_1 z_{t+1}-\Pr(S_{t+1}=1|S_t)\mu_1-(1-\Pr(S_{t+1}=1|S_t))\mu_0)^2|S_t]+\\
&\quad+\Pr(S_{t+1}=0|S_t)E[(\mu_0+\sigma_0 z_{t+1}-\Pr(S_{t+1}=1|S_t)\mu_1-(1-\Pr(S_{t+1}=1|S_t))\mu_0)^2|S_t]\\
&= \Pr(S_{t+1}=1|S_t)E[((1-\Pr(S_{t+1}=1|S_t))(\mu_1-\mu_0)+\sigma_1 z_{t+1})^2|S_t]+\\
&\quad+\Pr(S_{t+1}=0|S_t)E[(\Pr(S_{t+1}=1|S_t)(\mu_0-\mu_1)+\sigma_0 z_{t+1})^2|S_t]\\
&= \Pr(S_{t+1}{=}1|S_t)\sigma_1^2+(1-\Pr(S_{t+1}{=}1|S_t))\sigma_0^2+\Pr(S_{t+1}{=}1|S_t)(1-\Pr(S_{t+1}{=}1|S_t))(\mu_1-\mu_0)^2
\end{aligned}
\tag{9.D4}
$$

because $[(1-\Pr(S_{t+1}=1|S_t))+(\Pr(S_{t+1}=1|S_t))]=1$. Instead

$$
\begin{aligned}
Var[y_{t+1}] &= \bar{\xi}_1 E[(\mu_1+\sigma_1 z_{t+1}-\bar{\xi}_1\mu_1-(1-\bar{\xi}_1)\mu_0)^2]+\\
&\quad+(1-\bar{\xi}_1)E[(\mu_0+\sigma_0 z_{t+1}-\bar{\xi}_1\mu_1-(1-\bar{\xi}_1)\mu_0)^2]\\
&= \bar{\xi}_1 E[((1-\bar{\xi}_1)(\mu_1-\mu_0)+\sigma_1 z_{t+1})^2]-(1-\bar{\xi}_1)E[\bar{\xi}_1(\mu_1-\mu_0)+\sigma_0 z_{t+1})^2]\\
&= \bar{\xi}_1(1-\bar{\xi}_1)^2(\mu_1-\mu_0)^2+(1-\bar{\xi}_1)\bar{\xi}_1^2(\mu_1-\mu_0)^2+\bar{\xi}_1\sigma_1^2+(1-\bar{\xi}_1)\sigma_0^2\\
&= \bar{\xi}_1\sigma_1^2+(1-\bar{\xi}_1)\sigma_0^2+\bar{\xi}_1(1-\bar{\xi}_1)(\mu_1-\mu_0)^2.
\end{aligned}
\tag{9.D5}
$$

In both cases, notice that

$$
\begin{aligned}
Var_t[y_{t+1}] &\neq \Pr(S_{t+1}=1|S_t)\sigma_1^2+(1-\Pr(S_{t+1}=1|S_t))\sigma_0^2\\
Var_t[y_{t+1}] &\neq \bar{\xi}_1\sigma_1^2+(1-\bar{\xi}_1)\sigma_0^2,
\end{aligned}
\tag{9.D6}
$$

with the difference represented by the terms $\Pr(S_{t+1}=1|S_t)(1-\Pr(S_{t+1}=1|S_t))(\mu_1-\mu_0)^2$ in the case of the conditional variance and $\bar{\xi}_1(1-\bar{\xi}_1)(\mu_1-\mu_0)^2$ in the case of the unconditional variance. This means that in a MSIH(2) model, not only the regime-specific variances will be weighted in the overall variance across regimes, but also the (squared) size of the between-regime jumps in regime-specific means, $\mu_1-\mu_0$, will contribute to the variability of the process.

We now move to compute conditional and unconditional skewness:

$E[S_{t+1}\sigma_j z_{t+1}]=E[S_{t+1}]E[\sigma_j z_{t+1}]=0 \;\; j=0,1$.

$$E_t[(y_{t+1}-E_t[y_{t+1}])^3]=\Pr(S_{t+1}=1|S_t)E[(\mu_1+\sigma_1 z_{t+1}-\Pr(S_{t+1}=1|S_t)\mu_1+$$
$$-(1-\Pr(S_{t+1}=1|S_t))\mu_0)^3|S_t]+\Pr(S_{t+1}=0|S_t)E[(\mu_0+\sigma_0 z_{t+1}+$$
$$-\Pr(S_{t+1}=1|S_t)\mu_1-(1-\Pr(S_{t+1}=1|S_t))\mu_0)^3|S_t]$$
$$=\Pr(S_{t+1}=1|S_t)(1-\Pr(S_{t+1}=1|S_t))^3(\mu_1-\mu_0)^3+(1-\Pr(S_{t+1}=1|S_t))\times$$
$$\times(\Pr(S_{t+1}=1|S_t))^3(\mu_0-\mu_1)^3+\sigma_1^3 E[z_{t+1}^3|S_t]+\sigma_0^3 E[z_{t+1}^3|S_t]+$$
$$+3\Pr(S_{t+1}=1|S_t)(1-\Pr(S_{t+1}=1|S_t))(\mu_1-\mu_0)\sigma_1^2 E[z_{t+1}^2|S_t]+$$
$$-3\Pr(S_{t+1}=1|S_t)(1-\Pr(S_{t+1}=1|S_t))(\mu_1-\mu_0)\sigma_0^2 E[z_{t+1}^2|S_t]$$
$$=\Pr(S_{t+1}=1|S_t)(1-\Pr(S_{t+1}=1|S_t))(\mu_1-\mu_0)^3[(1-\Pr(S_{t+1}=1|S_t))^2+$$
$$-(\Pr(S_{t+1}=1|S_t))^2]+3\Pr(S_{t+1}=1|S_t)(1-\Pr(S_{t+1}=1|S_t))(\mu_1-\mu_0)(\sigma_1^2-\sigma_0^2) \quad (9.D7)$$

where $E[z_{t+1}^3|S_t]=0$, $E[z_{t+1}^2|S_t]=1$, $E[(1-\Pr(S_{t+1}=1|S_t))^2$ $(\mu_1-\mu_0)^2\sigma_1 z_{t+1}|S_t]=E[(\Pr(S_{t+1}=1|S_t))^2\,(\mu_0-\mu_1)^2\,\sigma_0 z_{t+1}|S_t]=0$ so that

$$Skewness_t[y_{t+1}]=(\mu_1-\mu_0)\times$$
$$\times\frac{\xi_{1,t+1}(1-\xi_{1,t+1})\left\{(\mu_1-\mu_0)^2[(1-\xi_{1,t+1})^2-\xi_{1,t+1}^2]+3(\sigma_1^2-\sigma_0^2)\right\}}{[\xi_{1,t+1}\sigma_1^2+(1-\xi_{1,t+1})\sigma_0^2+\xi_{1,t+1}(1-\xi_{1,t+1})(\mu_1-\mu_0)^2]^{3/2}} \quad (9.D8)$$

where we have shortened the notation by defining $\xi_{1,t+1}\equiv\Pr(S_{t+1}=1|S_t)$. Similarly, straightforward but tedious algebra reveals that

$$Skewness[R_{t+1}]=(\mu_1-\mu_0)\times$$
$$\times\frac{\bar{\xi}_1(1-\bar{\xi}_1)[(\mu_1-\mu_0)^2[(1-\bar{\xi}_1)^2-\bar{\xi}_1^2]+3(\sigma_1^2-\sigma_0^2)]}{[\bar{\xi}_1\sigma_1^2+(1-\bar{\xi}_1)\sigma_0^2+\bar{\xi}_1(1-\bar{\xi}_1)(\mu_1-\mu_0)^2]^{3/2}}. \quad (9.D9)$$

This finding is very interesting:

$$Skewness_t[y_{t+1}]\neq 0 \text{ if and only if } \mu_1\neq\mu_0$$
$$Skewness[y_{t+1}]\neq 0 \text{ if and only if } \mu_1\neq\mu_0, \quad (9.D10)$$

i.e., you need switching in conditional means in order for non-zero skewness to obtain. However, it is also clear that even when $\mu_1\neq\mu_0$ it is possible for both conditional and unconditional skewness coefficient to be zero when (this is just a sufficient condition): (i) $\xi_{1,t+1}=0$ or $\bar{\xi}_1=0$; (ii) $\xi_{1,t+1}=1$ or $\bar{\xi}_1=1$. The two sets of restrictions do not carry the same meaning though, as $\bar{\xi}_1=0$ or 1 really means we are not facing a MS model, in the sense that the underlying MC may be clearly reduced to a single state, while $\xi_{1,t+1}=0$ or 1 just means that as of time *t* you are certain that in the following period you are either in the first regime or in the second.[4]

Finally, we deal with conditional and unconditional excess kurtosis:

$$
\begin{aligned}
E_t[(R_{t+1}-E_t[R_{t+1}])^4] &= \Pr(S_{t+1}{=}1|S_t)E[(\mu_1+\sigma_1 z_{t+1}-\Pr(S_{t+1}{=}1|S_t)\mu_1+\\
&\quad -(1-\Pr(S_{t+1}{=}1|S_t))\mu_0)^4|S_t]+\Pr(S_{t+1}{=}0|S_t)E[(\mu_0+\sigma_0 z_{t+1}-\Pr(S_{t+1}{=}1|S_t)\mu_1+\\
&\quad -(1-\Pr(S_{t+1}{=}1|S_t))\mu_0)^4|S_t]\\
&= \Pr(S_{t+1}{=}1|S_t)(1-\Pr(S_{t+1}{=}1|S_t))(\mu_1-\mu_0)^4[(1-\Pr(S_{t+1}{=}1|S_t))^3+(\Pr(S_{t+1}{=}1|S_t))^3]+\\
&\quad +6\Pr(S_{t+1}{=}1|S_t)(1-\Pr(S_{t+1}{=}1|S_t))(\mu_1-\mu_0)^2[(1-\Pr(S_{t+1}{=}1|S_t))\sigma_1^2+\\
&\quad +\Pr(S_{t+1}{=}1|S_t)\sigma_0^2]+3\Pr(S_{t+1}{=}1|S_t)\sigma_1^4+3(1-\Pr(S_{t+1}{=}1|S_t))\sigma_0^4
\end{aligned}
\tag{9.D11}
$$

where $E[z_{t+1}|S_t]=E[z_{t+1}^3|S_t]=0$, $E[z_{t+1}^2|S_t]=1$, $E[z_{t+1}^4|S_t]=3$, so that

$$
\begin{aligned}
ExKurt_t[R_{t+1}] &= \frac{\xi_{1,t+1}(1-\xi_{1,t+1})\{(\mu_1-\mu_0)^4[(1-\xi_{1,t+1})^3+\xi_{1,t+1}^3]+6(\mu_1-\mu_0)^2}{[\xi_{1,t+1}\sigma_1^2+(1-\xi_{1,t+1})\sigma_0^2+\xi_{1,t+1}(1-\xi_{1,t+1})(\mu_1-\mu_0)^2]^2}\\
&\quad +\frac{[(1-\xi_{1,t+1})\sigma_1^2+\xi_{1,t+1}\sigma_0^2]\}+3\xi_{1,t+1}\sigma_1^4+3(1-\xi_{1,t+1})\sigma_0^4}{[\xi_{1,t+1}\sigma_1^2+(1-\xi_{1,t+1})\sigma_0^2+\xi_{1,t+1}(1-\xi_{1,t+1})(\mu_1-\mu_0)^2]^2}-3.
\end{aligned}
\tag{9.D12}
$$

Similarly, straightforward but tedious algebra reveals that

---

[4] If that seems more plausible, consider that $\Pr(S_{t+1}=1|S_t)=(1-p_{00})\times \Pr(S_t=0|\Im_t)+p_{11}\Pr(S_t=1|\Im_t)$ which can be 1 if and only if either $(1-p_{00})\Pr(S_t=0|\Im_t)=1$ (but that means that $p_{00}=0$), $p_{11}\Pr(S_t=1|\Im_t)=1$ (but that means that $p_{11}=1$), or the sum happens to be one. The first two cases do indicate problems with the irreducibility of the MC. The third case is more interesting. If

$$1=(1-p_{00})\Pr(S_t=0|\Im_t)+p_{11}\Pr(S_t=1|\Im_t)=\boldsymbol{e}'_2\boldsymbol{P}'\boldsymbol{\xi}_{t|t}=\boldsymbol{e}'_2E_t[\boldsymbol{\xi}_{t+1}]$$

where $\boldsymbol{e}'_2=[0\ 1]$, this means that as of time *t* we are forecasting with certainty that time *t*+1 will be dominated by regime 1. That is rather odd as it implies a very precise periodicity of the underlying MC.

$$ExKurt[R_{t+1}] = \frac{\bar{\xi}_1(1-\bar{\xi}_1)\Big\{(\mu_1-\mu_0)^4[(1\text{-}\bar{\xi}_1)^3+\bar{\xi}_1^3]+}{[\bar{\xi}_1\sigma_1^2+(1-\bar{\xi}_1)\sigma_0^2+\bar{\xi}_1(1-\bar{\xi}_1)(\mu_1-\mu_0)^2]^2}$$
$$+\frac{+6(\mu_1-\mu_0)^2[(1-\bar{\xi}_1)\sigma_1^2+\bar{\xi}_1\sigma_0^2]\Big\}3\bar{\xi}_1\sigma_1^4+3(1-\bar{\xi}_1)\sigma_0^4}{[\bar{\xi}_1\sigma_1^2+(1-\bar{\xi}_1)\sigma_0^2+\bar{\xi}_1(1-\bar{\xi}_1)(\mu_1-\mu_0)^2]^2}-3. \quad (9.D13)$$

This finding is once more very interesting. First of all, notice that also in this case, when $\mu_0 = \mu_1$,

$$Ex\ Kurt_t[R_{t+1}] = \frac{3\xi_{1,t+1}\sigma_1^4+3(1-\xi_{1,t+1})\sigma_0^4}{[\xi_{1,t+1}\sigma_1^2+(1\text{-}\xi_{1,t+1})\sigma_0^2]^2}-3$$
$$=3\left[\frac{\xi_{1,t+1}\sigma_1^4+(1-\xi_{1,t+1})\sigma_0^4}{\xi_{1,t+1}^2\sigma_1^4+(1\text{-}\xi_{1,t+1})^2\sigma_0^4+2\xi_{1,t+1}(1\text{-}\xi_{1,t+1})\sigma_0^2\sigma_1^2}\text{-}1\right] \quad (9.D14)$$

which is less than the expression found above: MS dynamics means simply adds to the excess kurtosis of a series. Moreover, in this case MS will generate positive excess kurtosis if and only if

$$\xi_{1,t+1}\sigma_1^4+(1-\xi_{1,t+1})\sigma_0^4 > \xi_{1,t+1}^2\sigma_1^4+(1-\xi_{1,t+1})^2\sigma_0^4+$$
$$+2\xi_{1,t+1}(1-\xi_{1,t+1})\sigma_0^2\sigma_1^2 \quad (9.D15)$$

Moreover, notice that if one also has $\sigma_0^2=\sigma_1^2=\sigma^2$, then

$$ExKurt_t[R_{t+1}] = \frac{\sigma^4[3\xi_{1,t+1}+3(1-\xi_{1,t+1})]}{\sigma^4[\xi_{1,t+1}+(1-\xi_{1,t+1})]^2}-3=0 \quad (9.D16)$$

as it should be because when $\mu_0=\mu_1$ and $\sigma_0^2=\sigma_1^2$, there is no MS left in the process.

Because in the single-regime case, the normality of the shocks z_{t+1} carries over to the series investigated, it is sensible to ask what are the conditional and unconditional distributions of returns under the two-state MS process. Here the point is that even a simple two-state MSIH model such as the one in this section, may generate substantial departures from normality. Given a MS model, it is clear that conditioning on S_{t+1} —which is equivalent to say that either the regime is observable (but this violates our assumptions) or that, again oddly, S_{t+1} may be perfectly predicted— $y_{t+1} \sim N(\mu_{S_{t+1}}, \sigma^2_{S_{t+1}})$, which is a simple Gaussian distribution. However, in a MS model, S_{t+1} is unobservable, while the case in which S_{t+1} may be perfectly predicted given time t information appears to be bizarre. In fact, even if you were to somehow know what the current, time t regime S_t is, notice that in general $\Pr(S_{t+1}=j|S_t=i)$ represents the generic $[i,j]$ element of the transition matrix **P**. If the Markov chain is ergodic and irreducible, we know that $\Pr(S_{t+1}=j|S_t=i)<1$, $i,j=1,2$. Because of this fact the conditional distribution of y_{t+1} is:

$$f(y_{t+1}|\Im_t)=f(y_{t+1}|S_t)=\Pr(S_{t+1}=1|S_t)\phi(\mu_1,\sigma_1^2)+ \\ +(1-\Pr(S_{t+1}=1|S_t))\phi(\mu_0,\sigma_0^2), \tag{9.D17}$$

where $\varphi(\mu_1,\sigma_1^2)$ is a normal density function with mean μ_1 and variance σ_1^2. As we know from Definition 9.2, the density in (9.D17) is a mixture, with probabilistic and time-varying weights $\Pr(S_{t+1}=1|S_t)$, $(1-\Pr(S_{t+1}=1|S_t)$, of two normal densities and it is *not* itself a normal density. Therefore, even conditioning on time t information and on knowledge (still difficult to obtain) of the current state S_t, returns in a two-state MS will not have a normal distribution, unless $\mu_0=\mu_1$ and $\sigma_0^2=\sigma_1^2$, when (trivially)

$$f(y_{t+1}|\Im_t)=\Pr(S_{t+1}=1|S_t)\varphi(\mu,\sigma^2)+(1-\Pr(S_{t+1}=1|S_t))\phi(\mu,\sigma^2) \\ =[\Pr(S_{t+1}=1|S_t)+(1-\Pr(S_{t+1}=1|S_t))]\phi(\mu,\sigma^2)=\phi(\mu,\sigma^2). \tag{9.D18}$$

In fact, we note that when $\mu_0=\mu_1$ and $\sigma_0^2=\sigma_1^2$, from results obtained above we have

$$Skew_t[y_{t+1}]=(\mu-\mu)\frac{\xi_{1,t+1}(1\text{-}\xi_{1,t+1})\left\{(\mu-\mu)^2[(1\text{-}\xi_{1,t+1})^2+\xi_{1,t+1}^2]+3(\sigma^2-\sigma^2)\right\}}{[\xi_{1,t+1}\sigma^2+(1-\xi_{1,t+1})\sigma^2+\xi_{1,t+1}(1-\xi_{1,t+1})(\mu-\mu)^2]^{3/2}}=0 \tag{9.D19}$$

$$ExKurt_t[R_{t+1}]=\frac{\xi_{1,t+1}(1\text{-}\xi_{1,t+1})\left\{(\mu\text{-}\mu)^4[(1\text{-}\xi_{1,t+1})^3+\xi_{1,t+1}^3]+6(\mu\text{-}\mu)^2[(1\text{-}\xi_{1,t+1})\sigma^2+\xi_{1,t+1}\sigma^2]\right\}}{[\xi_{1,t+1}\sigma^2+(1-\xi_{1,t+1})\sigma^2+\xi_{1,t+1}(1-\xi_{1,t+1})(\mu-\mu)^2]^2} \\ +\frac{3\xi_{1,t+1}\sigma^4+3(1\text{-}\xi_{1,t+1})\sigma^4}{[\xi_{1,t+1}\sigma^2+(1-\xi_{1,t+1})\sigma^2+\xi_{1,t+1}(1-\xi_{1,t+1})(\mu-\mu)^2]^2}-3=0, \tag{9.D20}$$

which is consistent with the conclusion that y_{t+1} follows a normal distribution.

As for the unconditional density of y_{t+1}, i.e., the density that does not condition on any precise prior information, it is logical to state that absent any information on either S_t or at least $\Pr(S_t \mid \Im_t)$, the best assessment we can make of each of the regimes is simply that $\Pr(S_t = 1) = \bar{\xi}_1$ and $\Pr(S_t = 0) = 1 - \bar{\xi}_1$. Therefore, on average, in the population, the data will come $\bar{\xi}_1$ percent of the time from $\varphi(\mu_1, \sigma_1^2)$ and $(1 - \bar{\xi}_1)$ percent of the time from $\varphi(\mu_0, \sigma_0^2)$. The result is that the unconditional distribution of y_{t+1} is:

$$f(y_{t+1}) = \bar{\xi}_1 \phi(\mu_1, \sigma_1^2) + (1 - \bar{\xi}_1)\phi(\mu_0, \sigma_0^2), \qquad (9.D21)$$

which is another mixture (in this case, not time-varying, being unconditional) of two normal distributions and that, as we know, this will imply (assuming $\bar{\xi}_1 \in (0,1)$)

$$Skew[y_{t+1}] = (\mu_1 - \mu_0)\frac{\bar{\xi}_1(1\text{-}\bar{\xi}_1)[(\mu_1 - \mu_0)^2[(1 - \bar{\xi}_1)^2 + \bar{\xi}_1^2] + 3(\sigma_1^2 - \sigma_0^2)]}{[\bar{\xi}_1\sigma_1^2 + (1 - \bar{\xi}_1)\sigma_0^2 + \bar{\xi}_1(1 - \bar{\xi}_1)(\mu_1 - \mu_0)^2]^{3/2}} \neq 0$$

$$ExKurt[y_{t+1}] = \frac{\bar{\xi}_1(1\text{-}\bar{\xi}_1)\Big\{(\mu_1 - \mu_0)^4[(1\text{-}\bar{\xi}_1)^3 + \bar{\xi}_1^3] + 6(\mu_1 - \mu_0)^2}{[\bar{\xi}_1\sigma_1^2 + (1 - \bar{\xi}_1)\sigma_0^2 + \bar{\xi}_1(1 - \bar{\xi}_1)(\mu_1 - \mu_0)^2]^2} + \qquad (9.D22)$$

$$+ \frac{\times[(1\text{-}\pi_1)\sigma_1^2 + \pi_1\sigma_0^2]\Big\} 3\pi_1\sigma_1^4 + 3(1 - \pi_1)\sigma_0^4}{[\bar{\xi}_1\sigma_1^2 + (1 - \bar{\xi}_1)\sigma_0^2 + \bar{\xi}_1(1 - \bar{\xi}_1)(\mu_1 - \mu_0)^2]^2} - 3 > 0$$

Additionally, when $\mu_0 \neq \mu_1$, notice that even the variance of $f(y_{t+1})$ fails to simply be the probability-weighted average of σ_1^2 and σ_0^2 because, as we know, $Var[y_{t+1}] = \bar{\xi}_1\sigma_1^2 + (1 - \bar{\xi}_1)\sigma_0^2 + \bar{\xi}_1(1 - \bar{\xi}_1)(\mu_1 - \mu_0)^2$.
The on-line Example 9E puts this ideas to work in an application to Value-at-Risk calculations, showing how dealing with moments and densities derived from MS models requires a degree of familiarity with Monte Carlo simulation techniques.
For instance, Rydèn, Teräsvirta, and Åsbrink (1998) show that MS mixtures estimated on daily S&P 500 returns for a long1928–1991 sample closely reproduces a range of properties of asset returns previously emphasized in the literature (see Chapter 5 and Granger and Ding, 1995): returns are not autocorrelated in levels (except, possibly, at lag one); the autocorrelation functions of absolute and squared returns decay slowly starting from the first autocorrelation and $Corr(|R_t|,|R_{t-}|) > Corr(|R_t|^{\theta},|R_{t-}|^{\theta}) > 0$ for $\theta \neq 1$; the decay in the autocorrelation functions of squares and absolute values of returns is much slower than the exponential rate of a stationary

AR(l) or ARMA(*p,q*) model; the autocorrelations of $sign(R_t)$ are insignificant; moreover, $|R_t|$ and $sign(R_t)$ are independent.

On-Line Ex. 9E. (**Value-at-Risk in MS Models**) As a case of the effectiveness and benefits of MS modelling in risk management applications, consider again the simple univariate MSIH(2) model,

$$R_{t+1} = S_{t+1}\mu_1 + (1-S_{t+1})\mu_0 + [S_{t+1}\sigma_1 + (1-S_{t+1})\sigma_0]z_{t+1} \quad z_{t+1} \sim N(0,1),$$

in which $S_{t+1} = 0$, 1 is *unobservable* at all points in time. In the limit case in which *K* = 1, which is a benchmark single-state linear model, to compute (say) 1% VaR is straightforward:

$$\begin{aligned} 0.01 &= \Pr(R_{t+1} < -VaR_{t+1}^{0.01}(k=1)) = \Pr\left(\frac{R_{t+1}-\mu}{\sigma} < -\frac{VaR_{t+1}^{0.01}(k=1)+\mu}{\sigma}\right) \\ &= \Pr\left(z_{t+1} < -\frac{VaR_{t+1}^{0.01}(k=1)+\mu}{\sigma}\right) = \Phi\left(-\frac{VaR_{t+1}^{0.01}(k=1)+\mu}{\sigma}\right) \end{aligned} \qquad (9.E1)$$

so that, after defining $\Phi^{-1}(\cdot)$ as the inverse CDF of a standard normal distribution, we have:

$$\begin{aligned} \Phi^{-1}(0.01) &= \Phi^{-1}\left(\Phi\left(-\frac{VaR_{t+1}^{0.01}(k=1)+\mu}{\sigma}\right)\right) = -\frac{VaR_{t+1}^{0.01}(k=1)+\mu}{\sigma} \\ &\Rightarrow VaR_{t+1}^{0.01}(K=1) = -\sigma\Phi^{-1}(0.01) - \mu. \end{aligned} \qquad (9.E2)$$

Now, moving to the *K* = 2 case, let's start from an approximate way to look at the problem of computing 1% VaR: one colleague in your risk management department is proposing to use the following *conditional* 1% VaR measure:

$$\begin{aligned} VaR_{t+1}^{0.01}(K=2) = &-\left[\Pr(S_{t+1}=1|S_t)\sigma_1^2 + (1-\Pr(S_{t+1}=1|S_t))\sigma_0^2\right]\Phi^{-1}(0.01) + \\ &-\left[\Pr(S_{t+1}=1|S_t)\mu_1 + (1-\Pr(S_{t+1}=1|S_t))\mu_0\right], \end{aligned} \qquad (9.E3)$$

in which the colleague is obviously conditioning with respect to the current state, S_t but still applying a normal distribution result. Unfortunately, you should not agree with his/her proposal, or at least should clarify to the team that this is simply an approximation. The reason is that in Section 8.3 of the book we have found that

$$\begin{aligned} f(R_{t+1}|\Im_t) &= f(R_{t+1}|S_t) = \Pr(S_{t+1}{=}1|S_t)\phi(\mu_1,\sigma_1^2) + (1-\Pr(S_{t+1}{=}1|S_t))\phi(\mu_0,\sigma_0^2) \\ &\neq \phi\left(\left[\Pr(S_{t+1}{=}1|S_t)\mu_1 + (1-\Pr(S_{t+1}{=}1|S_t))\mu_0\right], \Pr(S_{t+1}{=}1|S_t)\sigma_1^2 + (1-\Pr(S_{t+1}{=}1|S_t))\sigma_0^2\right), \end{aligned}$$

(9.E4)

and that $f(R_{t+1}|\Im_t)$ does not follow a Normal distribution, but a probability-weighted mixture of two normal distributions which is itself not a Normal distribution. As a result, the way of proceeding to VaR calculations proposed by the colleague may turn out to be grossly incorrect as it employs $\Phi^{-1}(0.01)$, where the use of the standard normal CDF was previously coming from the fact that $R_{t+1} \sim N(\mu,\sigma^2)$. When this assumption breaks down, the procedure is clearly invalid. Moreover, you know from Section 3 of Chapter 8 that

$$Var_t[R_{t+1}] = \Pr(S_{t+1}=1|S_t)\sigma_1^2 + (1\text{-}\Pr(S_{t+1}=1|S_t))\sigma_0^2 + \Pr(S_{t+1}=1|S_t)(1\text{-}\Pr(S_{t+1}=1|S_t))(\mu_1\text{-}\mu_0)^2$$
$$\neq \Pr(S_{t+1}=1|S_t)\sigma_1^2 + (1-\Pr(S_{t+1}=1|S_t))\sigma_0^2 \qquad (9.E5)$$

unless $\mu_0 = \mu_1$, which is generally not the case in a MSIH(2,0) model. After you have made your objection during his presentation, this colleague of yours revises his/her proposal to use the following *conditional* 1% VaR measure:

$$VaR_{t+1}^{0.01}(K=2) = -[\Pr(S_{t+1}=1|S_t)\sigma_1^2 + (1\text{-}\Pr(S_{t+1}=1|S_t))\sigma_0^2 + \Pr(S_{t+1}=1|S_t)(1\text{-}\Pr(S_{t+1}=1|S_t))\times$$
$$\times(\mu_1-\mu_0)^2]\Phi^{-1}(0.01) - \left[\Pr(S_{t+1}=1|S_t)\mu_1 + (1-\Pr(S_{t+1}=1|S_t))\mu_0\right]. \qquad (9.E6)$$

Your reaction should remain negative: unfortunately, making one claim less wrong does not make it correct. Even though it is now correct that

$$Var_t[R_{t+1}] = \Pr(S_{t+1}=1|S_t)\sigma_1^2 + (1-\Pr(S_{t+1}=1|S_t))\sigma_0^2 +$$
$$+\Pr(S_{t+1}=1|S_t)(1-\Pr(S_{t+1}=1|S_t))(\mu_1-\mu_0)^2, \qquad (9.E7)$$

the fact remains that

$$f(R_{t+1}|\Im_t) = f(R_{t+1}|S_t) = \Pr(S_{t+1}=1|S_t)\phi(\mu_1,\sigma_1^2) + (1-\Pr(S_{t+1}=1|S_t))\phi(\mu_0,\sigma_0^2)$$
$$\neq \phi(\left[\Pr(S_{t+1}=1|S_t)\mu_1 + (1\text{-}\Pr(S_{t+1}=1|S_t))\mu_0\right],$$
$$[\Pr(S_{t+1}=1|S_t)\sigma_1^2 + (1\text{-}\Pr(S_{t+1}=1|S_t))\sigma_0^2] + \Pr(S_{t+1}=1|S_t)(1\text{-}\Pr(S_{t+1}=1|S_t))(\mu_1-\mu_0)^2) \qquad (9.E8)$$

so that VaR cannot be computed in that simply way.

Finally, it seems time for you to suggest how this should be done correctly. Here you may be in trouble, though: unfortunately there is no closed-form solution which means that you will have to resort to simulation-based (Monte Carlo) methods. The problem is that

$$f(R_{t+1}|S_t) = \Pr(S_{t+1}=1|S_t)\phi(\mu_1,\sigma_1^2) + (1-\Pr(S_{t+1}=1|S_t))\phi(\mu_0,\sigma_0^2) \qquad (9.E9)$$

fails to have a closed-form representation and as such it impossible to simply draw from some well-specified PDF or CDF. This means that your proof of the functional form of 1% VaR in

$$0.01 = \Pr(R_{t+1} < -VaR_{t+1}^{0.01}(K=1)) = \Pr\left(\frac{R_{t+1}-\mu}{\sigma} < -\frac{VaR_{t+1}^{0.01}(K=1)+\mu}{\sigma}\right)$$
$$= \Pr\left(z_{t+1} < -\frac{VaR_{t+1}^{0.01}(K=1)+\mu}{\sigma}\right) = \Phi\left(-\frac{VaR_{t+1}^{0.01}(K=1)+\mu}{\sigma}\right) \quad (9.E10)$$

simply fails because it is not true that $\Pr(z_{t+1} < (-VaR_{t+1}^{0.01}(k=1)+\mu)/\sigma)$ can be measured using $\Phi(\cdot)$. What you can do is the following. First, simulate a large number *M* of one-month returns assuming $S_t = 0$ from

$$R_{t+1} = S_{t+1}\mu_1 + (1-S_{t+1})\mu_0 + [S_{t+1}\sigma_1 + (1-S_{t+1})\sigma_0]z_{t+1} \quad z_{t+1} \sim N(0,1), \quad (9.E11)$$

when $S_{t+1} = 1$ with probability $p_{01} = (1-p_{00})$ and $S_{t+1} = 0$ with probability p_{00}. Call these *M* one-month ahead returns $\left\{R_{t+1}^m(S_t=0)\right\}_{m=1}^{M}$.[5] Second, simulate a large number *M* of one-month returns assuming $S_t = 1$ from

$$R_{t+1} = S_{t+1}\mu_1 + (1-S_{t+1})\mu_0 + [S_{t+1}\sigma_1 + (1-S_{t+1})\sigma_0]z_{t+1} \quad z_{t+1} \sim N(0,1), \quad (9.E12)$$

When $S_{t+1} = 1$ with probability p_{11} and $S_{t+1} = 0$ with probability $1-p_{11}$. Call these *M* one-month ahead returns $\left\{R_{t+1}^m(S_t=1)\right\}_{m=1}^{M}$. Finally, you need to aggregate this 2*M* simulations in a unique set, using:

$$R_{t+1}^m = \Pr(S_t=1|\Im_t)R_{t+1}^m(S_t=1) + (1-\Pr(S_t=1|\Im_t))R_{t+1}^m(S_t=0) \quad m=1,2,\ldots,M. \quad (9.E13)$$

[5]This means that when $S_{t+1} = 1$ you will simulate from $R_{t+1} = \mu_1 + \sigma_1 z_{t+1}$; when $S_{t+1} = 0$ you will simulate from $R_{t+1} = \mu_0 + \sigma_0 z_{t+1}$. How do you simulate a two-point (also called Bernoulli) random variable that takes value 1 with probability $1-p_{00}$ and 0 with probability p_{00}? Simple, you draw a uniform defined on [0,1] and you set $S_{t+1} = 1$ if the uniform draw is less than (or equal to) $1-p_{00}$, and you set $S_{t+1} = 0$ otherwise.

At this point, your 1% VaR will be simply defined as: the simulated returns in the set $\left\{R_{t+1}^{m}\right\}_{m=1}^{M}$ that leaves exactly 1% of your total M simulations (after your aggregation step, i.e., $M/100$ simulations, which better be an integer) *below* the 1% VaR value.