EXERCISE ON SVAR'S, HIDDEN CHAINS

(1) Consider a model of Markov-switching mean and variance, as follows:

$$y_t = \mu(S_t) + \sigma(S_t)\varepsilon_t\,,\; t = 1, \dots, T \tag{1}$$

$$\varepsilon_t \sim N(0,1),\;\text{ i.i.d. across } t \tag{2}$$

$$S_t \in \{0,1\}\,,\;\text{all t} \tag{3}$$

$$P[S_{t+1} = j \mid S_t = i] = p_{ij} \tag{4}$$

$$P[S_1 = j] = \pi_j\,. \tag{5}$$

(a) Write down the pdf for the data $\{y_1, \dots, y_T\}$ conditional on $\{\mu_j, \sigma_j, \pi_j,\; j = 1, 2\}$, the matrix of p_{ij} values, and the sequence $\{S_t,\; t = 1, \dots, T\}$ values.
Conditional on all this, the data are jointly normal, so the pdf is

$$\prod_{t=1}^{T} \frac{\exp\left(-\frac{(y_t-\mu(S_t))^2}{2\sigma^2(S_t)}\right)}{\sqrt{2\pi\sigma^2(S_t)}}\,. \tag{$*$}$$

(b) Verify that this pdf is unbounded above for any fixed $\{y_t\}$ sequence. [Hint: consider S_t sequences where one of the two possible values for S occurs at only a single value of t.]
If $S_t = 0$ only at $t = t_0$, then if we set $\mu(0) = y_{t_0}$ and let $\sigma(0)$ approach zero, the likelihood component for t_0 goes to infinity, while the other components of the product in $(*)$ are unaffected.

(c) If we did not condition on $\{S_t\}$, but calculated the likelihood as a function of the μ_j, σ_j, p_{ij}, and π_j values alone, would we still have unboundedness?
To eliminate the conditioning on the S_t sequence, we have to sum the expression in $(*)$ over all S sequences. But since there is a finite sample and only two possible values of the state, there is a finite number (i.e. 2^T) of such sequences, all of which have non-negative probability. If any one of them is unbounded abaove, therefore, the entire sum is unbounded above.

(d) Does the unboundedness of the likelihood make it fail to be integrable?
Yes. Consider the part of the likelihood coming from terms in which $S_t = j$. Considered as a function of μ_j (this is alternate notation for $\mu(S_j)$), this has the form of a $N(\bar{y}_j, \sigma^2/T_j)$ pdf, where T_j is the number of observations for which $S_j = T_j$ and $\bar{y}_j$ is the sample mean of y_t over this part of the sample. When we integrate with respect to μ_j, we get

$$\sqrt{T_j}(2\pi\sigma_j^2)^{-(T_j-1)/2}\exp\left(-\frac{s_j^2}{2\sigma_j^2}\right)\,,$$

Date: January 19, 2006.

where s_j^2 is the sum of squared deviations of y_t from μ_j over the $(S_t = j)$ part of the sample. Considered as a function of $1/\sigma^2$, this is in the form of a $\chi^2(T_j + 1, s^2)$ pdf, and hence is clearly integrable for all positive integer T^j, so long as $s_j^2 > 0$. When $T_j > 1$, it is a zero probability event that $s_j^2 = 0$. But for $T_j = 1$, $s_j^2 \equiv 0$, and the whole expression is a constant, and thus not integrable with respect to σ_j^2. And of course when $T_j = 0$, the likelihood is constant as a function of μ_j and σ_j^2, so that when we integrate with respect to them we get infinity.

(e) The usual strategy, as we discussed in class, for making MCMC draws from a model like this is to draw μ_j, σ_j values and p_{ij} values from their posteriors conditional on the $\{S_t\}$ sequence, then draw $\{S_t\}$ conditional on the other parameters, and iterate. Does the unboundedness of the likelihood create problems for this approach, assuming a flat prior is used?
Yes. MCMC iterations do not generally converge at all when the target kernel is not integrable. In this case they would most likely get stuck at points near one of the poles in the likelihood.

(f) What kind of settings of the parameters of a conjugate prior would eliminate the unboundedness of the posterior?
One could eliminate the non-integrability by using a conjugate prior equivalent to the likelihood function with two observation on each state. Then the full posterior would have the shape of a likelihood with $T_j + 2$ observations on each state, so that even with $T_j = 0$, there is a well-defined, integrable posterior.

(2) Here is an identification scheme for a structural VAR in interest rates (r), money stock (m), the price level (p), and industrial production ip:

	r	m	p	ip
m policy	x	x	0	0
m demand	x	x	x	x
py 1	0	0	x	0
py 2	0	0	x	x

This matrix describes the constraints on A_0 in the system $A(L)y_t = \varepsilon_t$, with ε_t i.i.d. $N(0, I)$. The zeros in the matrix represent coefficients constrained to zero. The x's represent coefficients that are not constrainted.

(a) Is this system identified? Overidentified? How do you know?
It is identified. First note that no linear combination of equations in an orthonormal transformation of the system can mix either of the first two equations with either of the last two. Such a combination would have to violate the zero restrictions on the last two equations. (Note that an orthonormal transformation, if it replaces one linear equation with a linear combination of itself and another equation, must replace the other equation also, with an orthognal linear combination.) The same triangularity structure implies that the two last equations cannot be transformed without violating the last zero restriction in the third row. The first two equations can't be altered because any transformation of those two would violate the zero restrictions on the first equation.

One can also point to the order conditions. Since $(n^2 - n)/2 = 6$ here, we have one more restriction than is needed for local identification, so we expect overidentification. The inverse of the reduced form covariance matrix Σ is $A_0' A_0$, and one can show that the submatrix of that formed by the 3rd and 4th rows, first and second columns, is necessarily singular, so Σ is indeed restricted.

(b) Suppose two of the coefficients corresponding to x's turned out to be zero in fact, even though they are not constrained to be zero. Would this raise problems for identification? Would your answer differ according to which two x's turned out to be zeros?
Any pair of additional zeros that made A_0 singular of course implies that the system is ***incomplete*** *(i.e. that it can't be solved for y_t as a function of past y's and current shocks). Examples would be single zeros at (3,3) or 4,4), or pairs of zeros in the same row and column in the upper left 2x2 matrix. But this is not an identification problem. There would be an identification problem if the (2,3) and (2,4) coefficients were both zero. In that case, though the system is generically complete, the first and second equations are no longer distinct. This is an example of failure of the rank condition for identification on a particular low-dimensional subset of the parameter space. I believe that any other pair of zeros that doesn't imply singularity raises no problems for identification. (E.g., zeros at (1,1) and (2,4), or (1,1) and (2,2).)*

(c) Suppose we estimate A_0 by first applying OLS to

$$y(t) = \sum_{s=1}^{k} B_s y_{t-s} + u(t)\,;$$

forming $S = \hat{u}'\hat{u}$, the crossproduct of the OLS residuals; and then maximizing

$$(T+4)\log|A_0| - \tfrac{1}{2}\operatorname{trace}\big(A_0' A_0 (S + 4I)\big)\,.$$

Does this produce maximum likelihood estimates? Does it produce estimates that share the asymptotic properties of maximum likelihood? If it is not maximum likelihood, does it differ in a way that has an interpretation?
What is displayed differs from the concentrated (with respect to reduced form coefficients) likelihood function by the logged factor

$$4\log|A_0| - \tfrac{1}{2}\operatorname{trace}\big(4A_0' A_0 I\big)\,.$$

So maximizing the expression displayed in the question would not deliver the MLE. But this extra factor can be interpreted as dummy observations, and thus as part of a conjugate prior. In particular, it would be generated by four observations (the factor in front of $\log|A_0|$) in all of which all the right-hand-side variables are zero and in each of which a different one of the four variables in the system on the left-hand-side is 1, with the others zero. Four dummy observations do not generate a proper prior on the whole system, but they would help rule out anomalous behavior of the likelihood in case S were near-singular.

(d) Describe how you would generate posterior probability bands for the structural impulse responses of this system.

This is covered in the "Error Bands for Impulse Responses" paper by me and Zha. Error bands for reduced form impulse responses can be generated by direct drawing from the posterior, since the posterior is normal-inverse-wixhart, a standard form. In the exactly identified case, one can draw from the posterior on the reduced form parameters and calculate A_0 from $A_0' A_0 = TS^{-1}$ for each draw. The apparent analogue of this for the overidentified case, in which the step of calculating A_0 from the equations connecting it to S/T is replaced by maximizing the concentrated likelihood as if the drawn Σ were S/T, is incorrect. The problem is that in the overidentified case the posterior distribution on Σ is not the same as the posterior we would obtain by ignoring the implied restrictions on Σ. Drawing Σ from the reduced-form posterior ignores the restriction. "Estimating" A_0 by treating the drawn Σ as if it were S/T does get us back in to the restricted parameter space, but in a way that does not reflect the posterior distribution on that restricted space.

So, to do it right, one needs MCMC here. A simple scheme is to use "independence Metropolis-Hastings", with the proposal distribution based on the usual Gaussian approximation to the log likelihood. That is, one makes draws from a distribution centered at the posterior mode and with covariance matrix minus the inverse of the second derivative of the log posterior at the maximum. Usually, to avoid getting stuck through the proposal pdf becoming small relative to the posterior kernel in the tails, one replaces the normal with a corresponding multivariate t distribution with the same or slightly larger variance.

Note that it is only the A_0 draw that needs an MCMC approach. The overall sampling scheme would be Gibbs. We would alternate between drawing from the Gaussian posterior for the reduced form parameters conditional on A_0, and taking an MCMC step to generate a draw from the conditional distribution of A_0 given the reduced form parameters. One could imagine making the MCMC step use a different proposal distribution at each iteration, since the conditional distribution will have a different peak for each draw of the reduced form parameters. However this would require rerunning the posterior maximization algorithm at each draw. It is likely that just using the fixed proposal distribution generated from the posterior mode would work well enough.