14.384 Problem Set 2 Solutions
Fall, 2004
Professor: Michael Jansson
T.A.: Pierre Yared

1 Modeling Trend (continued)

Consider the following time series model

$$\begin{aligned} x_t &= \beta_0 + \beta_1 t + \epsilon_t \\ \epsilon_t &\sim \text{White Noise } (0, \sigma^2) \end{aligned}$$

(a) Use least squares to detrend x_t(Assume β_0 and β_1 in the population are known and determine the predicted $\widehat{x}_t$) What is the autocovariance function of the residual component?

Let me create a new variable $\widetilde{x}_t = x_t - \beta_0 - \beta_1 t$ which represents our detrended variable. Note that since ϵ_t is white noise, we can determine the first

and second moments of $\widetilde{x}_t$ very easily:

$$\begin{aligned} E(\widetilde{x}_t) &= E(\epsilon_t) = 0 \\ Var(\widetilde{x}_t) &= Var(\epsilon_t) = \sigma^2 \\ Cov(\widetilde{x}_t, \widetilde{x}_{t-h}) &= Cov(\epsilon_t, \epsilon_{t-h}) = 0 \ \forall h \neq 0 \end{aligned}$$

(b) Use first differences to detrend x_t. What is the autocovariance function of the residual component?

Note that the above detrending required us to know β_0 and β_1a priori. Since we do not know population parameters typically, another way to isolate the cyclical component of a stochastic process is by taking first differences. Let me create $\Delta x_t = x_t - x_{t-1} = \beta_1 + \epsilon_t - \epsilon_{t-1} = \beta_1 + \Delta\epsilon_t$. Note that this is not an invertible MA process. In particular,

$$\begin{aligned} E(\Delta x_t) &= \beta_1 + E(\Delta\epsilon_t) = \beta_1 \\ Var(\Delta x_t) &= Var(\beta_1 + \Delta\epsilon_t) = Var(\Delta\epsilon_t) = 2Var(\epsilon_t) - Cov(\epsilon_t, \epsilon_{t-1}) = 2\sigma^2 \\ Cov(\Delta x_t, \Delta x_{t-h}) &= Var(\Delta\epsilon_t, \Delta\epsilon_{t-h}) = \begin{cases} -\sigma^2 \text{ if } h = 1 \\ 0 \quad \text{if } h > 1 \end{cases} \end{aligned}$$

Note: in taking first differences, we actually increased the variance of the component we are interested in estimating.

(c) Use first differences to detrend x_t assuming that $\epsilon_t = \rho\epsilon_{t-1} + u_t$. What is the autocovariance function of the residual component?

Note that since $(1-\rho L)\epsilon_t = \mu_t$ that I can now write the process for x_t as:

$$\begin{aligned} x_t &= \beta_0 + \beta_1 t + \frac{\mu_t}{1-\rho L} \implies \\ (1-\rho L)\, x_t &= (1-\rho L)(\beta_0 + \beta_1 t) + \mu_t \implies \\ x_t &= \rho x_{t-1} + (1-\rho L)(\beta_0 + \beta_1 t) + \mu_t \implies \end{aligned}$$

Taking first differences of x_t yields:

$$\begin{aligned} \Delta x_t &= x_t - x_{t-1} = (1-\rho L)\,\beta_1 + \rho \Delta x_{t-1} + \Delta\mu_t \\ &= (1-\rho)\,\beta_1 + \rho\Delta x_{t-1} + \Delta\mu_t \end{aligned}$$

Where I used the fact that $L\beta_1 = \beta_1$ since β_1 is constant in every period. The above equation suggests that Δx_t follows an ARMA(1,1) process. Note that $\Delta x_t - \beta_1 = \rho\Delta x_{t-1} - \rho\beta_1 + \Delta\mu_t$ and we also know that $E(\Delta\mu_t) = 0$. We cano now solved for the first two moment:

$$E(\Delta x_t - \beta_1) = \rho E(\Delta x_{t-1} - \beta_1) + E(\Delta\mu_t) \implies E(\Delta x_{t-1} - \beta_1) = 0 \implies E(\Delta x_{t-1}) = \beta_1$$

$$\begin{aligned} Var(\Delta x_t) &= E(\Delta x_t - \beta_1)^2 = E(\rho(\Delta x_{t-1} - \beta_1) + \Delta\mu_t)^2 \\ &= \rho^2 E(\Delta x_{t-1} - \beta_1) + E\left(\Delta\mu_t^2\right) + 2\rho E(\Delta\mu_t(\Delta x_{t-1} - \beta_1)) \\ &= \rho^2 Var(\Delta x_t) + 2\sigma_\mu^2 + 2\rho E\left(\Delta\mu_t \Delta\mu_{t-1}\right) = \rho^2 Var(\Delta x_t) + 2\sigma_\mu^2 - 2\rho\sigma_\mu^2 \implies \\ Var(\Delta x_t) &= 2\sigma_\mu^2 \frac{1-\rho}{1-\rho^2} = \frac{2\sigma_\mu^2}{1+\rho} \\ Cov(x_t, x_{t-h}) &= E(\Delta x_t - \beta_1)(\Delta x_{t-h} - \beta_1) = E\left(\rho^h(\Delta x_{t-h} - \beta_1) + \sum_{k=0}^{h-1} \rho^k \Delta\mu_{t-k}\right)(\Delta x_{t-h} - \beta_1) \\ &= \rho^h Var(\Delta x_t) + E\left(\left(\sum_{k=0}^{h-1} \rho^k \Delta\mu_{t-k}\right)(\Delta x_{t-h} - \beta_1)\right) \\ &= \rho^h Var(\Delta x_t) + \rho^{h-1} E\left(\Delta\mu_{t-h+1}\Delta\mu_{t-h}\right) \\ &= \sigma_\mu^2 \left(\frac{2\rho^h}{1+\rho} - \rho^{h-1}\right) \text{ for } h > 0 \end{aligned}$$

Now assume that the true model is given by a random walk with drift

$$\begin{aligned} x_t &= \mu + x_{t-1} + \epsilon_t, \qquad t = 1, ...T \\ \epsilon_t &\sim \text{White Noise } (0, \sigma^2) \end{aligned}$$

(d) Use first differences to detrend x_t.What is the autocovariance function of the residual component?

Taking first differences, we obtain: $\Delta x_t = \mu + \epsilon_t - \epsilon_{t-1} = \mu + \Delta\epsilon_t$

$$\begin{aligned} E(\Delta x_t) &= \mu + E(\epsilon_t) = \mu \\ Var(\Delta x_t) &= E(\Delta x_t - \mu)^2 = E\left(\epsilon_t^2\right) = \sigma^2 \\ Cov(\Delta x_t, \Delta x_{t-h}) &= Var(\epsilon_t\,, \epsilon_{t-h}) = 0 \text{ if } h > 0 \end{aligned}$$

(e) Use a linear trend to detrend x_t (write the process introduced in part (d) in a similar way as the one used in part (a)) What is the autocovariance function of the residual component? (Hint: denote $x_0 = \beta_0$ and write the process in terms of μ, β_0 and the ϵ_t's. You can do this using an iterative method starting from $t = 1$).

We let B_0 denote the initial value of x_0 and iterate the process forward:

$$\begin{aligned}
x_1 &= \mu + \beta_0 + \epsilon_t \\
x_2 &= \mu + x_1 + \epsilon_t = 2\mu + \beta_0 + \epsilon_1 + \epsilon_2 \\
x_t &= \beta_0 + \mu t + \sum_{s=1}^{t} \epsilon_s
\end{aligned}$$

Therefore, the detrended process becomes:

$$\widetilde{x}_t = x_t - (\beta_0 + \mu t) = \sum_{s=1}^{t} \epsilon_s$$

and we can write the moments for $\widetilde{x}_t$.

$$\begin{aligned}
E(\widetilde{x}_t) &= 0 \\
Var(\widetilde{x}_t) &= E\left(\widetilde{x}_t^2\right) = E\left(\sum_{s=1}^{t} \epsilon_s\right)^2 = E\left(\sum_{s=1}^{t} \epsilon_s^2\right) = tE\left(\epsilon_t^2\right) = t\sigma^2 \\
Cov(\widetilde{x}_t, \widetilde{x}_{t-h}) &= E(\widetilde{x}_t \widetilde{x}_{t-h}) = E\left(\sum_{s=1}^{t} \epsilon_s\right)\left(\sum_{j=1}^{t-h} \epsilon_j\right) = E\left(\sum_{s=1}^{t-h} \epsilon_s\right)^2 = (t-h)\,\sigma^2
\end{aligned}$$

Therefore, if we try to fit a trend to this non-stationary process, we're going to have an infinitely increasing error term.

(f) Simulate the results obtained for $n = 100, \sigma^2 = 1$, and plot the autocorrelation function for the 2 models considered using the two detrending methods. Comment on the results.

Here is an example of a code you could have used in stata:

```
. set obs 101
obs was 0, now 101
. gen e = invnorm(uniform())
. gen x = .
(101 missing values generated)
. replace x = 13 in 1
(1 real change made)
. replace x = 2 + x[_n-1] + e in 2/101
(100 real changes made)
. keep in 2/101
(1 observation deleted)
. gen time=_n
. tsset time
time variable: time, 1 to 100
. gen detrend1=x[_n]-x[_n-1]
(1 missing value generated)
. gen detrend2=13+2*time
. corrgram detrend1, lags(6)
. graph export "Detrend1.emf", replace
. corrgram detrend2, lags(6)
. graph export "Detrend2.emf", replace
```

As expected, these graphs and the autocorrelation functions show that indeed, the first detrending process by differencing generates a residual term which exhibits little autocorrelation beyond the first partial autocorrelation whereas detrending using a linear trend as in the second case generates a residual term which arithmetically declines over the lag used. See tables and graphs on the next page.

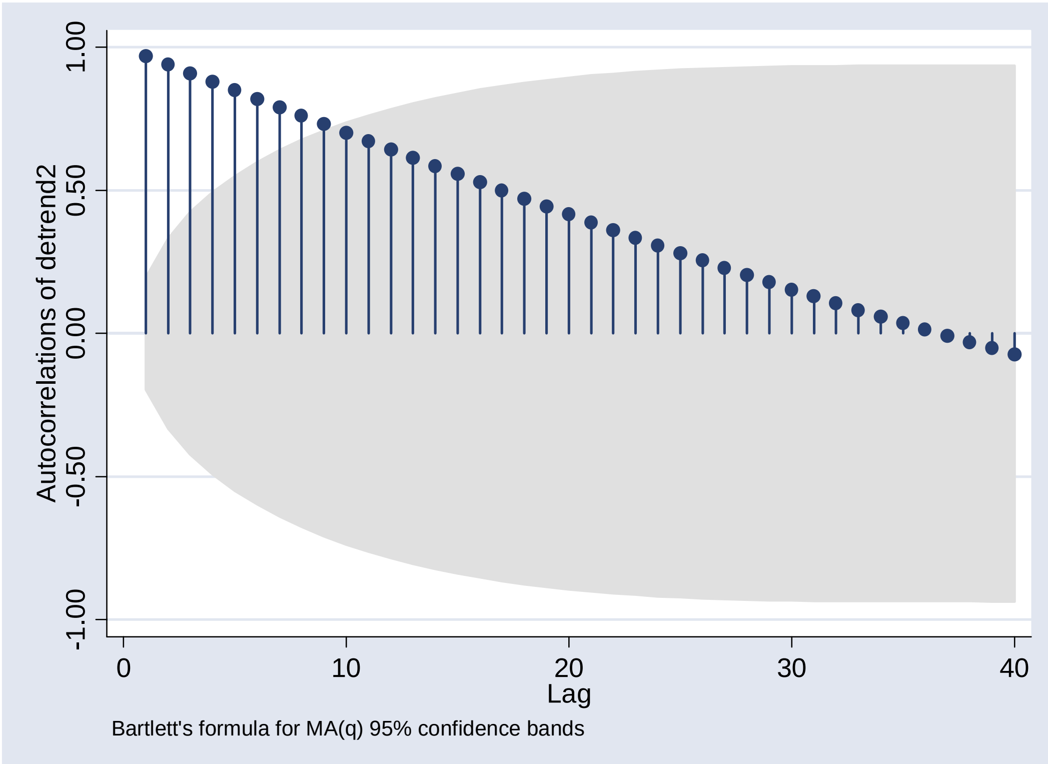
1f

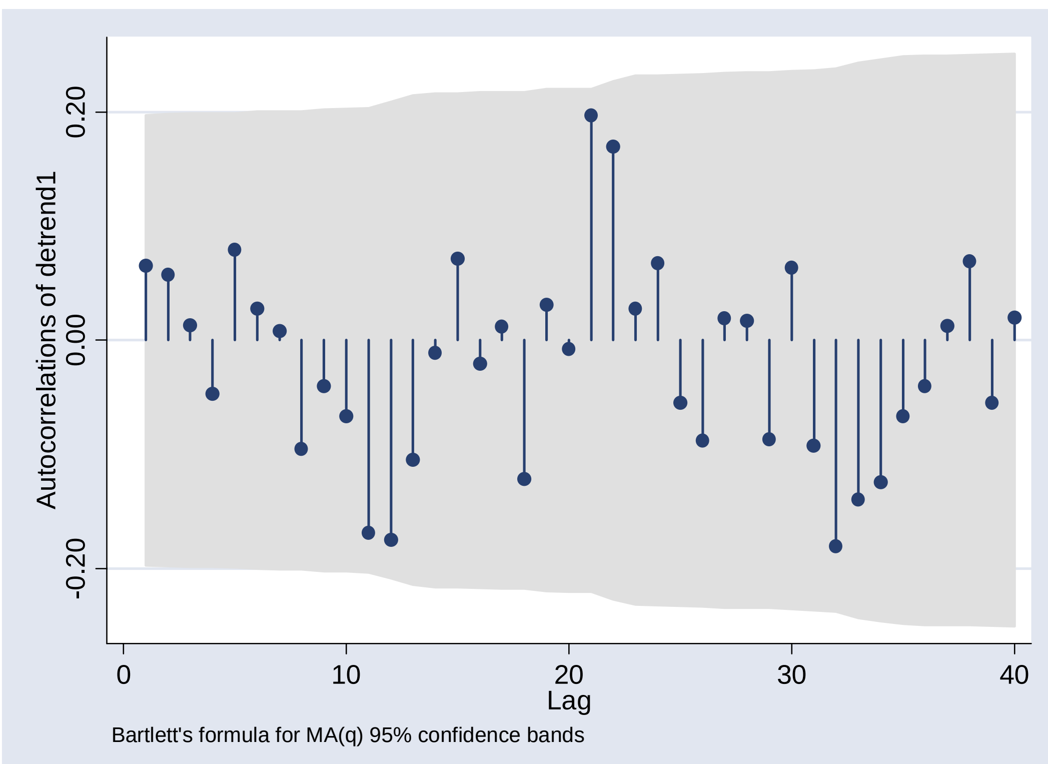
```
. corrgram detrend1, lags(6)

                                          -1       0       1 -1       0
> 1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor
> ]
-------------------------------------------------------------------------------
> -
1        0.0653   0.0655   .4354  0.5093         |                  |
>
2        0.0573   0.0511  .77422  0.6790         |                  |
>
3        0.0133   0.0032  .79252  0.8513         |                  |
>
4       -0.0466  -0.0530  1.0215  0.9065         |                  |
>
5        0.0791   0.0896  1.6866  0.8906         |                  |
>
6        0.0279   0.0215  1.7704  0.9396         |                  |
>

. corrgram detrend2, lags(6)

                                          -1       0       1 -1       0
> 1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor
> ]
-------------------------------------------------------------------------------
> -
1        0.9700        .   96.941  0.0000          |-------
2        0.9400        .   188.91  0.0000          |-------
3        0.9100        .      276  0.0000          |-------
4        0.8801        .    358.3  0.0000          |-------
5        0.8502        .   435.92  0.0000          |------
6        0.8204        .   508.96  0.0000          |------
```


Autocorrelations of detrend2
1.00
0.50
0.00
-0.50
-1.00
0
10
20
30
40
Lag
Bartlett's formula for MA(q) 95% confidence bands


Autocorrelations of detrend1
0.20
0.00
-0.20
0
10
20
30
40
Lag
Bartlett's formula for MA(q) 95% confidence bands

2 Forecasting

Assume that ϵ_t is *iid* with $E\epsilon_t^4 < \infty$, $E\epsilon_t = 0$ and $E\epsilon_t^2 = 1$. Let ϕ_1 and ϕ_2 be such that $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$ has all its roots outside the unit circle. Assume that x_t is the stationary solution to

$$\phi(L)X_t = \epsilon_t$$

(a) You are given a sample $X_1, ..., X_T$ of observations. Find the best linear predictor in the mean square sense of X_{T+1} and X_{T+2}.

Let $\widehat{X}_{T+1}$ and $\widehat{X}_{T+2}$ represent the optimal linear predictor. Since we know $X_1, ..., X_T$, the answer here is very simple:

$$\begin{aligned}
\widehat{X}_{T+1} &= E\left(\phi_1 X_T + \phi_2 X_{T-1} + \epsilon_{T+1} | X_1, ..., X_T\right) = \phi_1 X_T + \phi_2 X_{T-1}, \text{ and} \\
\widehat{X}_{T+2} &= E\left(\phi_1 X_{T+1} + \phi_2 X_T + \epsilon_{T+2} | X_1, ..., X_T\right) \\
&= E\left(\phi_1\left(\phi_1 X_T + \phi_2 X_{T-1} + \epsilon_{T+1}\right) + \phi_2 X_T + \epsilon_{T+2} | X_1, ..., X_T\right) = \left(\phi_1^2 + \phi_2\right) X_T + \phi_1 \phi_2 X_{T-1}
\end{aligned}$$

(b) Calculate the MSE of your forecast conditional on $X_1, ..., X_T$, i.e. calculate $E((X_{T+h} - \hat{X}_{T+h})^2 | X_T, X_{T-1}, ...)$ where $\hat{X}_{T+h}$ is the h-step ahead forecast.

This part is easy to solve using the MA(∞) representation for X_{T+h} (this means that we will assume that we have an infinite number of time period going back).

$$X_{T+h} = \phi_1 X_{T+h-1} + \phi_2 X_{T+h-2} + \epsilon_{T+h} = \sum_{j=0}^{\infty} \Psi_j \epsilon_{T+h-j}$$

The best linear predictor for this process can be written as:

$$\widehat{X}_{T+h} = E\left(\sum_{j=0}^{\infty} \Psi_j \epsilon_{T+h-j} | X_{-\infty}, ..., X_T\right) = \sum_{j=h}^{\infty} \Psi_j \epsilon_{T+h-j}$$

where we have used the fact that we have observed $\epsilon_{-\infty}$ to ϵ_T and that $E\left(\epsilon_{T+h-j} | X_{-\infty}, ..., X_T\right) = 0$. We can therefore write the prediction error $\widehat{\epsilon}_{T+h}$ very easily:

$$\widehat{\epsilon}_{T+h} = X_{T+h} - \widehat{X}_{T+h} = \sum_{j=0}^{\infty} \Psi_j \epsilon_{T+h-j} - \sum_{j=h}^{\infty} \Psi_j \epsilon_{T+h-j} = \sum_{j=0}^{h-1} \Psi_j \epsilon_{T+h-j}$$

so that prediction error $\widehat{\epsilon}_{T+h}$ is a $MA\left(h-1\right)$ process. Now to calculate the MSE of the process:

$$\begin{aligned}
Var\left(\widehat{\epsilon}_{T+h} | X_{-\infty}, ..., X_T\right) &= E\left(\widehat{\epsilon}_{T+h}^2 | X_{-\infty}, ..., X_T\right) = E\left(\left(\sum_{j=0}^{h-1} \Psi_j \epsilon_{T+h-j}\right)^2 | X_{-\infty}, ..., X_T\right) \\
&= \sum_{j=0}^{h-1} \Psi_j^2 E\left(\epsilon_{T+h-j}^2 | X_{-\infty}, ..., X_T\right) = \sum_{j=0}^{h-1} \Psi_j
\end{aligned}$$

((()))

(c) Find the unconditional MSE of the forecast.

Since $E\left(\epsilon_{T+h-j}^2 | X_{-\infty}, ..., X_T\right) = E\left(\epsilon_{T+h-j}^2\right)$ the conditional and uncondtional MSE must be equal. This is a consequence of ϵ_{T+h-j} being iid. Therefore

$$Var\left(\widehat{\epsilon}_{T+h}\right) = \sum_{j=0}^{h-1} \Psi_j$$

3 Simulation

Consider the following time series model

$$x_{t+1} = 0.5x_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{iid } N(0,1).$$

(a) Compute the theoretical ACF for the process.

We first note that the process is stationary by determining the roots of the polynomial:

$$\begin{aligned}\left(1 - \frac{1}{2}L^2\right) x_{t+1} &= \epsilon_t \implies \left(1 - \frac{1}{2}z^2\right) x_{t+1} = \epsilon_t \implies \\ z &= \pm\sqrt{2} \text{ and since } |z| > 1, \text{ the process is stationary}\end{aligned}$$

To determine the ACF,note that $E\left(x_{t+1}\right) = E\left(0.5x_{t-1} + \epsilon_t\right) = 0.5E\left(x_{t-1}\right)$, which implies that $E\left(x_t\right) = 0$. To determine the ACF, we follow the same steps as we did in the first problem set:

$$\begin{aligned}\gamma_0 &= Var\left(x_t\right) = E\left(x_t^2\right) = (0.5)^2 E\left(x_{t-1}^2\right) + E\left(\epsilon_t^2\right) + 2E\left(x_{t-1}\epsilon_t\right) = 0.25\gamma_0 + 1 \\ &= 4/3 \\ \gamma_1 &= Cov\left(x_{t+1}, x_t\right) = 0.5E\left(x_{t-1}x_t\right) + E\left(\epsilon_t x_t\right) = 0.5\gamma_1 \implies \gamma_1 = 0 \\ \gamma_h &= Cov\left(x_{t+1}, x_{t+1-h}\right) = 0.5E\left(x_{t-1}x_{t+1-h}\right) + 0.5E\left(\epsilon_t x_{t+1-h}\right) = 0.5\gamma_{h-2} \text{ for } h > 1\end{aligned}$$

We therefore have a difference equation $\gamma_h = \gamma_{h-2}$ with initial conditions $\gamma_0 = 4/3$ and $\gamma_1 = 0$. Such a difference equations as noted in the first problem set has the general solution

$$\gamma_h = c_1\left(\frac{1}{\sqrt{2}}\right)^h + c_2\left(-\frac{1}{\sqrt{2}}\right)^h$$

Using this equation and our initial conditions, this yields: $c_1 = c_2 = 2/3$. Therefore, the process yields: $\gamma_h = \frac{2}{3}\left(\left(\frac{1}{\sqrt{2}}\right)^h + \left(-\frac{1}{\sqrt{2}}\right)^h\right)$, so that

$$\gamma_h = \begin{cases} \frac{4}{3}\left(\frac{1}{\sqrt{2}}\right)^h & \text{if } h \text{ is even} \\ 0 & \text{if } h \text{ is odd}\end{cases}$$

(b) Simulate a series of length 1000 from this process. Calculate the sample autocovariance. Does the theoretical autocovariance look close to the sample one?

Here is a sample program written for stata:

```
. set obs 2000
obs was 0, now 2000
. gen e = invnorm(uniform())
. gen x = .
(2000 missing values generated)
. replace x = 13 in 1
(1 real change made)
. replace x = 10 in 2
(1 real change made)
. replace x = .5*x[_n-2] + e in 3/2000
(1998 real changes made)
. keep in 1001/2000
(1000 observations deleted)
. gen time=_n
. tsset time
time variable: time, 1 to 1000
. corrgram x, lags(6)
. graph export "ps2_2.emf", replace
```

The table and graph are displayed at the end of the problem set. As you can see autocorrelations are very high for even lags but close to zero for odd lags. This can be seen both with the autocorrelation table and the graph of the autocorrelations.

(c) Simulate 100 different paths of size 1000 from this process. Estimate the first order autocovariance. Plot an histogram of the sample estimated first order autocovariance from this process.

Here is a sample program written for stata:

```
clear
set seed 1
program sim, rclass
// Generate 1000 observations
drop _all
set obs 2000
gen t = _n
tsset t
gen e = invnorm(uniform())
gen x = 0
forvalues i = 3/2000 {
replace x = 0.5*x[`i'-1] + e in `i'
}
```

```
keep in 1001/2000
// Estimate the autocorrelation
corrgram x, noplot lags(1)
return scalar autocorr = r(ac1)
end
simulate "sim" ac = r(autocorr), reps(100) dots
kdensity ac
graph export "ps2_3.emf", replace
```

Note that this looks like a normal distribution.

(d) Find the theoretical asymptotic distribution of the sample first order autocovariance.

Note that for a vector of T observations from X_2 to X_{T+1}, we can write the process in question as $X_{T+1} = \frac{1}{2}X_{T-1} + \epsilon_T$ where

$$X_{T+1} = \begin{bmatrix} x_{t+1} \\ x_t \\ \cdots \\ x_2 \end{bmatrix}, \; X_{T-1} = \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ \cdots \\ x_0 \end{bmatrix}, \text{ and } \epsilon_T = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \\ \cdots \\ \epsilon_1 \end{bmatrix}$$

We can write the first order sample autocovariance (which corresponds to the theoretical $Cov(x_{t+1}, x_{t-1})$ as:

$$\begin{aligned}
\widehat{\gamma}_{T+1}(1) &= \frac{1}{T-1}X'_{T+1}X_T = \frac{1}{T-1}\left(\frac{1}{2}X'_{T-1}X_T + \epsilon'_T X_T\right) \\
&= \frac{1}{T-1}\left(\frac{1}{2}X'_{T+1}X_T + \frac{1}{2}x_0x_1 - \frac{1}{2}x_{T+1}x_T + \epsilon'_T X_T\right) \\
&= \frac{1}{2}\widehat{\gamma}_{T+1}(1) + \frac{1}{T-1}\left(\frac{1}{2}x_0x_1 - \frac{1}{2}x_{T+1}x_T + \epsilon'_T X_T\right) \implies \\
&= \frac{1}{T-1}\left(x_0x_1 - x_{T+1}x_T\right) + \frac{2}{T-1}\epsilon'_T X_T
\end{aligned}$$

We would like to determine the asymptotic distribution of $\widehat{\gamma}_{T+1}(1)$. In order to do this, we look at the distribution of its components. We focus on $\frac{2}{T-1}\epsilon'_T X_T$. Note the following:

$$\begin{aligned}
E\left(\epsilon'_T X_T\right) &= \frac{2}{T-1}E\left(\sum_{t=1}^{T} x_t\epsilon_t\right) = 0 \\
Var\left(\epsilon'_T X_T\right) &= E\left(\epsilon'_T X_T\right)^2 = E\left(\sum_{t=1}^{T} x_t\epsilon_t\right)^2 = E\left(\sum_{t=1}^{T} x_t^2\epsilon_t^2\right) \text{ (since } \epsilon_t \text{ is iid)} \\
&= \left(\sum_{t=1}^{T} x_t^2(1)\right) = E\left(x_t^2\right) = T\gamma(0) \\
&\quad \text{where } \gamma(0) \text{ is the variance of } x_t\text{, which we know from part 1 is } 4/3
\end{aligned}$$

By standard CLT,

$$\frac{1}{\sqrt{T-1}}E\left(\sum_{t=1}^{T} x_t\epsilon_t\right) \to_d N\left(0, \frac{T\gamma(0)}{T-1}\right) = N(0, \gamma(0))$$

Now, looking back at our expression for $\widehat{\gamma}_{T+1}(1)$, we premultiply it by $\sqrt{T-1}$ and achieve:

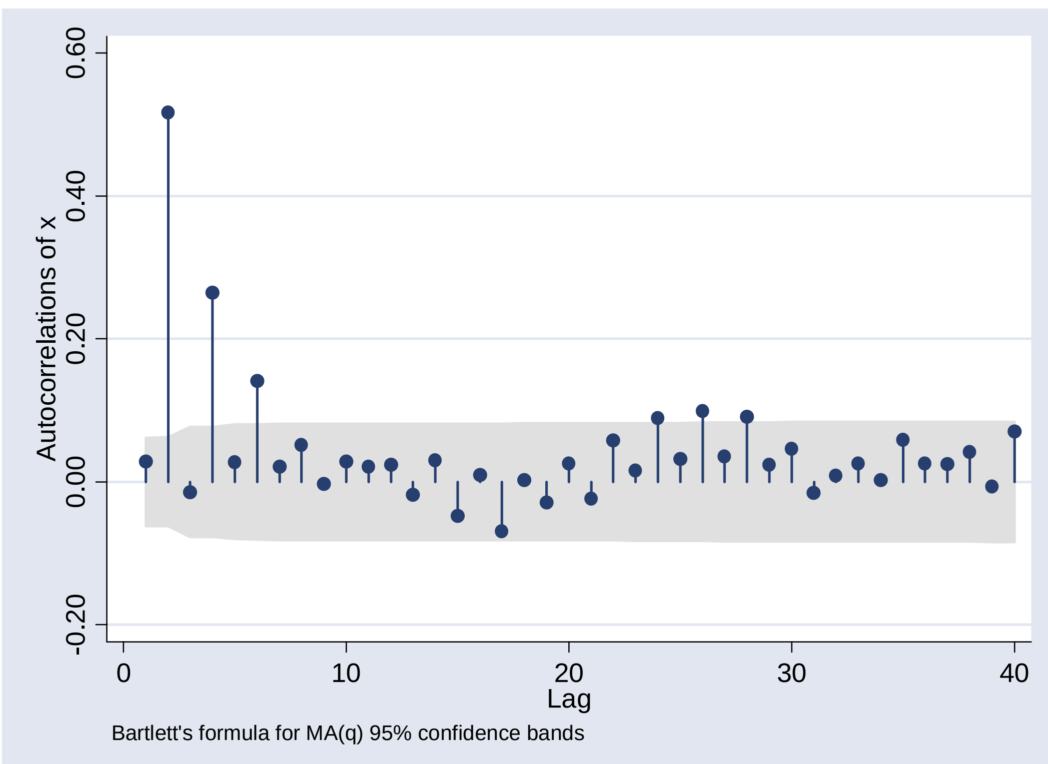
$$\sqrt{T-1}\widehat{\gamma}_{T+1}(1) = \frac{1}{\sqrt{T-1}}\left(x_0x_1 - x_{T+1}x_T\right) + \frac{2}{\sqrt{T-1}}\epsilon'_T X_T$$

The first part converges in probability to zero, and the second part converges in distribution to $N\left(0, 4\gamma\left(0\right)\right) = N\left(0, 16/3\right)$. Therefore the asymptotic distribution of $\widehat{\gamma}_{T+1}\left(1\right)$ can be characterized as

$$\widehat{\gamma}_{T+1}\left(1\right) \sim N\left(0, \frac{16}{3\left(T-1\right)}\right)$$

3b

```
                                                            -1       0       1 -1       0
> 1
 LAG         AC        PAC        Q      Prob>Q  [Autocorrelation]  [Partial Autocor
> ]
-------------------------------------------------------------------------------
> -
1         0.0287   0.0287   .82678  0.3632          |                  |
>
2         0.5174   0.5178   269.57  0.0000          |----              |----
>
3        -0.0149  -0.0507    269.8  0.0000          |                  |
>
4         0.2643  -0.0018   340.09  0.0000          |--                |
>
5         0.0275   0.0694   340.85  0.0000          |                  |
>
6         0.1412   0.0018   360.95  0.0000          |-                 |
>
```



3c