ANSWERS TO EXERCISE ON MODEL CHOICE

Here, as a reminder, are the three models.

$$y_t = y_{t-1} + \varepsilon_t \qquad \sigma^2 \sim \sigma^{-2} e^{-1/\sigma^2} \tag{a}$$

$$y_t = c + \rho y_{t-1} + \varepsilon_t \qquad \begin{cases} \sigma^2 \sim \sigma^{-2} e^{-1/\sigma^2} \\ \rho \mid \sigma^2 \sim U(-1,1) \\ c \mid \sigma^2 \sim N(0, 10^4 \sigma^2) \\ \rho \mid \sigma^2 \text{ independent of } c \mid \sigma^2 \end{cases} \tag{b}$$

$$y_t = c + \alpha y_{t-1} + \beta y_{t-2} + \varepsilon_t \qquad \begin{cases} \sigma^2 \sim \sigma^{-2} e^{-1/\sigma^2} \\ \alpha, \beta \sim \text{uniform on } |\beta| < 1,\ \alpha + \beta < 1,\ \beta - \alpha < 1 \\ c \mid \sigma^2 \sim N(0, 10^4 \sigma^2) \\ (\alpha, \beta) \mid \sigma^2 \text{ independent of } c \mid \sigma^2 \end{cases} \tag{c}$$

There was an error in the formulation of this problem: If the prior pdf for σ^2 is actually a prior on σ^2 (not σ as I was thinking when I wrote it down), it is not integrable, and in general model comparison does not work with non-integrable priors. The problem is that a non-integrable prior is unique only up to a scale factor, which is arbitrary. Here, though, because the non-integrable prior is the same for all three models, and because all three models generate integrable posteriors with this non-integrable prior, it is possible to proceed. The results can be thought of as limiting results that would be obtained with priors proportional to $\sigma^{-2-\delta} \exp(-1/\sigma^2)$, as $\delta \downarrow 0$.

The posterior for model (a) is proportional to

$$(2\pi)^{-T/2} \sigma^{-T-2} \exp\left(-\tfrac{1}{2\sigma^2}(u'u + 2)\right), \tag{1}$$

where u is the $T \times 1$ vector with typical element $y_{t+1} - y_t$, $t = 1, \dots, T$. Noting that as a function of σ^2 this is in the shape of an inverse-$\Gamma(T/2, \frac{1}{2}(u'u+2)$, we can see that the σ^2 can be integrated out to yield

$$(2\pi)^{-T/2} \Gamma(T/2) \left(\frac{2}{u'u + 2}\right)^{T/2}. \tag{2}$$

This expression, which now contains no unknown parameters, is the posterior weight on this model.

The posterior for model (b) is

$$\frac{1}{2}(2\pi)^{-(T+1)/2} \sigma^{-T-3} 10^{-2} \exp\left(-\tfrac{1}{2\sigma^2}(u'u + 2 + 10^{-4} c^2)\right), \tag{3}$$

where now the typical element of u is $u_t = y_{t+1} - \rho y_t - c$. Note the leading $1/2$ in the expression, which is the density for the prior on ρ, which spreads evenly over an interval of length 2.

This answer sheet will go beyond what you were asked to do in the problem set by considering the case where the prior on c has a general precision λ^2, rather than the fixed numerical precision $\lambda^2 = 10^{-4}$. This leads to the posterior pdf

$$\frac{1}{2}(2\pi)^{-(T+1)/2}\sigma^{-T-3}\lambda \exp\left(-\tfrac{1}{2\sigma^2}(u'u + 2 + \lambda^2 c^2)\right), \tag{4}$$

With ρ and σ^2 held fixed, the posterior has a Gaussian shape, so it is convenient to integrate out c analytically. If we set $w = y - \rho y_{-1}$ and take $\bar{w}$ to be the sample mean of w, then with c integrated out the posterior pdf in ρ and σ^2 is

$$\frac{1}{2}(2\pi)^{-T/2}\sigma^{-T-2}\lambda\sqrt{T+\lambda^2}\exp\left(-\tfrac{1}{2\sigma^2}((w-\theta\bar{w})'(w-\theta\bar{w})+2)\right), \tag{5}$$

where $\theta = 1 - \lambda/\sqrt{T+\lambda^2}$. Note that with $\lambda = 10^{-2}$, the only place in this formula where λ has an important effect is its appearance as a factor in front of the whole density. The $T+\lambda^2$ term is very close to T and θ is very close to one. So in your answers to this question it would have been OK to consider only the λ factor.

We now have an expression that is almost exactly in the same form as the likelihood for model (a), and we can integrate out the σ^2 as before to obtain

$$\frac{1}{2}(2\pi)^{-T/2}\Gamma(T/2)\lambda\sqrt{T+\lambda^2}\left(\frac{2}{(w-\theta\bar{w})'(w-\theta\bar{w})+2}\right)^{T/2}. \tag{6}$$

This expression is not free of unknown parameters. It still depends, via w and $\bar{w}$, on ρ, and we therefore need to integrate it numerically to obtain the posterior weight on the model.

For the third model (c), the posterior pdf with c and σ^2 integrated out is in exactly the same form (6) as for the second model, except that for the third model we reinterpret w as $w = y - \alpha y_{-1} - \beta y_{-2}$ and the one-half factor in front of the density becomes one-fourth, since in this case the uniform density over the stable region for α and β is spread over a triangle with base 4 and height 2, and thus area 2, so that the uniform density is .25 over this region.

To get the posterior weights on the latter two models requires numerical integration. A program that does the integration, largely due to Piotr Eliasz, is below. Note that though varying the concentration of the prior, here λ, can in principal have a big effect on odds ratios, here the effect of varying λ is limited. I found no value of λ that made the posterior odds on model (a) less than .995. Apparently here the very agnostic priors on ρ and on α,β make the latter two models less plausible.

```
% Piotr Eliasz, Princeton University, 11.13.2002, Modified by Chris Sims 11.19.2002
%clear
lambda=1e-2;
%load ex1data.txt -ascii
%y=log(ex1data(:,2));
y=log(rgdp(:,2));
dates=rgdp(:,1);
T=size(y,1)-2;

% model 1
z=y(3:end)-y(2:end-1);
```

```
%-------------
%plot(dates(3:end),z);pause
z=sum(z.^2)+2;
m1=z^(-T/2);
%m1=(z+2)^(-T/2);

% model 2
% need a grid on rho
rho=[.2:.001:.999];
z1=zeros(size(rho)); z2=zeros(size(rho));
for s=1:T
    insum=repmat(y(2+s),size(rho))-rho.*y(1+s);
    z1=z1+insum.^2;
    z2=z2+insum;
end
z=z1-z2.^2/(T+lambda^2)+2*ones(size(rho));
m2=z.^(-T/2);
m2=m2/2/sqrt(1+T/lambda^2);
%figure; plot(rho,m2)
m2=sum(m2)*.001;
% model 3
% need a grid on alpha and beta
[alpha,beta]=meshgrid(-.5:.01:1.99, -.999:.001:.999);
index=find(beta+abs(alpha) >= 1);
z1=zeros(size(alpha)); z2=zeros(size(alpha));
for s=1:T
    insum=repmat(y(2+s),size(alpha))-alpha.*y(1+s)-beta.*y(s);
    z1=z1+insum.^2;
    z2=z2+insum;
end
z=z1-z2.^2/(T+lambda^2)+2*ones(size(alpha));
m3=z.^(-T/2);
m3=m3/4/sqrt(1+T/lambda^2);
m3(index)=0;
%figure; mesh(alpha,beta,m3)
m3=sum(sum(m3))*.01*.001;

% posterior odds
po=[m1 m2 m3]/(m1+m2+m3)


%This gives:

%po =
%
%    0.99957    0.00014759   0.00028265
```