%Benfordness simulations

%%BEPress results allow heterogeneity in starting populations, currency
%% conversion (to $) rates, and per capita income
%%Below code allows for heterogeneity in initial growth rates of population
%% and per capita income
%%Below code allows for random walks of growth rates of population and
%% and per capita income, as well as random walk of level of currency
%% conversion rates

%start out with N1 countries for N2 years (N=N1*N2)
%run for N3 years before observing

N1=180;%number of countries
N2=50;%years of observation
N3=10;%years of evolution before observation
N4=100;%number of repetitions

rand('state',0)
randn('state',0)

%creating initial conditions: population, currency (thru PPP), & cgdp
lnPop=8.5+1.7*randn(N1,N4);
Pop=exp(lnPop);
lnPPP=1.7+1.5*randn(N1,N4);
PPP=exp(lnPPP);
lnCGD=6+1*randn(N1,N4);
cgdp=exp(lnCGD);

%average initial growth rates
popgr=0.020;
lnCGD=0.07;

%allowing dissimilar initial growth rates
popgr=popgr+0*randn(N1,N4);%s2_PopGR
lnCGD=lnCGD+0*randn(N1,N4);%s2_cgdpGR
cgdpe=exp(lnCGD)-1;

%setting up random walk shocks to initial growth rates & levels
popsd=0;%s2_PopRW
pppsd=0;%s2_PPIRW
cgdp=0;%s2_cgdpRW

pope=popsd*randn(N1,N2+N3,N4);
pppe=pppsd*randn(N1,N2+N3,N4);
cgdp=cgdpsd*randn(N1,N2+N3,N4);

ObsBen=0;
ObsChi2=0;
Rej90=0;
Rej95=0;
Rej99=0;

for r=1:N4;
%population growth rates
g1=popgr(:,r);
for t=1:N2+N3;
g1=[g1 popgr(:,r)+sum(pope(:,1:t,r),2)];
end;
g1=1+g1;

% cgdp change rates
g3=cdgpu(:,r);
for t=1:N2+N3;
    g3=[g3 cdgpu(:,r)+sum(cgdpe(:,1:t,r),2)];
end;
g3=1+g3;

H0=Pop(:,r).*PPP(:,r).*cdgp(:,r); % initial GDP

% finding new GDPS by multiplying initial GDP with appropriate aggregated growth
% rates and currency conversion levels
GDP=H0;
for t=1:N2+N3;
    GDP=[GDP H0.*prod(g1(:,1:t),2).*prod(1+pppe(:,1:t,r),2).*prod(g3(:,1:t),2)];
end;

% removing GDPS before observed sample
ObsGDP=GDP(:,N3+2:N2+N3+1);

yy=benfordcheck(ObsGDP); % subroutine that counts frequency of first digits
Ct=yy(1:9); % frequency of first digits 1 thru 9
N=yy(10); % total number of observations

% Benford first-digit predictions
TrueBen=zeros(9,1);
for i=1:9;
    TrueBen(i)=log10(1+(1/i));
end;

% chi-squared statistic calculation
chi2=N*sum(((Ct/N-TrueBen).^2)./TrueBen);

% counting rejections at various confidence levels
Rej90=Rej90+[chi2>13.36];
Rej95=Rej95+[chi2>15.51];
Rej99=Rej99+[chi2>20.09];

ObsBen=ObsBen+Ct/N; % summing observed first-digit frequencies
ObsChi2=ObsChi2+chi2; % summing chi-squared statistics
disp(r)
end;

ObsBen=ObsBen/N4; % average observed first-digit frequencies
ObsChi2=ObsChi2/N4; % average chi-squared statistic
Rej90=Rej90/N4;
Rej95=Rej95/N4;
Rej99=Rej99/N4;

disp('    Avg%s   PredBenford')
disp([ObsBen TrueBen])
disp('Avg Chi-2 stat')
disp(ObsChi2)
disp(' % rejections (90%, 95% & 99%)')
disp([Rej90 Rej95 Rej99])

% thresholds w/ dof=8
% 95.0% ~ 15.51
% 97.5% ~ 17.54
% 99.0% ~ 20.09
% 99.5% ~ 21.96