Forecasting & Predictive Analytics

ESSEC-CentraleSupélec Master in Data Sciences & Business Analytics
Guillaume Chevillon

EXAMINATION

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
   - Show that the process $y_t = \Delta x_t$ 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).
   - 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?

   (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?
   - Show that the series x_t is invertible for $|\theta| < 1$ and find the coefficients in the representation

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

   - What is the optimal forecast (using the Mean-Square Forecast Error criterion) for x_{t+1} given its infinite past $x_t, x_{t-1}, ...$?
   - 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.

   (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.

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?

(b) When x_t is not stationary, do x_t and y_t cointegrate? Please argue simply and find the cointegration relation.

(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}.

(d) Define $w_t = y_t - c y_{t-1}$. 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$(1, 1)$.

(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 $\mathbf{z}_{t+1|t}$.

(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

$$\widehat{y}^*_{t+1|t} = b x_t + \rho \left(x_{t+1} - c x_t \right), \tag{3}$$

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)?

(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

$$\widetilde{y}_{t+1|t} = \gamma x_{t+1}, \tag{4}$$

The associated MSFE is $\mathsf{E}\left[\left(y_{t+1} - \widetilde{y}_{t+1|t}\right)^2\right] = 1 - \rho^2 + \left(b - c\rho\right)^2$. What conclusion do you reach about the relative accuracy of the two conditional forecasts?

(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?

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).

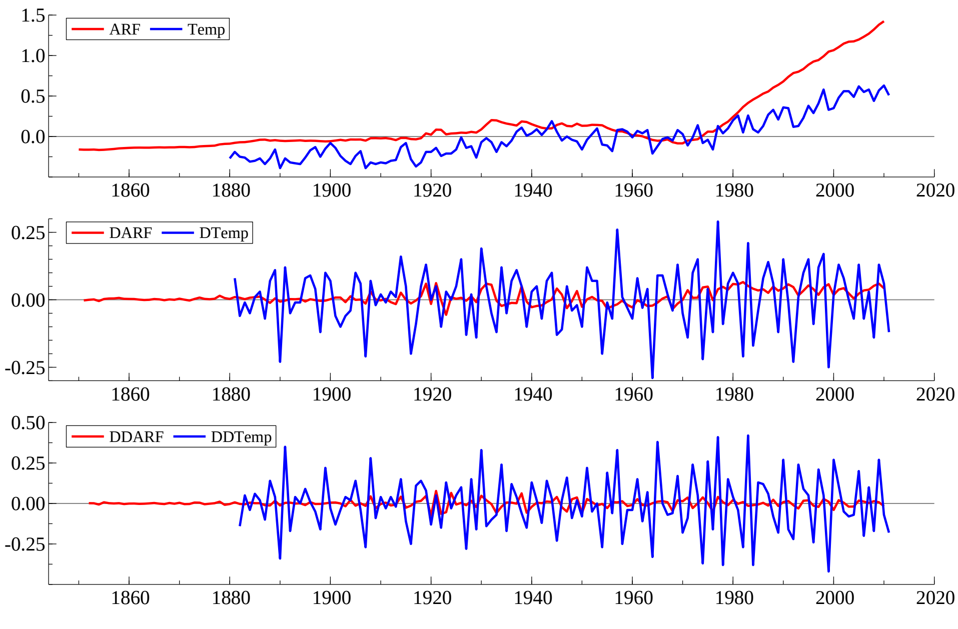

Figure 1: Time Series plots of $Temp$ and ARF and their differences

(a) Figures 1, 2 and 3 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.

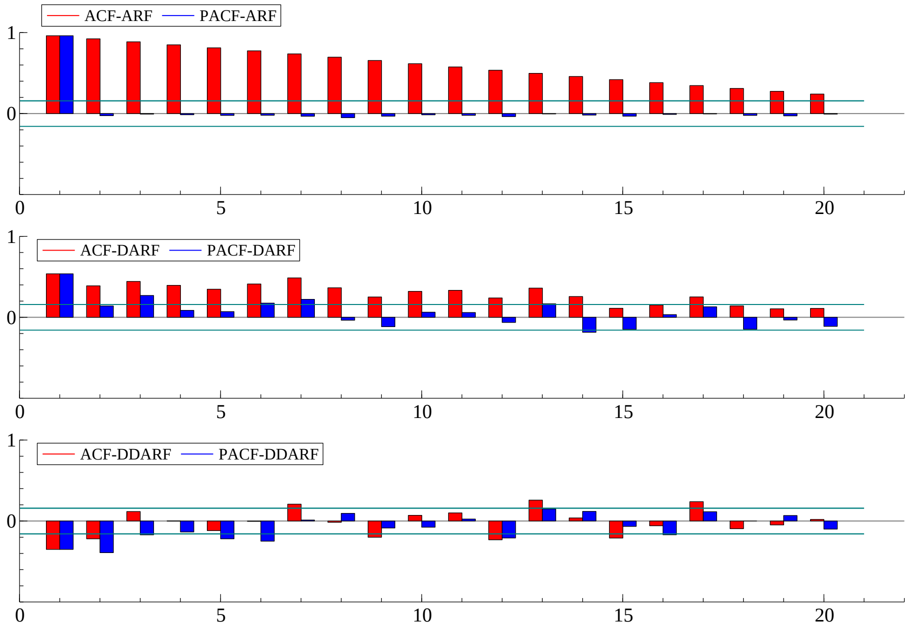

Figure 2: Autocorrelogram and Partial Autocorrelogram of *ARF* and its differences

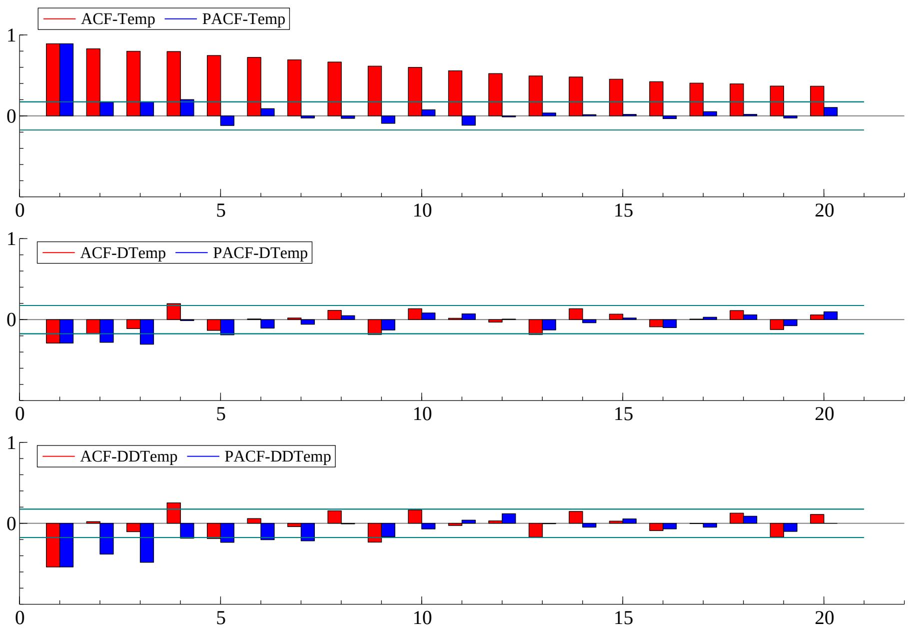

Figure 3: Autocorrelogram and Partial Autocorrelogram of *Temp* and its differences

(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).
- The test is carried out with a constant and a trend included in the model, explain what this means.

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

(c) Let's think further about the peculiar shape of the ACF for $DARF$. When we look at the time series plot of Figure 1, 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 4 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?
- 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

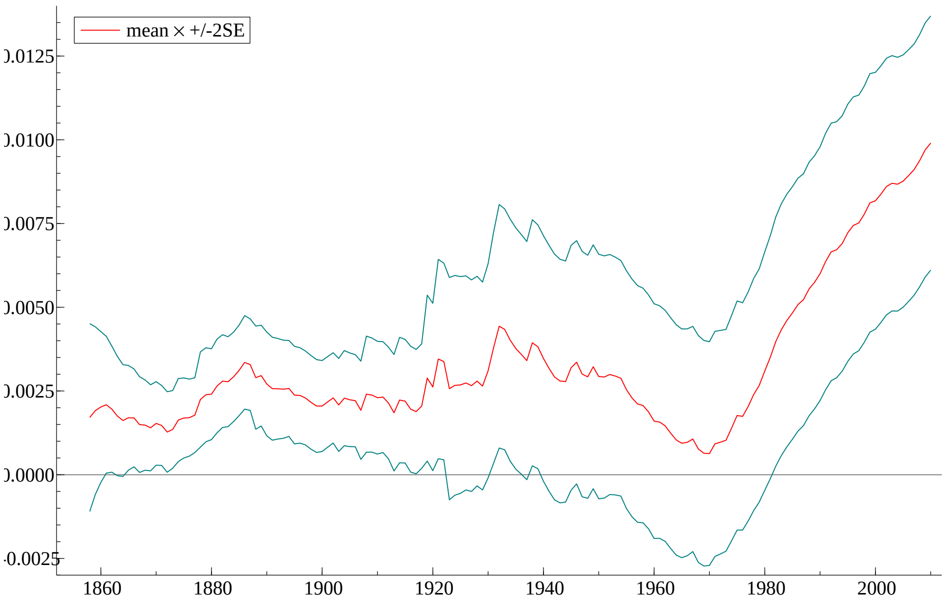

Figure 4: Recursive mean of $DARF$ together with a two-sided 95% confidence interval

the following result (t-statistics are in parentheses below the parameters):

$$DARF_t = \underset{(0.387)}{6.4 \times 10^{-4}} + \underset{(11.0)}{0.036} \times dumm1970_t + DARF\_dum \qquad (5)$$

where $DARF\_dum$ is defined as the residual from the regression. Figure 5 plots $DARF$ and $DARF\_dum$, together with the ACF/PACF of $DARF\_dum$.

i. Explain, in Equation (5), whether there is evidence of a break in mean for $DARF$.

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.

(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 6.

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.

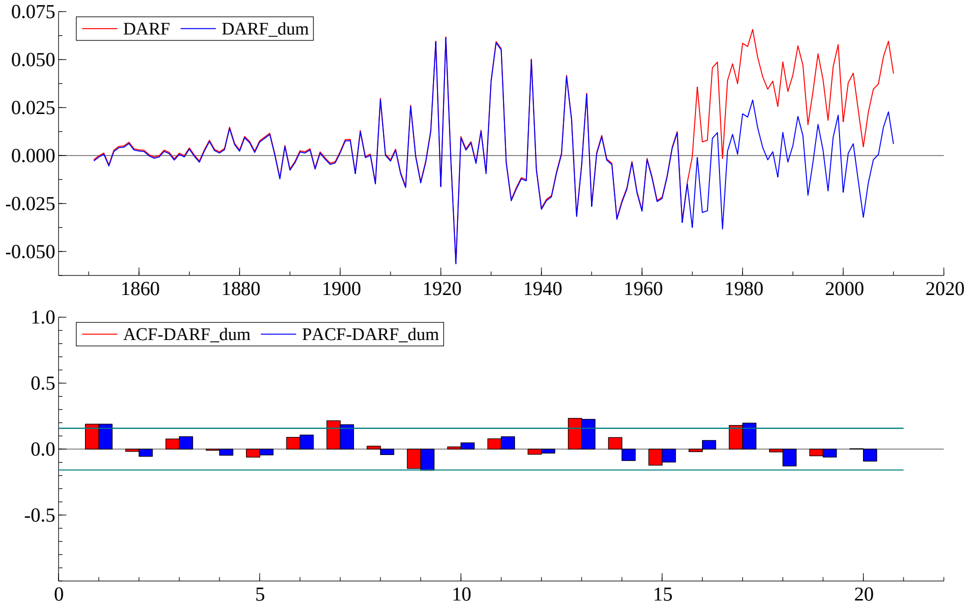

Figure 5: Estimation of a shift in the mean of $DARF$ and ACF/PACF of $DARF$ once corrected for the estimated shift.

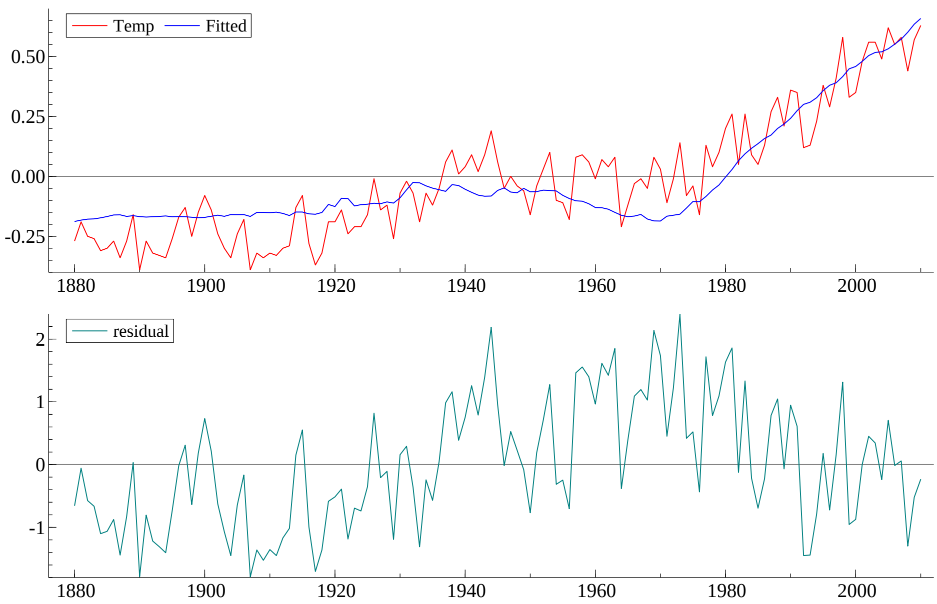

Figure 6: Estimation of the contemporaneous link between temperature and human greenhouse gas emissions.

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?

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).

(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.

ii. We obtained the IRF presented in Figure 7 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.

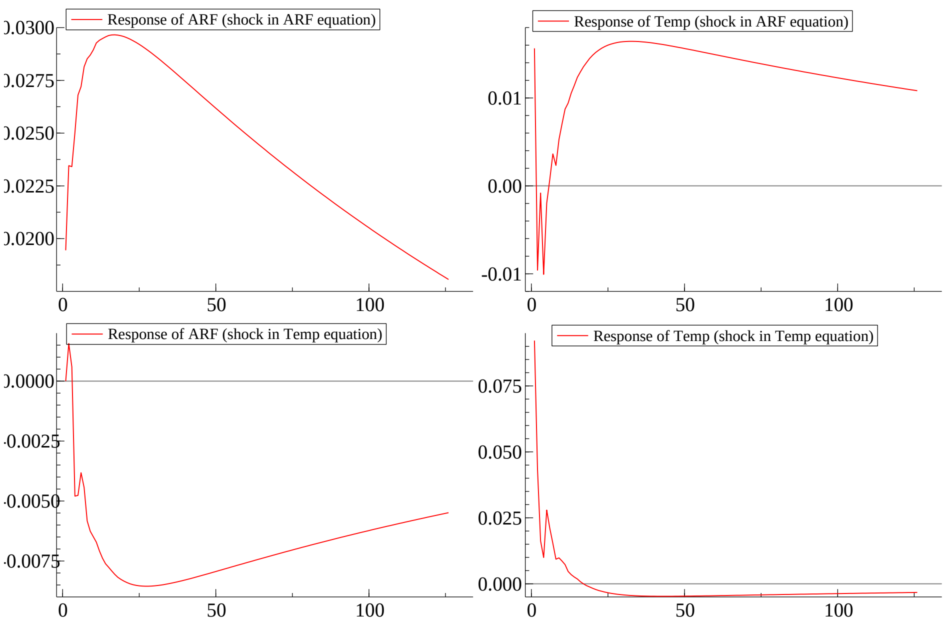

Figure 7: Impulse response functions to shocks in the ARF and $Temp$ equations.

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.