MS&E349 Homework 1

Tracy Cao, Sven Lerner & Jordan Rowley

---

# Theoretical Questions

## Question 1

Given the ACF of the undifferenced data, the time series does not appear stationary.

When we difference the data, both the ACF and PACF indicate that the model depicts a moving average process. In particular, the fact that the ACF "dies off" past lag 1 indicates that the model is of the form
ARIMA(0, 1, 1).

As a rule of thumb, we know that the likely signs of an $MA(q)$ model consist in an ACF that "dies off" after lag $q$ and a PACF that exhibits exponential decay. On the other hand, the likely signs of an $AR(p)$ model consist in an ACF that exponentially decays and a PACF that "dies off" after lag $p$.

## Question 2

### 2.1

Note that

$$\begin{aligned}\mathbb{E}[h_1(X_t,\phi)] &= \mathbb{E}[Y_{t-1}Y_t - \phi Y_{t-1}^2]\\ &= \mathbb{E}[Y_{t-1}Y_t] - \phi\mathbb{E}[Y_{t-1}^2]\\ &= \gamma(0) - \phi\gamma(1)\end{aligned}$$

This is only satisfied when $\phi = \phi_0$

Now seeing that $Y_t = \epsilon_t + \phi_0 Y_{t-1}$

$$\begin{aligned}\mathbb{E}[h_2(X_t,\phi)] &= \mathbb{E}[Y_{t-2}(\epsilon_t + \phi_0 Y_{t-1} - \phi Y_{t-1})]\\ &= \mathbb{E}[Y_{t-2}(\phi_0 Y_{t-1} - \phi Y_{t-1})]\\ &= \phi_0\gamma(1) - \phi\gamma(1)\end{aligned}$$

This is only satisfied when $\phi = \phi_0$

We see that in this case $\Phi = \Phi_1 \cap \Phi_2 = \{\phi_0\}$

### 2.2

We see that

$$\hat{\phi}_1 = \arg\min \phi(\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}Y_t - \phi Y_{t-1}^2))^2$$

First order conditions give us that

$$\hat{\phi}_1 = \frac{\sum_{t=1}^T Y_{t-1}Y_t}{\sum_{t=1}^T Y_{t-1}^2}$$

We now see that

$$\begin{aligned}\sqrt{T}(\hat{\phi}_1 - \phi_0) &= \sqrt{T}\frac{\sum_{t=1}^T Y_{t-1}Y_t - \phi_0\sum_{t=1}^T Y_{t-1}^2}{\sum_{t=1}^T Y_{t-1}^2}\\ &= \sqrt{T}\frac{\sum_{t=1}^T Y_{t-1}(Y_t - \phi_0 Y_{t-1})}{\sum_{t=1}^T Y_{t-1}^2}\\ &= \sqrt{T}\frac{\sum_{t=1}^T Y_{t-1}(\epsilon_t)}{\sum_{t=1}^T (Y_{t-1}^2}\\ &= \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T Y_{t-1}(\epsilon_t)}{\frac{1}{T}\sum_{t=1}^T Y_{t-1}^2}\end{aligned}$$

We note that the numerator converges in distribution to $\mathcal{N}(0, Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_{t-1}\epsilon_t))$ and the denominator converges in probability to $\gamma_0$ so

$$\sqrt{T}(\hat{\phi}_1 - \phi_0) \xrightarrow{d} \mathcal{N}(0, \frac{1}{\gamma(0)^2} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_{t-1}\epsilon_t))$$

At this point, let's get a better handle on that variance term

We see that $\mathbb{E}[Y_{t-1}\epsilon_t] = 0$ so we have

$$\mathbb{E}[\sum_{t=1}^{T} Y_{t-1}\epsilon_t] = \mathbb{E}[\sum_{t=1}^{T} Y_{t-2}\epsilon_t] = 0$$

This gives us

$$Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}) \begin{pmatrix} Y_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} \mathbb{E}[(\sum_{t=1}^{T} Y_{t-1}\epsilon_t)^2] & \mathbb{E}[(\sum_{t=1}^{T} Y_{t-1}\epsilon_t)(\sum_{t=1}^{T} Y_{t-2}\epsilon_t)] \\ \mathbb{E}[(\sum_{t=1}^{T} Y_{t-1}\epsilon_t)(\sum_{t=1}^{T} Y_{t-2}\epsilon_t)] & \mathbb{E}[(\sum_{t=1}^{T} Y_{t-1}\epsilon_t)^2] \end{pmatrix}$$

Now we see

$$\begin{aligned}
\mathbb{E}[(\sum_{t=1}^{T} Y_{t-1}\epsilon_t)^2] &= \mathbb{E}[(\sum_{t=1}^{T} Y_{t-1}\epsilon_t)(\sum_{t=1}^{T} Y_{t-1}\epsilon_t)] \\
&= \mathbb{E}[(\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} Y_{t_1-1}\epsilon_{t_1} Y_{t_2-1}\epsilon_{t_2})] \\
&= (\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} \mathbb{E}[Y_{t_1-1}\epsilon_{t_1} Y_{t_2-1}\epsilon_{t_2})] \\
&= (\sum_{t_1=1}^{T} \mathbb{E}[Y_{t_1-1}\epsilon_{t_1} Y_{t_1-1}\epsilon_{t_1})] \\
&= \sum_{t_1=1}^{T} \gamma(0)
\end{aligned}$$

Similar steps give us

$$\mathbb{E}[(\sum_{t=1}^{T} Y_{t-2}\epsilon_t)^2] = \sum_{t=1}^{T} \gamma(0)$$

$$\mathbb{E}[(\sum_{t=1}^{T} Y_{t-1}\epsilon_t)(\sum_{t=1}^{T} Y_{t-2}\epsilon_t)] = \sum_{t_1=1}^{T} \gamma(1)$$

We can also see that

$$\gamma(0) = \sum_{i=0}^{\infty} (\phi_0^i)^2 = \frac{1}{1-\phi_0^2}$$

$$\gamma(1) = \sum_{i=0}^{\infty}(\phi_0^{i+1})\phi_0^i = (\phi_0)\sum_{i=0}^{\infty}(\phi_0^i)^2 = \frac{\phi_0}{(1-\phi_0^2)}$$

Coming back to the initial quantity of interest we have

$$\sqrt{T}(\hat{\phi}_1 - \phi_0) \xrightarrow{d} \mathcal{N}(0, \frac{1}{\gamma(0)}$$

or

$$\sqrt{T}(\hat{\phi}_1 - \phi_0) \xrightarrow{d} \mathcal{N}(0, 1-\phi_0^2)$$

**2.3**

Similarly we note

$$\hat{\phi}_2 = \arg\min \phi(\frac{1}{T}\sum_{t=1}^{T}(Y_{t-2}Y_t - \phi Y_{t-1}Y_{t-2}))^2$$

First order conditions give us that

$$\hat{\phi}_2 = \frac{\sum_{t=1}^{T} Y_{t-2}Y_t}{\sum_{t=1}^{T} Y_{t-1}Y_{t-2}}$$

so we have

$$\begin{aligned}\sqrt{T}(\hat{\phi}_2 - \phi_0) &= \sqrt{T}\frac{\sum_{t=1}^{T} Y_{t-2}Y_t - \phi_0\sum_{t=1}^{T} Y_{t-1}Y_{t-2}}{\sum_{t=1}^{T} Y_{t-1}Y_{t-2}} \\ &= \sqrt{T}\frac{\sum_{t=1}^{T} Y_{t-2}(Y_t - \phi_0 Y_{t-1})}{\sum_{t=1}^{T} Y_{t-1}Y_{t-2}} \\ &= \sqrt{T}\frac{\sum_{t=1}^{T} Y_{t-2}\epsilon_t}{\sum_{t=1}^{T} Y_{t-1}Y_{t-2}}\end{aligned}$$

reusing the derivations from the last part we have

$$\sqrt{T}(\hat{\phi}_1 - \phi_0) \xrightarrow{d} \mathcal{N}(0, \frac{\gamma(0)}{\gamma(1)^2})$$

or

$$\sqrt{T}(\hat{\phi}_1 - \phi_0) \xrightarrow{d} \mathcal{N}(0, \frac{1-\phi_0^2}{\phi_0^2})$$

As we must have that $|\phi_0| < 1$ the first estimator is asymptotically more efficient.

### 2.4

A sufficient condition for $g(\phi_W) = 0$ would be that the matrix $W$ is positive definite, and the estimates for $\phi$ obtained in 2.2 and 2.3 are actually equal, that is, $\hat{\phi}_1 = \hat{\phi}_2$.

Now the first order conditions give that $g(\hat{\phi}_1) = g(\hat{\phi}_2) = 0$, and since $W$ is positive definite, we know that $g(\phi_W) = 0$ is the unique minimizer of the scalar product $g^T W g$, hence $\hat{\phi_k} = \phi_W$, for $k = 1, 2$.

Note that we have imposed a pathological amount of structure on $Y_t$ in requiring that $\hat{\phi}_1 = \hat{\phi}_2$.

### 2.5

The CLT tells us that $\sqrt{T}(g_T(\phi)) \xrightarrow{d} N(0, S)$ for some covariance matrix $S$. Consider $d^T W g_T(\phi)$, where $d = \frac{\partial}{\partial \phi} g_T(\phi)$. Then as $T \to \infty$

$$\mathrm{Var}[d^T W g_T(\phi)] = d^T W S [d^T W]^T$$

Since

$$d^T W g_T(\phi) \approx d^T W g_T(\phi_0) + d^T W d(\phi - \phi_0)$$

Then

$$\mathrm{Var}[d^T W g_T(\phi_0) + d^T W d(\phi - \phi_0)] = d^T W d \mathrm{Var}[\phi][d^T W d]^T$$

Therefore, for sufficiently large $T$,

$$d^T W S [d^T W]^T \approx d^T W d \mathrm{Var}[\phi][d^T W d]^T$$

Now $d \xrightarrow{p} [\gamma(0), \gamma(1)] = \gamma(0)[1, \phi_0]$. Therefore, $dWd^T > 0$. This allows us to solve for $\mathrm{Var}[\phi]$:

$$\mathrm{Var}[\phi] = (d^T W d)^{-1}(d^T W) S (W d)(d^T W d)^{-1}$$

### 2.6

How can we choose $W$ to minimize the variance of our estimator?. When we plug in $W = S^{-1}$, we obtain an elegant expression for the variance:

$$\mathrm{Var}[\phi] = (d^T S^{-1} d)^{-1}$$

We will show that this is indeed the minimal variance that can be obtained for any choice of positive definite matrix $W$. Let the matrix function $V$ be defined by $V(W) = (d^T W d)^{-1}(d^T W) S (W d)(d^T W d)^{-1}$. We will show that the difference $V(W) - V(S^{-1})$ is a positive semi-definite matrix. Let $U = (d^T W d)^{-1}$. For ease of notation, denote $U = d^T W d$. Then it is sufficient to show that

$$U^{-1}d^TWSWdU^{-1} - (d^TS^{-1}d)^{-1} \geq 0$$

Borrowing a few terms, we can rewrite this difference into the form

$$U^{-1}d^TWS^{1/2}[I - S^{-1/2}dU^{-1}U(d^TS^{-1}d)^{-1}UU^{-1}d^TS^{-1/2}]S^{1/2}WdU^{-1}$$

which simplifies to

$$U^{-1}d^TWS^{1/2}[I - S^{-1/2}d(d^TS^{-1}d)^{-1}d^TS^{-1/2}]S^{1/2}WdU^{-1}$$

Let $M = U^{-1}d^TWS^{1/2}$ and $N = S^{-1/2}d(d^TS^{-1}d)^{-1}d^TS^{-1/2}$. Notice that

$$N^2 = S^{-1/2}d(d^TS^{-1}d)^{-1}d^TS^{-1/2}S^{-1/2}d(d^TS^{-1}d)^{-1}d^TS^{-1/2} = S^{-1/2}d(d^TS^{-1}d)^{-1}d^TS^{-1/2} = N$$

Hence $N$ is idempotent. Therefore, so also is $I - N$. Since $N$ is symmetric, our original difference may be re-expressed into the form

$$M(I - N)M^T = M(I - N)(I - N)M^T = [M(I - N)][M(I - N)]^T$$

Whether $M$ is a matrix or a vector, we know that $[M(I-N)][M(I-N)]^T \geq 0$ (in the first case the inequality means positive-semi-definiteness, while in it indicates a non-negative scalar). Therefore, $S^{-1}$ is indeed the value for $W$ that minimizes the variance of the estimator!