Forecasting & Predictive Analytics

ESSEC-CentraleSupélec Master in Data Sciences & Business Analytics

Guillaume Chevillon

EXAM ANSWERS

December 11, 2018

1:15PM-4.15PM

Calculators are authorized

Please answer the three exercises below.

You can always assume one result proven and proceed to the next, even within questions.

Theory exercises

1. In this exercise, we explore the properties of various time series models.

   (a) Consider the process $x_t = \alpha + \beta t + w_t$ where (α, β) are constants and w_t is a white noise process.

   - Determine whether x_t is stationary
     answer: $E[x_t] = \alpha + \beta t$ is not constant so x_t is not stationary
   - Show that the process $y_t = \Delta x_t$ is stationary
     answer: $\Delta x_t = \beta + \Delta w_t$, i.e. it is a constant plus Δw_t, an MA(1) process which it is itself stationary, hence Δx_t is stationary.
   - Show that the expectation of $v_t = (x_{t+1} + x_t + x_{t-1})/3$ equals $\alpha + \beta t$ and give a simple description of the autocovariance and partial autocovariance functions of v_t (no need to compute them).
     answer: at each t, $E[x_t] = \alpha + \beta t$ so

$$E[v_t] = \frac{\alpha + \beta(t+1) + \alpha + \beta t + \alpha + \beta(t-1)}{3} = \frac{3}{3}(\alpha + \beta t) = \alpha + \beta t$$

     and

$$v_t = \alpha + \beta t + \frac{1}{3}w_{t+1} + \frac{1}{3}w_t + \frac{1}{3}w_{t-1}$$

     i.e. v_t is a sum of a linear deterministic trend and an MA(2). Since the constant/trend do not affect the ACF/PACF, the latter are those of an MA(2), i.e. the first two ACF terms are nonzero, the following are zero and the PACF decays exponentially towards zero. Notice that

$$\begin{aligned}
var(v_t) &= \frac{3}{9}\sigma_w^2 = \frac{1}{3}\sigma_w^2 \\
corr(v_t, v_{t-1}) &= \frac{cov\left(\frac{1}{3}w_t + \frac{1}{3}w_{t-1}, \frac{1}{3}w_{t+1} + \frac{1}{3}w_t\right)}{var(v_t)} = \frac{2}{3} \\
corr(v_t, v_{t-2}) &= \frac{1}{3}
\end{aligned}$$

   - We consider the following two forecasting models that do not require knowledge of the model parameters:

$$\begin{aligned}
\widehat{x}_{t+1|t} &= v_{t-1} + \Delta x_t, \\
\widetilde{x}_{t+1|t} &= x_t + \Delta x_t.
\end{aligned}$$

     Are these forecasts biased? Which one yields the lower (unconditional) Mean-Square Forecast Error?
     answer:

$$\begin{aligned}
\widehat{x}_{t+1|t} &= \alpha + \beta(t-1) + \frac{1}{3}w_t + \frac{1}{3}w_{t-1} + \frac{1}{3}w_{t-2} + \beta + \Delta w_t \\
&= \alpha + \beta t + \frac{4}{3}w_t - \frac{2}{3}w_{t-1} + \frac{1}{3}w_{t-2} \\
\widetilde{x}_{t+1|t} &= \alpha + \beta(t+1) + 2w_t - w_{t-1}
\end{aligned}$$

The first forecast is biased since

$$\begin{aligned}
E\left[x_{t+1}-\widehat{x}_{t+1|t}\right] &= \beta, \\
var\left[x_{t+1}-\widehat{x}_{t+1|t}\right] &= var\left[w_{t+1}-\frac{4}{3}w_t+\frac{2}{3}w_{t-1}-\frac{1}{3}w_{t-2}\right]=\frac{10}{3}\sigma_w^2 \\
E\left[x_{t+1}-\widetilde{x}_{t+1|t}\right] &= 0, \\
var\left[x_{t+1}-\widetilde{x}_{t+1|t}\right] &= var\left[w_{t+1}-2w_t+w_{t-1}\right]=6\sigma_w^2
\end{aligned}$$

so the first yields a lower MSFE if $\beta^2<\frac{8}{3}\sigma_w^2$. Notice that one can easily get an unbiased forecast with lower MSFE than $\widehat{x}_{t+1|t}$ by setting $\widehat{x}^*_{t+1|t}=v_{t-1}+2\Delta x_t$.

(b) Consider the ARIMA model

$$x_t=\varepsilon_t+\theta\varepsilon_{t-2}, \tag{1}$$

where ε_t denotes a white noise process.

- What type of model is this? Explain simply how you would estimate it?
  answer: This is an MA(2) which can be estimated simply by Maximum Likelihood if we are willing to assume a distribution for ε_t, in which case the log likelihood function is

$$\begin{aligned}
\ell(\theta)=\log L(\theta) &= \log f_{\varepsilon_0}(\varepsilon_0)+\sum_{t=1}^{T}\log f_{\varepsilon}(\varepsilon_t|x_{t-1},...) \\
&= \log f_{\varepsilon_0}(\varepsilon_0)+\sum_{t=1}^{T}\log f_{\varepsilon}(x_t-\theta\varepsilon_{t-2}|x_{t-1},...)
\end{aligned}$$

  where f_ε is the density of ε_t and ε_{t-2} is computed as $x_{t-2}-\theta\varepsilon_{t-4}=x_{t-2}-\theta\left[x_{t-4}-\theta\left[x_{t-6}-\theta\left[...\right]\right]\right]$ and so on. Notice that we can also used Nonlinear Least Squares where the ε_t are computed using the latter non-linear expression.
- Show that the series x_t is invertible for $|\theta|<1$ and find the coefficients in the representation

$$\varepsilon_t=\sum_{k=0}^{\infty}\pi_k x_{t-k}.$$

  answer: when $|\theta|<1$ the expansion $\left(1+\theta L^2\right)^{-1}=1-\theta L^2+\theta^2L^4-\theta^3L^6...$ is defined so we can invert

$$\begin{aligned}
x_t &= \left(1+\theta L^2\right)\varepsilon_t \\
\left(1+\theta L^2\right)^{-1}x_t &= \varepsilon_t
\end{aligned}$$

  i.e.

$$\varepsilon_t=\left(1-\theta L^2+\theta^2L^4-\theta^3L^6\right)x_t=\sum_{k=0}^{\infty}(-\theta)^k x_{t-2k}$$

  for all $k\geq 0$

$$\pi_{2k}=(-\theta)^k, \quad \pi_{2k+1}=0.$$

- What is the optimal forecast (using the Mean-Square Forecast Error criterion) for x_{t+1} given its infinite past $x_t, x_{t-1}, ...$?
  answer: The optimal forecast is the conditional mean $E_t(x_{t+1}) = E_t(\varepsilon_{t+1} + \theta\varepsilon_{t-1}) = \theta\varepsilon_{t-1}$, i.e.

$$E_t(x_{t+1}) = \theta \sum_{k=0}^{\infty} \pi_k x_{t-k-1} = \theta \sum_{k=0}^{\infty} (-\theta)^k x_{t-1-2k}$$

- Same question for the optimal forecasts of x_{t+2} and x_{t+3} given $x_t, x_{t-1}, ...$ This should be reasonably simple keeping equation (1) in mind.
  answer: Now $E_t(x_{t+2}) = \theta\varepsilon_t = \sum_{k=0}^{\infty} (-\theta)^{k+1} x_{t-2k}$ and $E_t(x_{t+3}) = 0$.

(c) Describe and sketch the ACF and PACF of the seasonal ARIMA$(0,1)\times(1,0)_{12}$ model with autoregressive coefficient $\phi = 0.8$ and moving average coefficient $\theta = 0.5$. Please justify the shape but do not compute the values.
answer: Here we have a combination of an MA(1) and a seasonal AR(1) at lag 12, i.e. in fact an ARMA(12, 1) with all first eleven AR coefficients set to zero. A priori the MA(1) implies an ACF which is nonzero at lag 1 and a PACF that decays progressively, the AR(12) will induce an ACF that decays and a PACFthat is nonzero at lag 12.

$$\left(1 - \phi L^{12}\right) x_t = (1 + \theta L)\,\epsilon_t$$

The combination of both will also imply that the ACF will be nonzero at lags 13 (due to the interaction with the MA(1)) and also 24 and 25, and so on, see Figure 1.

2. In an empirical study of multistep forecasting models, McCracken & McGillicuddy (2018) explore the simple VAR(1) bivariate model:

$$\begin{aligned}\mathbf{z}_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} 0 & b \\ 0 & c \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}, \\ &= \mathbf{A}\mathbf{z}_{t-1} + \epsilon_t,\end{aligned} \tag{2}$$

where

$$\epsilon_t = \begin{bmatrix} u_t \\ v_t \end{bmatrix} \overset{i.i.d.}{\sim} \mathsf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

Throughout, we assume that forecast errors are evaluated using the Mean-Square Forecast Error (MSFE) as a loss function.

(a) Although $\mathbf{z}_t$ is bivariate, show that x_t follows a simple *univariate* process. What is its name? Under what condition is x_t stationary?
answer: We see that the process for x_t is

$$x_t = c x_{t-1} + v_t, \tag{3}$$

i.e., an AR(1) which is stationary if $|c| < 1$.

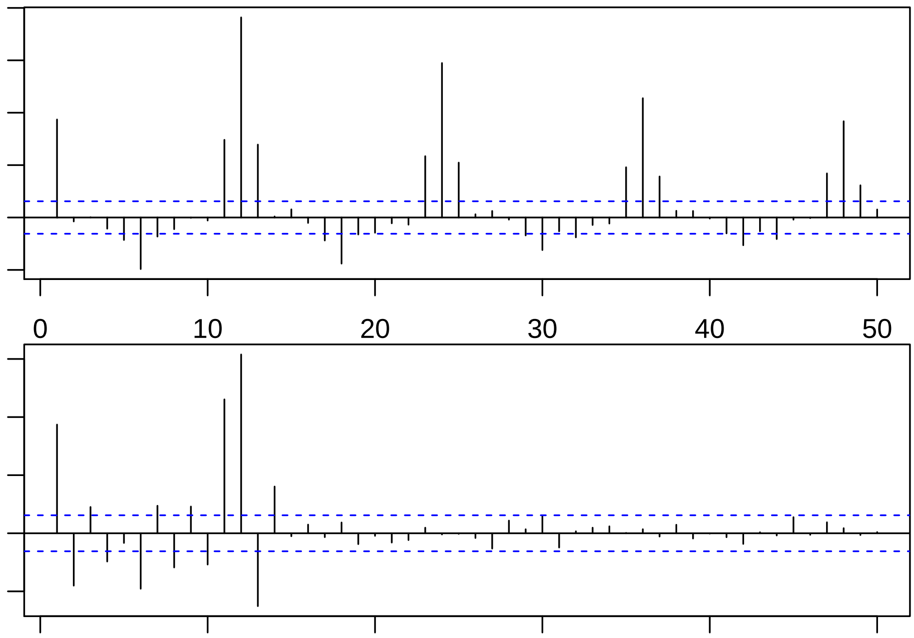

Figure 1: Simulated ACF and PACF of an ARMA$(0,1) \times (1,0)_{12}$.

(b) When x_t is not stationary, do x_t and y_t cointegrate? Please argue simply and find the cointegration relation.
answer: When x_t is not stationary, x_t and y_t cointegrate if x_t is integrated, i.e. $c = 1$. Then the first equation in the VAR shows that

$$y_t = bx_{t-1} + u_t$$

i.e. y_t is not stationary if $b \neq 0$ and x_t is not stationary and we can simply express $x_{t-1} = x_t - v_t$ from equation (3) with $c = 1$ so the cointegration relation is

$$y_t - bx_t = u_t - v_t$$

since $u_t - v_t$ is stationary (even $i.i.d$):

$$y_t - bx_t \sim \mathsf{I}(0)\,.$$

(c) Using equation (2), observe that, if $b \neq 0$, $x_{t-1} = b^{-1}y_t - b^{-1}u_t$. Since x_{t-1} can itself be expressed as a function of x_{t-2} (when $c \neq 0$) and v_{t-1}, solve the expression to obtain y_t as a function of y_{t-1}, u_t, u_{t-1} and v_{t-1}.
answer: The AR(1) together with $b^{-1}y_t - b^{-1}u_t$ imply:

$$\begin{aligned} x_{t-1} &= cx_{t-2} + v_{t-1} \\ b^{-1}y_t - b^{-1}u_t &= c\left(b^{-1}y_{t-1} - b^{-1}u_{t-1}\right) + v_{t-1} \end{aligned}$$

i.e.,

$$y_t = cy_{t-1} + u_t - cu_{t-1} + bv_{t-1}.$$

(d) Define $w_t = y_t - cy_{t-1}$. Using your answer to question (c), show that w_t is covariance stationary. Compute the autocovariance function of w_t and show that this implies that w_t follows an MA(1). We therefore write $w_t = \eta_t + \theta\eta_{t-1}$. Show (simply) that y_t follows an ARMA(1, 1).
answer: Now $w_t = u_t - cu_{t-1} + bv_{t-1}$ is the sum of three *iid* processes so it is stationary. It therefore admits a Wold decomposition into an infinite moving average representation

$$w_t = \eta_t + \sum_{k=1}^{\infty} \psi_k \eta_{t-k}$$

but since $cov\,(w_t, w_{t-1}) = -c\sigma_u^2 = -c$ and $cov\,(w_t, w_{t-k}) = 0$ for $k > 1$, we must have $\psi_k = 0$ for $k > 1$, i.e. letting $\psi_1 = \theta$ we see that

$$w_t = \eta_t + \theta\eta_{t-1}$$

and by definition

$$y_t = cy_{t-1} + \eta_t + \theta\eta_{t-1} \sim ARMA\,(1, 1)\,.$$

To express θ and σ_η^2 as a function of ρ, b and c, we simply need to equate $\mathsf{Var}\,(w_t)$ and $\mathsf{Cov}\,(w_t, w_{t-1})$ using the two decompositions and then $\eta_t = (1 + \theta L)^{-1} w_t = (1 + \theta L)^{-1} (u_t - cu_{t-1} + bv_{t-1})\,.$

(e) Under MSFE loss, how do we define the optimal forecast of $\mathbf{z}_{t+1}$ conditional on information $\mathcal{I}_t = \{\mathbf{z}_t, \mathbf{z}_{t-1}, ...\}$ observed at time t? We denote this forecast by $\mathbf{z}_{t+1|t}$. Find an expression for the value of $\mathbf{z}_{t+1|t}$ in terms of the parameters of the VAR(1) model. Find also an expression for the optimal forecast $\mathbf{z}_{t+2|t}$. What is the value of the MSFE associated with the forecast for y_{t+1}, $\mathsf{E}\left[\left(y_{t+1} - y_{t+1|t}\right)^2\right]$, where $y_{t+1|t}$ denotes the first entry of the vector $\mathbf{z}_{t+1|t}$.
answer: Under MSFE loss, the optimal forecast is the conditional expectation

$$\mathbf{z}_{t+1|t} = \begin{bmatrix} y_{t+1|t} \\ x_{t+1|t} \end{bmatrix} = \mathsf{E}\left[\mathbf{z}_{t+1}|\mathcal{I}_t\right] = \mathbf{A}\mathbf{z}_t = \begin{bmatrix} 0 & b \\ 0 & c \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix}$$

and $\mathbf{z}_{t+2|t} = \mathbf{A}^2\mathbf{z}_t$.
The forecast error associated with y_{t+1} is $y_{t+1} - y_{t+1|t} = y_t - bx_{t-1} = u_t$ with MSFE

$$\mathsf{E}\left[\left(y_{t+1} - y_{t+1|t}\right)^2\right] = \mathsf{Var}\,[u_t] = 1.$$

(f) Imagine now that at time t, you want to forecast y_{t+1} but that you actually know the future value x_{t+1}. This is called a conditional forecast. It can be shown then that the optimal conditional forecast is given by

$$\widehat{y}^*_{t+1|t} = bx_t + \rho\,(x_{t+1} - cx_t)\,, \tag{4}$$

where the value of the MSFE relative to $\widehat{y}^*_{t+1|t}$, $\mathsf{E}\left[\left(y_{t+1} - \widehat{y}^*_{t+1|t}\right)^2\right] = 1 - \rho^2$. Does this imply that the conditional forecast of y_{t+1} is more accurate than the

unconditional forecast obtained in question (e)?
answer: We see here that if $\rho \neq 0$ then there is contemporaneous correlation between u_{t+1} and v_{t+1}, the (unconditional) forecast errors $\left(y_{t+1} - y_{t+1|t}\right)$ and $\left(x_{t+1} - x_{t+1|t}\right)$. We see that in this case, having information about x_{t+1} helps forecasting since $\mathsf{E}\left[\left(y_{t+1} - \widehat{y}^{*}_{t+1|t}\right)^2\right] < \mathsf{E}\left[\left(y_{t+1} - y_{t+1|t}\right)^2\right]$. When $\rho = 0$, the two are identical, we cannot do better here than the unconditional forecast (short of actually knowing something that correlates with u_{t+1}).

(g) We now try a different *conditional* forecast, using the regression $y_t = \gamma x_t + \varepsilon_t$, where $\gamma = \mathsf{Cov}\left(y_t, x_t\right)/\mathsf{Var}\left(x_t\right) = bc + \rho\left(1 - c^2\right)$. This new conditional forecast is obtained as

$$\widetilde{y}_{t+1|t} = \gamma x_{t+1}, \tag{5}$$

The associated MSFE is $\mathsf{E}\left[\left(y_{t+1} - \widetilde{y}_{t+1|t}\right)^2\right] = 1 - \rho^2 + (b - c\rho)^2$. What conclusion do you reach about the relative accuracy of the two conditional forecasts?
answer: Here the new forecast is not the optimal conditional forecast and we see that if $(b - c\rho)^2 > 0$ then $\mathsf{E}\left[\left(y_{t+1} - \widetilde{y}_{t+1|t}\right)^2\right] > \mathsf{E}\left[\left(y_{t+1} - \widetilde{y}_{t+1|t}\right)^2\right]$ so the new forecast is less accurate. Notice that if $b = c\rho$ so $\gamma = \rho$ then both forecasts are as accurate since they coincide: $\widehat{y}^{*}_{t+1|t} = \widetilde{y}_{t+1|t} = \rho x_{t+1}$.

(h) Assume the processes undergo a structural break that is unknown to the modeler: b shifts between dates T and $T+1$ from the value b_0 to b_1. All other parameters remain constant. The same conditional forecasts then yield at time T

$$\mathsf{E}\left[\left(y_{T+1} - \widehat{y}^{*}_{T+1|T}\right)^2\right] = 1 - \rho^2 + \frac{(b_0 - b_1)^2}{1 - c^2},$$

$$\mathsf{E}\left[\left(y_{T+1} - \widetilde{y}_{T+1|T}\right)^2\right] = \mathsf{E}\left[\left(y_{T+1} - \widehat{y}^{*}_{T+1|T}\right)^2\right] + (b_1 - c\rho)^2 - (b_1 - b_0)^2.$$

What can you say about the relative accuracy of the two conditional forecasts when $b_1 = c\rho$? Does this contradict the notion of "optimal" forecast?
answer: When $b_1 = c\rho$ then we see that if $b_1 \neq b_0$ then

$$\mathsf{E}\left[\left(y_{T+1} - \widetilde{y}_{T+1|T}\right)^2\right] < \mathsf{E}\left[\left(y_{T+1} - \widehat{y}^{*}_{T+1|T}\right)^2\right]$$

i.e. the supposedly non-optimal conditional forecast performs better than the "optimal" one. It's because the "optimal" forecast is constructed based on the conditional expectation given information at time T but the data generating process shifts so this is no longer the true mathematical expectation. It so happens that when $b_1 = c\rho$, the conditional forecast $\widetilde{y}_{T+1|T}$ may be closer to the true optimal forecast based on the post-break DGP. Notice that even if $b_1 \neq c\rho$, there is a range of parameter values for which $\widetilde{y}_{T+1|T}$ is more accurate.

EMPIRICAL ANALYSIS

3. In this exercise, you are asked to reflect on an empirical analysis of the link between global temperature anomalies (i.e. difference from a simple seasonal model, available from the international GIEC consortium as Giss v3) and human emissions of greenhouse gases (so call Anthropogenic Radiative Forcing). The temperature series (variable $Temp$) and the emissions (variable ARF) are presented in Figure 1, together with their first (prefix D) and second (prefix DD) differences (Δ and Δ^2).

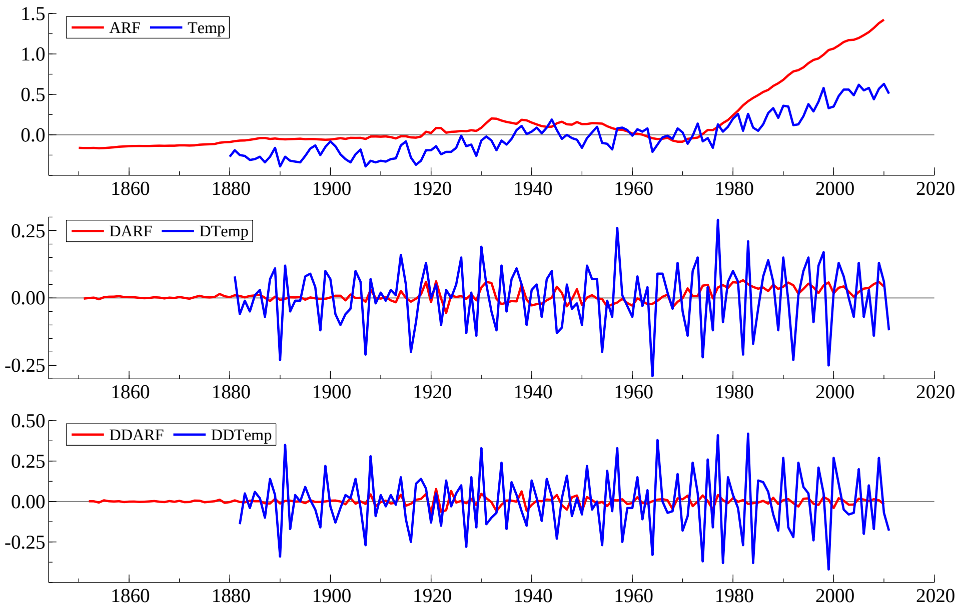

Figure 2: Time Series plots of $Temp$ and ARF and their differences

(a) Figures 2, 3 and 4 present the time series, ACF and PACF for the two variables and their differences. Using these three figures, what is, according to you, the degree of integration of $Temp$ and ARF? Propose (and justify) ARIMA(p, d, q) models for $Temp$ and ARF. Notice that $DARF$ presents a peculiar ACF that you may want to describe and attempt to understand.

answer: The variables $Temp$ and ARF are clearly non-stationary since they present an overall upward trend. The first differences seem "closer to" being stationary although the mean of $DARF$ seems to increase towards the end of the sample. So there is a possibility that $DARF$ might not be stationary. By contrast, the distribution of $DTemp$ seems stationary throughout.

Now, regarding the AC/PACF. For $Temp$, the ACF decays linearly as is typical for nonstationary processes. The ACF of $DTemp$ decays, yet with a non zero fourth term, the first three PACF terms are nonzero. This could indicate an AR(3) or an MA(4), or probably a mix of both. So I would start estimating an ARMA$(3, 4)$ for $DTemp$, i.e. an ARIMA$(3, 1, 4)$ for $Temp$.

As for $DARF$, it's not clear whether we need to consider it $\mathsf{I}(2)$ or $\mathsf{I}(1)$. If it is integrated of order 1, then we see that the ACF of $DARF$ decays very slowly and remains significant until long lags (around 15) and also the PACF is significant for the first seven lags (plus lag 14). So it could be a long AR(7) with possible additional MA structure or a very long MA(14). We cannot preclude the possibility of fractional integration either, but the sample size is small to detect it. Unfortunately, this means that we need to start estimating an ARIMA$(7, 1, q)$ for ARF with a q that we need to try out ($q = 14$ is unreasonable and won't be estimable). Alternatively, if we want to consider $DDARF$, then an MA(4) seems to work well, i.e. an ARIMA0, 2, 4 for ARF.

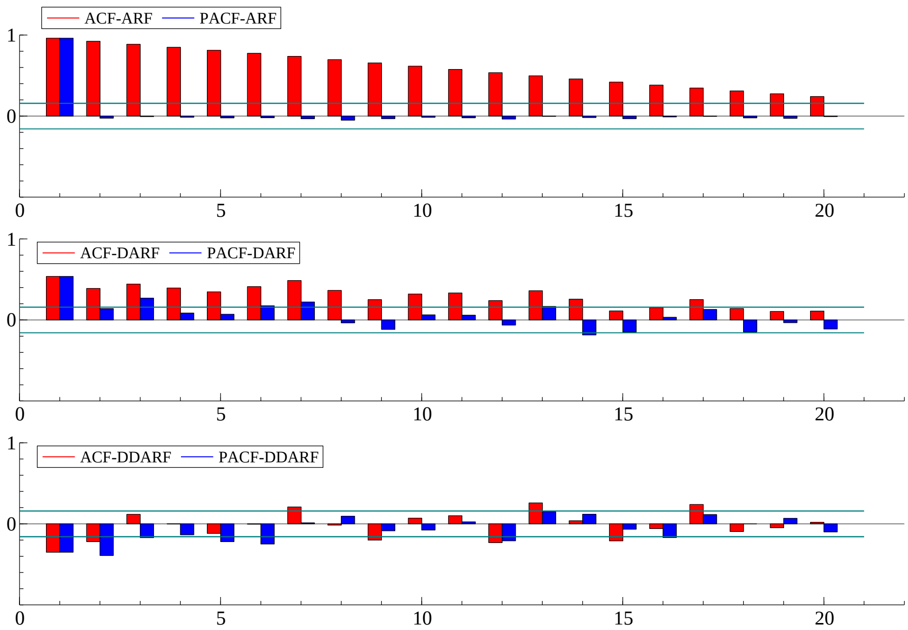

Figure 3: Autocorrelogram and Partial Autocorrelogram of ARF and its differences

(b) Table 1 presents the output from a series of Augmented Dickey-Fuller (ADF) tests for ARF and $Temp$. Each row represents an ADF test based on a different $AR(p)$ model, for p ranging from 1 to 12 (Column D-lag). For either ARF and $Temp$, three columns report the following: *t-ADF* is the test statistic, $\beta\ y_{t-1}$ is the inverse of the root of the lag polynomial with smallest modulus (i.e. the closest to unity), and AIC is the Aikaike Information Criterion.

- Explain the Augmented Dickey-Fuller test and conclude here about the null hypothesis (the critical values are adjusted for the sample size).

  answer: The Augmented Dicky-Fuller test (ADF) is a test for the null hypothesis that the process x_t presents a unit-root. The test statistic is the t statistic for $\phi = 0$ (against the alternative that $\phi < 0$) in the regression

$$\Delta x_t = \phi x_{t-1} + \sum_{j=1}^{p-1} \psi_j \Delta x_{t-j} + \varepsilon_t,$$

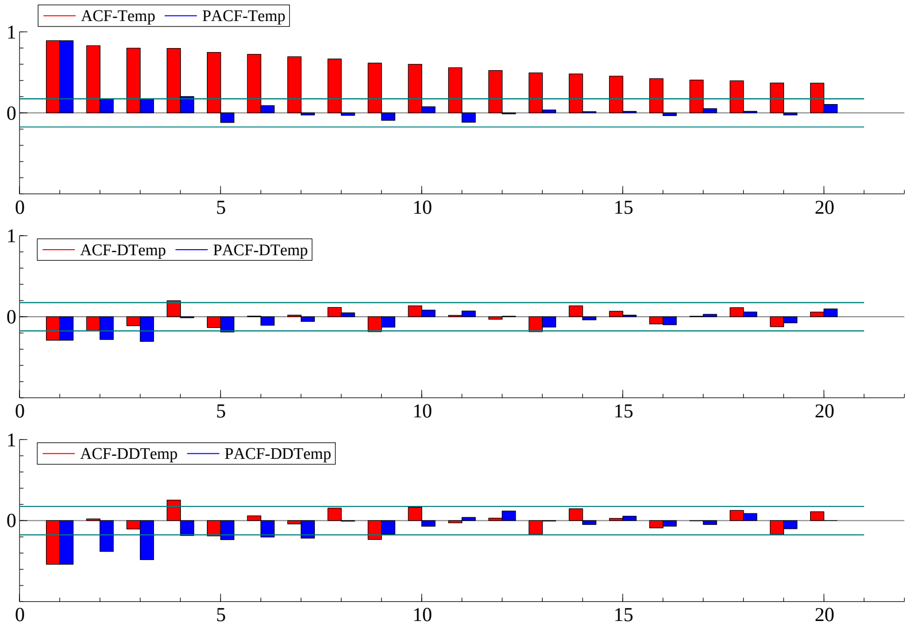

Figure 4: Autocorrelogram and Partial Autocorrelogram of $Temp$ and its differences

i.e. assuming x_t follows an AR(p) with a unit root. We can add a constant and a linear trend in the regression as well. The t-test for $\phi = 0$ is then assessed using a Dickey-Fuller distribution with corresponding deterministic components. The choice of p is made by estimating various AR(p) and choosing p that minimizes the Akaike Information Criterion (AIC).
In this example we first need to find the values of p for ARF and $Temp$, they are respectively 7 and 3. Then the t-statistics are -0.86 and -2.25 which are both greater than the 5% critical value of -3.45 (or -3.44 depending on the sample size adjustment). Hence we do not reject the null for either variable and conclude that they are nonstationary.

- The test is carried out with a constant and a trend included in the model, explain what this means.
  answer: As mentioned above, we have to choose the deterministic components included in the regression (none, constant, constant+trend). Unfortunately, these do not play the same role under the null and alternative hypotheses (e.g. a constant under the null implies an linear trend in the process while it implies a nonzero mean under the alternative). The general principle is that if we suspect the presence of a given deterministic element, we need to make sure it is present both under the null and the alternative. Here, given the upward trend, if there is no unit root, we need to make sure a trend is estimated under the alternative (so its presence can then be tested).

(c) Let's think further about the peculiar shape of the ACF for $DARF$. When we

Table 1: The sample is: 1863 - 2010 for ARF (161 observations) and 1893-2011 for $Temp$ (132 observations)
ARF: ADF tests (T=148, Constant+Trend; 5%=-3.44 1%=-4.02)
$Temp$: ADF tests (T=119, Constant+Trend; 5%=-3.45 1%=-4.04)

	ARF			Temp		
D-lag	t-ADF	$\beta\ y_{t-1}$	AIC	t-ADF	$\beta\ y_{t-1}$	AIC
12	-0.5576	0.99385	-7.802	-1.474	0.8204	-4.542
11	-0.755	0.99212	-7.814	-1.434	0.83089	-4.557
10	-0.5118	0.99499	-7.822	-1.249	0.85734	-4.566
9	-0.3079	0.99714	-7.832	-1.054	0.88392	-4.576
8	-0.7634	0.99324	-7.833	-1.447	0.84636	-4.578
7	-0.8593	0.9928	-7.846	-1.374	0.85934	-4.593
6	-0.05774	0.99953	-7.804	-1.516	0.84997	-4.608
5	0.4396	1.0035	-7.791	-1.79	0.82883	-4.619
4	0.5921	1.0046	-7.802	-2.325	0.78372	-4.608
3	0.7834	1.0059	-7.811	-2.253	0.79789	-4.621
2	1.459	1.011	-7.773	-3.174	0.72032	-4.569
1	1.693	1.0123	-7.783	-4.174	0.65786	-4.557
0	3.36	1.0246	-7.647	-5.188	0.62038	-4.565

look at the time series plot of Figure 2, we notice that ARF seems to experience a break in trend around 1970. Correspondingly, $DARF$ experiences a shift in its mean at the same period. To assess this, we compute the recursive mean of $DARF$, i.e. $\mu_t = \frac{1}{t}\sum_{j=1}^{t} DARF_j$, the mean up until instant t (here we consider $t = 1$ corresponds to the beginning of the sample), see Figure 5 that reports the recursive mean μ_t together with a 95% confidence interval.

- Explain why a shift in the deterministic trend of ARF corresponds to a shift in the mean of $DARF$. Is there evidence of such a shift here?
  answer: Assume a process presents a linear trend that presents a break at time $t_0 > 0$, i.e. such that

$$y_t = \alpha + \beta t + (\alpha_0 + \beta_0 t) \times 1_{\{t \geq t_0\}} + x_t$$

  where x_t does not have deterministic elements (constant or trend). Then

$$\begin{aligned}
\Delta y_t &= \beta + (\alpha_0 + \beta_0 t)\left(1_{\{t\geq t_0\}} - (\alpha_0 + \beta_0 (t-1))\, 1_{\{t-1\geq t_0\}}\right) + \Delta x_t \\
&= \beta + (\alpha_0 + \beta_0 t)\left(1_{\{t\geq t_0\}} - 1_{\{t\geq t_0+1\}}\right) + \beta_0 1_{\{t-1\geq t_0\}} + \Delta x_t \\
&= \beta + (\alpha_0 + \beta_0 t_0)\, 1_{\{t=t_0\}} + \beta_0 1_{\{t\geq t_0+1\}} + \Delta x_t
\end{aligned}$$

  so we see that the mean is β prior to the break and that it shifts to $\beta + \beta_0$ at $t_0 + 1$. There is also a "blip" or outlier $\alpha_0 + \beta_0 t_0$ at time t_0.
  We observe such a shift around 1970. We do see oscillations in μ_t beforehand, but the mean is overall relatively stable.

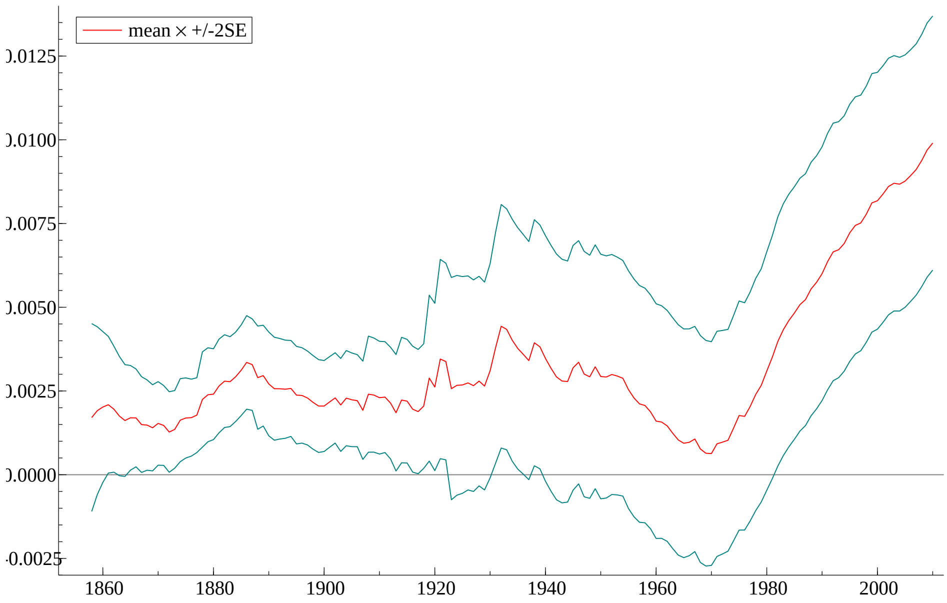

Figure 5: Recursive mean of $DARF$ together with a two-sided 95% confidence interval

- We now create a variable $dumm1970$ that takes value 0 before 1969 and 1 after 1970. We regress $DARF$ on a constant and $dumm1970$ and obtain the following result (t-statistics are in parentheses below the parameters):

$$DARF_t = \underset{(0.387)}{6.4 \times 10^{-4}} + \underset{(11.0)}{0.036} \times dumm1970_t + DARF\_dum \qquad (6)$$

where $DARF\_dum$ is defined as the residual from the regression. Figure 6 plots $DARF$ and $DARF\_dum$, together with the ACF/PACF of $DARF\_dum$.

i. Explain, in Equation (6), whether there is evidence of a break in mean for $DARF$.

answer: Here we see that the coefficient in front of the dummy variable is significantly different from zero at almost any significance level. There is very strong evident here of a mean shift in $DARF$.

ii. Compare and describe the difference between the ACF/PACF of $DARF$ and $DARDF\_dum$. Can you come up with an explanation for why moving from $DARF$ *to* $DARF\_dum$ modifies the ACF/PACF.

answer: Figure 6 shows that there is a mean shift towards the end of the sample. When taking this shift into account and correcting for it, we notice that the ACF/PACF no longer exhibits slowly decaying ACF at long lags. Instead, we see that most values are non-signficant, except a form of seasonality at 6 or 7 lags. This may actually correspond to business cycle frequencies (recessions tend to occur on average every six years, and they reduce economic activity, hence global emissions).

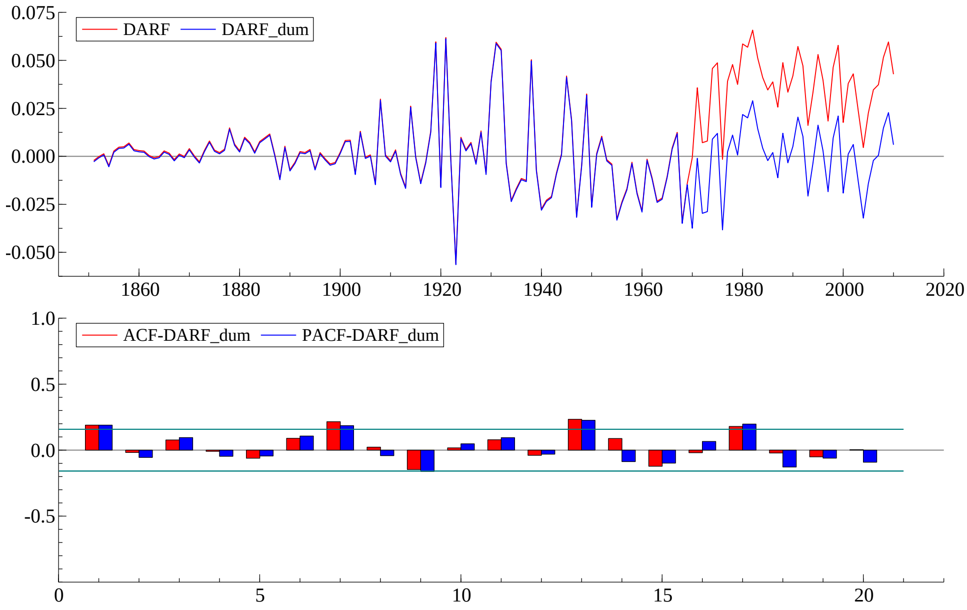

Figure 6: Estimation of a shift in the mean of $DARF$ and ACF/PACF of $DARF$ once corrected for the estimated shift.

Now, why does the shift in mean generate slowly decaying autocorrelations, it sufficiens to look at figure 5: the recursive mean looks like a stochastic trend, just attenuated. If it is not taken into account, it will appear (not directly though) in the computation of the ACF/PACF.

(d) We now regress $Temp$ on ARF to see whether there is a link between the two, as in the following equation (standard-errors are in parentheses below the parameters)

$$Temp_t = -\underset{(0.013)}{0.14} + \underset{(0.027)}{0.56}\ ARF_t + residual_t.$$

The temperature anomalies, the fitted values together with the residuals are reported in Figure 7.

i. Is there evidence that the recent increase in greenhouse gas emissions of human origin has impacted global temperatures? Justify very carefully, using the regression output and visual inspection of the corresponding figure.
answer: In the figure, we clearly see the fit of the variables, ARF seems to explain well $Temp$. This seems confirmed by the regression. Yet, we know that we need to be careful since both variables are nonstationary, hence there is a risk of spurious correlation/fit. For the relation to be meaningful in practice, the residuals need to be stationary. Visual inspection of Figure 7 is not fully conclusive: the residuals oscillate around their mean, but they are also very persistent, so a proper test is needed.

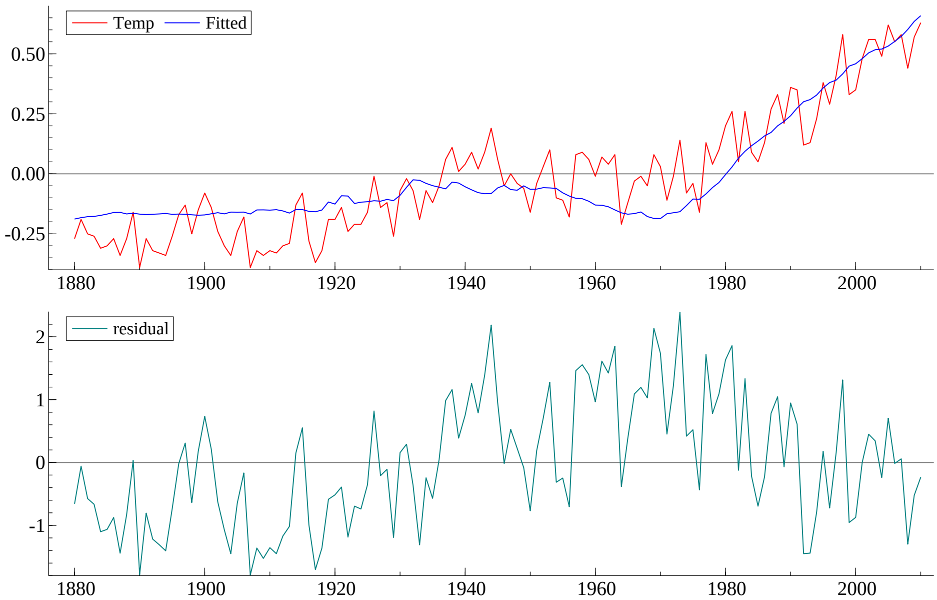

Figure 7: Estimation of the contemporaneous link between temperature and human greenhouse gas emissions.

ii. In other words, is there evidence of cointegration between the two variables? Argue simply, explaning the concept of cointegration. What extra information do you need in order to conclude?
answer: Well, we have concluded before that the two variables are nonstationary, possibly integrated of order 1 (if you correct for the break in trend in ARF). If that is true, and if the residuals discussed above are stationary, then $Temp_t - 0.56 \times ARF_t$ defines a stationary linear combination of the nonstationary variables: this is the definition of cointegration, the variables share a common source of non stationarity.

iii. Can you relate this to the shift/break discussed in question *c.* (above) regarding the sources of non-stationarity in this system? (It is not straightforward so argue simply).
answer: Several remarks can be made.

A. The break we had seen in the trend in ARF, so that $DARF$ appears non-stationary whereas $DTemp$ is stationary seem to imply that ARF and $Temp$ are not of the same order of integration and so cannot relate directly, unless $Temp$ also shows a break in trend, which Figure 7 shows to be plausible.

B. In fact it does indeed seem that the upward trend starts around 1970 for both ARF and $Temp$, so this source of nonstationarity appears common to both.

C. The trend in $Temp$ seems to increase earlier than that in ARF though on Figure 7, so further analysis is warranted.

(e) We now use a VAR(4) model in order to model the impact of a positive increase in ARF on future temperatures (and greenhouse gas emissions). We intend to do an Impulse Response Function (IRF) analysis but are worried about the issue of exogenous shocks. The VAR(4) is

$$\begin{bmatrix} ARF_t \\ Temp_t \end{bmatrix} = A_0 + A_1 \begin{bmatrix} ARF_{t-1} \\ Temp_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} ARF_{t-2} \\ Temp_{t-2} \end{bmatrix}$$
$$+ A_3 \begin{bmatrix} ARF_{t-3} \\ Temp_{t-3} \end{bmatrix} + A_4 \begin{bmatrix} ARF_{t-4} \\ Temp_{t-4} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix},$$

where

$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \overset{i.i.d}{\sim} \mathsf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Omega \right).$$

i. Explain the notion of IRF and why we need an extra assumption.
answer: The IRF is a way to see the impact of a change in a variable on future values of all variables. A way to consider a change in $(ARF_t, Temp_t)$ without changing the past is through a change in the errors $(\epsilon_{1t}, \epsilon_{2t})$. The issue is that if the two errors are correlated, we cannot consider the change in one of them only (and hence in one element of $(ARF_t, Temp_t)$ only) since they both co-move.
A solution consists in orthogonalizing $(\epsilon_{1t}, \epsilon_{2t})$ using any matrix Σ such that $\Sigma\Sigma' = \Omega$:

$$\mathsf{Var}\left(\Sigma^{-1} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \right) = \Sigma^{-1}\Omega\left(\Sigma'\right)^{-1} = \Sigma^{-1}\Sigma\Sigma'\left(\Sigma'\right)^{-1} = \mathbf{I}_2$$

i.e. such that, letting

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \Sigma^{-1} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \overset{i.i.d}{\sim} \mathsf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

(e_{1t}, e_{2t}) are orthogonal (even independent). We can therefore consider a change in either e_{1t} or e_{2t} alone.
The issue is that Σ is not unique, since for any orthonormal matrix O such that $OO' = \mathbf{I}_2$ then

$$\Sigma^{-1}O \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \overset{i.i.d}{\sim} \mathsf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

The question is to identify one Σ.

ii. We obtained the IRF presented in Figure 8 for shocks that take the value of one times the standard deviation. Comment on the implied long run impact of an sudden increase in ARF.
answer: The long run impact of a change in ARF on $Temp$ appears in

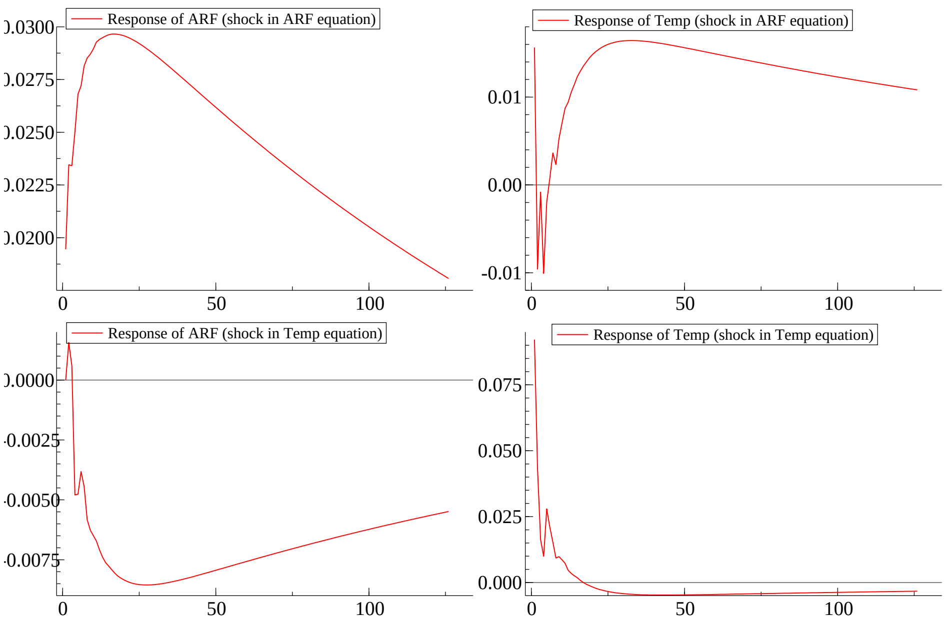

Figure 8: Impulse response functions to shocks in the ARF and $Temp$ equations.

the top-right hand side corner where we see that when there is a sudden increase in ARF, oscillations follow and then stabilize at a higher level of $Temp$ which then decays extremely slowly. The long run impact of a change in ARF onto itself seems to decay much faster. Yet, we observe a build up at first that leads to 150% the initial shock and the level of the initial shock is only reached again after 100 years, so the correction is still very slow.

iii. Whereas we believe that a shock in ARF has an immediate impact on $Temp$, we believe that a shock in $Temp$ does not have an immediate impact on ARF, but only a delayed one. We use a Choleski Decomposition of Ω into a product of triangular matrices but hesitate between the ordering of the variables. Denote (e_{1t}, e_{2t}) the structural shocks such that we have the choice between

$$(a) \quad \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} = \begin{bmatrix} s_1 & 0 \\ s_3 & s_2 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}, \quad \text{and} \quad (b) \quad \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} = \begin{bmatrix} c_1 & c_3 \\ 0 & c_2 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

which of the two orderings (a) or (b) describes best the narrative above.

answer: Here the question is whether there is the possibility of a shock only on ϵ_{1t} or ϵ_{2t} (i.e. ARF or $Trends$). In (a) we see that e_{1t} affects both ϵ_{1t} and ϵ_{2t} and e_{2t} only affects ϵ_{2t}. Hence $Temp_t$ can experience a shock when ARF is not modified at first (via e_{2t}) whereas any shock to ARF_t appears through e_1 and then $Temp_t$ is also affected. This is what is explained in the narrative above. The Choleski decomposition of Ω into $\Sigma\Sigma'$ we need to use is therefore (a) since (b) gets the opposite effect. We observe in Figure 8 that in the top row, both variables are shocked at time 0 (a shock to ARF), whereas in the bottom row, only $Temp$ is shocked at time zero: the initial response of ARF is zero.