Sheet 7

1 $X_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t$ where ε_t is white noise. $\perp\!\!\!\perp X_0$

a) Note that in the above formula, we can express X_{t-1} as

$$\begin{cases} X_0 & \text{if } t=1 \\ \beta_0 + \beta_1 X_{t-2} + \varepsilon_{t-1} & \text{otherwise} \end{cases}$$

(Letting $s=0$ for ease)

and continue "rolling back" and re-substituting (*) along the lines of the following:

$$X_t = \beta_0 + \beta_1(\beta_0 + \beta_1 \overbrace{X_{t-2}}^{(*)} + \varepsilon_{t-1}) + \varepsilon_t$$

(*) What happens when we have $|\beta_1|<1$ but Y_0 doesn't meet the relevant conditions?

The effect of the initial condition becomes negligible at t gets large.

But then we can see that once all the nested brackets are expanded, we have

$$= (\beta_0 + \beta_1\beta_0 + \beta_1^2\beta_0 + \dots) + (\beta_1^t X_0) + (\varepsilon_t + \beta_1\varepsilon_{t-1} + \dots)$$

$$= \beta_0 \sum_{i=0}^{t-1} \beta_1^i + \beta_1^t X_0 + \sum_{i=0}^{t-1} \beta_1^i \varepsilon_{t-i}$$

Also, for AR(p) models why is the stationarity rule $[\sum \beta_i < 1$, without any abs(·) operators?

This is necessary but not sufficient for stationarity.

And for the case where $s>0$, we're just shifting where we decide to start counting from, i.e. we don't necessarily need to roll back all the way to X_0 since X_s contains all the relevant information to generate X_{s+k}.

So clearly X_t does not have any dependence on future values of $\{\varepsilon_t\}$, as one would expect, since our sum includes only past terms. If we took the conditional expectation of ε_{t+1} given X_t, we'd just get $0=\mu$.

b) $\mathbb{E}[X_t] = \beta_0 \sum_{i=0}^{t-1} \beta_1^i + \beta_1^t \mathbb{E}[X_0] + \mathbb{E}[\varepsilon_t] + \underbrace{\mathbb{E}[\beta_1\varepsilon_{t-1}]}_{\beta_1\mathbb{E}[\varepsilon_{t-1}]} + \dots$ by linearity of $\mathbb{E}$

$$= \cancel{\beta_0 \sum_{i=0}^{t-1} \beta_1^i} + \frac{\beta_1^t \beta_0}{1-\beta_1} + \beta_0 \cdot \frac{1-\beta_1^t}{1-\beta_1}$$ as ε has $\mu=0$, iid

$$= \frac{\beta_0}{1-\beta_1} \text{ for every } t$$

$$\mathrm{Var}(X_t) = \mathrm{Var}(\beta_1^t X_0) + \mathrm{Var}\left(\sum_{i=0}^{t-1} \beta_1^i \varepsilon_{t-i}\right) \quad \text{since } X_0 \perp\!\!\!\perp \varepsilon$$

$$= \beta_1^{2t}\,\mathrm{Var}(X_0) + \sigma_\varepsilon^2 \cdot \sum_{i=0}^{t-1} \beta_1^{2i} \quad \text{as } \mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2$$

$$= \sigma_\varepsilon^2 \cdot \beta_1^{2t} / (1-\beta_1^2) + \sigma_\varepsilon^2 \cdot \frac{(1-\beta^{2t})}{(1-\beta_1^2)}$$

$$= \frac{\sigma_\varepsilon^2}{1-\beta_1^2} \quad \text{for all } t$$

$$\mathrm{Cov}(X_s, X_{s+t}) = \mathrm{Cov}\Big(X_s, \underbrace{\beta_0 \sum_{i=0}^{t-1} \beta_1^i}_{\text{constant}} + \beta_1^t X_s + \sum_{i=0}^{t-1} \beta_1^i \varepsilon_{s+t-i}\Big)$$

(technically)

$$= \underbrace{\beta_1^{|t|}\,\mathrm{Var}(X_s)}_{} + \underbrace{\mathrm{Cov}(X_s, \beta_1 \varepsilon_{s+t}) + \mathrm{Cov}(X_s, \beta_1 \varepsilon_{s+t-1}) + \cdots}_{}$$

$\mathrm{Var}(X_s)$ is constant for all s as shown above, so ACF depends only on lag (here, t), not time.

as established, these covariance terms $= 0$ because X_s is independent of future values of ε

So yes, $\{X_t\}$ is weakly stationary: μ and σ^2 are $\perp\!\!\!\perp$ t, and ACF is time-invariant.

c) i. $X_t = t\beta_0 + \sum_{i=1}^{t} \varepsilon_i$

ii. $\mathbb{E}[X_t] = t\beta_0$ which depends on t, so not time-invariant and thus already fails stationarity

$\mathrm{Var}(X_t) = \mathrm{Var}\left(\sum_{i=1}^{t} \varepsilon_t\right) = t \cdot \sigma_\varepsilon^2$ (as iid)

d) $\mathbb{E}[X_{s+h} \mid X_s, X_{s-1}, \dots] = \mathbb{E}\left[\beta_0 \sum_{i=0}^{h-1} \beta_1^i + \beta_1^h X_s + \sum_{i=0}^{h-1} \beta_1^i \varepsilon_{s+h-i} \mid X_s, X_{s-1}, \dots\right]$

Nice!

$$= \beta_0 \cdot \frac{1-\beta_1^h}{1-\beta_1} + \beta_1^h X_s + \cancel{\mathbb{E}\left[\sum \beta_1^i \varepsilon_{s+h-i}\right]}$$

$$= (1-\beta_1^h)\mu + \beta_1^t X_s \quad \text{where } \mu := \mathbb{E}[X_s] = \frac{\beta_0}{1-\beta_1} \text{ for in (b)}$$

e) $\mathbb{E}[(X_{s+h} - X_{s+h|s})^2] = \mathbb{E}[(\sum_{i=0}^{h-1} \beta_1^i \varepsilon_{s+h-i})^2]$

expanding this seems tricky, not sure how to.

~~= $\mathbb{E}$[someso~~

· All the cross terms, i.e.

$\mathbb{E}\ (\beta_1^i \varepsilon_{s+h-i})(\beta_1^j \varepsilon_{s+h-j}),\ i \neq j$

have mean zero, by independence of each ε_t and properties of joint distributions.

· Only the squared terms remain, i.e. $\sum_{i=0}^{h-1} (\beta_1^i \varepsilon_{s+h-i})^2$

So we have $\mathbb{E}[(\varepsilon_{s+h})^2] + \beta_1^2 \mathbb{E}[(\varepsilon_{s+h-1})^2] + \dots$

Note that $Var(\varepsilon_t) := \mathbb{E}[\varepsilon_t^2] - \mathbb{E}[\varepsilon_t]^2 = \mathbb{E}[\varepsilon_t^2]$, and they're iid.

Great

Which means this MSFE $= \left(\sum_{i=0}^{h-1} \beta_1^{2i}\right) \cdot \sigma_\varepsilon^2$ ∎

For $|\beta_1| < 1$, our geometric series reaches a limit $\in \mathbb{N}$ over longer horizons. For $\beta_1 = 1$, MSFE is unbounded and $= h\sigma_\varepsilon^2$.
In either case, error increases with horizon.

In particular, the best long-term forecast is just the mean μ, and our MSFE $\to \frac{1}{1-\beta_1^2} \cdot \sigma_\varepsilon^2 = Var(X_t)$ as established in (b).

2. $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$

a) $\mathbb{E}[X_t] = \mathbb{E}[\cdot] = \mathbb{E}[\varepsilon_t] + \theta_1 \mathbb{E}[\varepsilon_{t-1}] + \dots = 0$ by linearity

$Var(X_t) = Var(\varepsilon_t) + \theta_1^2 Var(\varepsilon_{t-1}) + \dots + \theta_q^2 Var(\varepsilon_{t-q})$ by $\{\varepsilon_t\}$ iid so 0 covariance
$= \left(\sum_{i=0}^{q} \theta_q^2\right) \cdot \sigma_\varepsilon^2$ letting $\theta_0 = 1$

b) $X_{t-h} = \varepsilon_{t-h} + \theta_1 \varepsilon_{t-h-1} + \dots + \theta_q \varepsilon_{t-h-q}$

So $Cov(X_t, X_{t-h}) = Cov(\varepsilon_t + \dots + \theta_q \varepsilon_{t-q},\ \varepsilon_{t-h} + \dots + \theta_q \varepsilon_{t-h-q})$
We can break this up into a grid of cov ~~and~~ ~~we~~ terms. But all the off-

diagonals are 0 as ε is iid, and there are no "on-diagonal" terms since $h > q$ and so the first term ε_{t-h} in the expression for X_{t-h} comes "before" the last term ε_{t-q} for X_t. i.e., there's no overlap in the relevant shocks so we don't see any nonzero variance terms appear.

c). Let $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$.

Then $\mathrm{Cov}(X_t, X_{t-h}) = \mathrm{Cov}(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \varepsilon_{t-h} + \theta_1 \varepsilon_{t-h-1} + \theta_2 \varepsilon_{t-h-2})$

$\cancel{= \mathrm{Cov}(\varepsilon_t, \varepsilon_{t-h} + \dots) + \theta_1 \mathrm{cov}(\varepsilon_{t-1}, \dots) + \theta_2 \mathrm{cov}(\varepsilon_{t-2},}$

$\cancel{= \mathrm{cov}(\varepsilon_t, \varepsilon_{t-h})}$

Like above, we only care about the ~~terms~~ cases where we have the same shock terms as arguments to cov.

If $h = 0$, then $\mathrm{cov}(\cdot) = \mathrm{Var}(X_t) = (1 + \theta_1^2 + \theta_2^2) \cdot \sigma_\varepsilon^2$

If $h = 1$, $\mathrm{cov}(\cdot) = (\theta_1 + \theta_1 \theta_2) \cdot \sigma_\varepsilon^2$

If $h = 2$, $\mathrm{cov}(\cdot) = (\theta_2)\, \sigma_\varepsilon^2$

If $h > 2$ by (b) $\mathrm{cov}(\cdot) = 0$

Not sure how to argue for even-ness. I think in 1(b) I needed to also?

(*) And when we have $h < 0$, we can consider this as $\mathrm{Cov}(X_{t-s}, X_t)$ for $s = -h$, and since Cov is symmetric, the above is true for $h = -1, -2$, etc.

So the μ, σ^2 and ACF don't vary with time, meaning it's Stationary.

d). i. All the previous observations depend only on ε_0 and then shocks up to $t-h$. But ~~each~~ ε_t is $\perp\!\!\!\perp$ of ε_0 and every other ε_{t-s} for $s \geq 1$, as they're iid. So then $\varepsilon_t \perp\!\!\!\perp X_{t-h}$.

ii. $\varepsilon_t = -\theta_1 \varepsilon_{t-1} + X_t = -\theta_1(-\theta_1 \varepsilon_{t-2} + X_{t-1}) + X_t$, ~~etc~~

$= -\theta_1(-\theta_1(-\theta_1 \varepsilon_{t-3} + X_{t-2}) + X_{t-1}) + X_t$

$$= \sum_{i=0}^{t-1} X_t - \theta_1 X_{t-1} + \theta_1^2 X_{t-2} - \theta_1^3 \varepsilon_{t-3}$$

Suppose this is $\varepsilon_0 = 0$, then we stop at the θ_1^2 term.

$$= \sum_{i=0}^{t-1} (-\theta_1)^i X_{t-i}$$

[how to argue that more formally?] — this has all the info from prev. obs.

iii. $\mathbb{E}[X_{T+1} \mid X_T, X_{T-1}, \dots] = \mathbb{E}[\varepsilon_{t+1} + \theta_1 \varepsilon_t \mid \varepsilon_t + \theta_1 \varepsilon_{t-1}]$

$$= \mathbb{E}[\varepsilon_{t+1} \mid \dots] + \mathbb{E}[\theta_1 \varepsilon_t \mid \dots]$$

$$= 0 + \theta_1 \varepsilon_t$$ as each ε is iid, so conditioning on the past doesn't help us forecast the next shock.

$$= \theta_1 \cdot \sum_{i=0}^{T-1} (-\theta_1)^i X_{T-i}$$

3 i. a) Very low persistence.
b) Appears to be stationary, effectively white noise.
c) No transformations needed.

ii. a) Has much stronger persistence than you'd expect from a stationary series.
b) Variance seems lower now than in the past, and there are several trends visible (e.g. downward from 1980 - 2000). Might have stationarity from 1970-90 but unlikely given persistence.
c) Taking a differenced series would help remove the trend and likely help address the lack of stationarity.

iii. a) As ii.
b) Various trends e.g. downward from 1980-2000, upward since then (2000-2010). Within epochs e.g. 1985-95, might have stationarity.
c) Again, take a first difference and possibly split up into

seperate series to account for breaks.

iv. a) Moderate but rapidly decaying persistence.
b) Appears stationary, though maybe ↓ variance now than previously.
c) None needed, we already have a growth rate via logs.

4. $Y_t := 1200 \times \log(IP_t / IP_{t-1})$

a) No, it's not quite the same as a monthly % change.

$$Y_t = 1200 \times (\log IP_t - \log IP_{t-1})$$
$$= 1200 \times (\Delta \log IP_t)$$

A change in logs is equal to a growth rate, and here we've ~~turned~~ scaled that up by $12\times$ to annualise. So Y_t tells us the annualised growth rate of industrial production based on that month, in %.

(Also under this accounting a "100% ↑" and then "100% ↓" would take you back to the original level, not to 0?)

hmm, maybe not exactly because the denominator in the log changes from period t to $t+1$?

b) ~~Jan: $\hat{Y}_t$ = 0.79 + 0.05 × 101.4 + 0.19 × 101.0 + 0.23 × 100.4 + 0.16 × 100.2~~

=

	M9	M10	M11	M12	M13	M14
Y_t	7.20	2.39	7.14	4.74	–	–
$\hat{Y}_t$	–	–	–	–	4.09	3.92

c) Because there might be seasonal ~~patterns~~ noise in how growth occurs, so the observation from a year ago helps you predict this year's value better, beyond what you can do from the previous months alone.

Probably not worth including this in the model: the -ve coefficient is a priori surprising and magnitude is small relative to SE (90% CI crosses 0).

d) $BIC := \log \frac{SSR_m}{T} + m \frac{\log T}{T}$ where $m = p+1$ and $T = (2013-1986+1) \times 12 = 336$.

e.g. $BIC_0 = 4.08$ (rest in spreadsheet).

We want to find $\underset{p}{\text{argmin}}[BIC]$, which is when $p=4$, at ~~[illegible]~~ a 3.940

$AIC := \log \frac{SSR_m}{T} + m \frac{2}{T}$ and in this case agrees on $p=4$ (though in general AIC always chooses a weakly larger model))

for 3.983

So we should include 4 lags, as she originally did.

e) Our ADL model is $Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \gamma_1 \Delta R_{t-1} + \dots + u_t$

~~[illegible]~~ W Granger causes X iff lags of W help us improve our forecast of X, vs using its own lags alone.

We can perform an F-test to check for this (specifically, whether the MSFE is smaller when incorporating lags of R_t into our IP forecast), by comparing against a null that $H_0: \gamma_1 = \dots = \gamma_4 = 0$, vs $H_1: \exists l \ \gamma_l \neq 0$

$$F := \frac{SSR_{rs} - SSR_{un}}{SSR_{un}} \times \frac{\cancel{n-k-1}}{q}$$

$$= \frac{15842 - 13147}{\cancel{15842}\ 13147} \times \frac{336-8-1}{4} = 16.8$$

and here the limiting distribution is $F \sim F_{4,\infty}$ ~~so we have~~ ~~an p-value of~~ so we can reject H_0 at the 0.1% level ($c_\alpha = 4.616$)

f) Because Y_t and ΔR_t are plausibly stationary, but IP_t and R_t are not, out-of-sample forecasting is more credible under her model specification than using the raw data. Also, the F statistic has the usual limiting distribution under stationarity (but I think not otherwise? Though this all feels a bit alchemical, if that's a word...) hahaha!