CALENDARS IN JDEMETRA+

DEFINITION OF CALENDAR EFFECTS

We consider in this document the regression variables that can be defined in JD+ to model calendar effects. We call them hereafter "calendar variables (CV)".

A natural way for modelling calendar effects consists in distributing the days of each period in different groups. The regression variable corresponding to a type of day (a group) is simply defined by the number of days it contains for each period. Usual classifications are:

- Trading days (7 groups): each day of the week defines a group (Mondays...Sundays)
- Working days (2 groups): week days and weekends

But we could also consider:

- 3 groups: week days (Mondays->Fridays), Saturdays, Sundays; that solution will be called TD3 hereafter
- ...

The definition of a group could involve partial days. For instance, we could consider that one half of Saturdays belong to week days and the second half to weekends

Specific holidays are often handled as Sundays: they go to the group that contains the Sundays (or more generally to the group corresponding to "non working days"), but they also could be considered separately: for instance 8 groups: trading days (outside specific holidays) + specific holidays.

Example:

Week days + Saturdays + Sundays

	Total	**Week days**	**Saturdays**	**Sundays**
Jan-2013	31	23	4	4
Feb-2013	28	20	4	4
Mar-2013	31	21	5	5
Apr-2013	30	22	4	4
May-2013	31	23	4	4
Jun-2013	30	20	5	5
Jul-2013	31	23	4	4

MEAN AND SEASONAL EFFECTS OF CV

The definition proposed above will lead to regression variables that have a mean effect and a seasonal effect. In the usual decomposition of a series, the mean effect (independent of the period) should be allocated to the trend-cycle component and the fixed seasonal effect (dependent of the period, 0 on average) should be affected to the corresponding component. So, the actual calendar effect should only contain effects that don't belong to the other components. Such a choice is just a sensible convention.

By mean effect and seasonal effect, we consider in JDemetra+ long term theoretical effects and not effects computed on the time span of the considered series (which should be continuously revised).

The mean effect of a CV is the average number of days in its group.

Taking into account that one year has on average 365.25 days, the monthly mean effects for TD3 are:

Week days	365.25/12*5/7 = **21.7411**
Saturdays	365.25/12*1/7 = **4.3482**
Sundays	365.25/12*1/7 = **4.3482**
Total	*365.25/12 =* ***30.4375***

The number of days by period is highly seasonal. So, any set of CVs will contain, at least in some variables, a significant seasonal effect.

The seasonal effect is defined as the average number of days by period (Januaries..., first quarters...), outside the mean effect. Removing that fixed seasonal effects consists in removing for each period the long term average of days that belong to it.

The different definitions are applied below to the week days of TD3 (TD2). If we consider the variable (v) "week days" in TD3 (or TD2), we have:

Period (p)	**Average days**	**Average week days**	**Mean effect ~cm(v)**	**"Seasonal effect" ~ cs(v, p)**
Jan	31	31*5/7=22.1429	21.7411	0.4018
Feb	28.25	28.25*5/7=20.1786	21.7411	-1.5625
Mar	31	31*5/7=22.1429	21.7411	0.4018
Apr	30	30*5/7=21.4286	21.7411	-0.3125
May	31	31*5/7=22.1429	21.7411	0.4018
Jun	30	30*5/7=21.4286	21.7411	-0.3125
Jul	31	31*5/7=22.1429	21.7411	0.4018

For a given time span, the actual "calendar effects" is then easily derived.

Time period (t)	**Week days**	**Mean effect ~cm(v)**	**"Seasonal effect" ~ cs(v, p)**	**"Calendar effect" ~ cc(v, t)**
Jan-2013	23	21.7411	0.4018	0.8571
Feb-2013	20	21.7411	-1.5625	-0.1786
Mar-2013	21	21.7411	0.4018	-1.1429
Apr-2013	22	21.7411	-0.3125	0.5714
May-2013	23	21.7411	0.4018	0.8571
Jun-2013	20	21.7411	-0.3125	-1.4286
Jul-2013	23	21.7411	0.4018	0.8571

Making the splitting between the "mean effect" and the "seasonal effect" is usually unnecessary. We can consider them together (they will be simply called "mean effects") and compute them simply by removing from each CV its average number of days by period. We will only consider those global mean effects in the next paragraph.

LINEAR TRANSFORMATIONS OF THE CV

As far as RegArima model are considered, we can use any non degenerated linear transformation of the CVs; it will give the same results (likelihood, residuals, parameters, joint effect of CV, joint F-

test on the coefficients of the CV...). The linearized series that will be further decomposed is invariant to any linear transformation of the CVs.
However, it should be mentioned that choices of calendar corrections based on tests on the individual T-stats are dependent on the transformation, which is rather arbitrary. This is the case in old versions of Tramo-Seats. That is why joint test (as in the new version of Tramo-Seats) should be preferred

Examples of linear transformation: *contrast variables*

The usual trading days variables are defined by the following transformation: the contrast variables (Mondays - Sundays ... Saturdays - Sundays) are used with the length of periods.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} M \\ T \\ W \\ T \\ F \\ Sat \\ S \end{bmatrix} = \begin{bmatrix} M-S \\ T-S \\ W-S \\ T-S \\ F-S \\ Sat-S \\ Length\ of\ period \end{bmatrix}$$

For the usual working days variables, we use:

$$\begin{bmatrix} 1 & -{}^{5}/_{2} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} Week \\ Weekend \end{bmatrix} = \begin{bmatrix} Contrast\ week \\ Length\ of\ period \end{bmatrix}$$

For TD3, we could use:

$$\begin{bmatrix} 1 & 0 & -6 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} Week \\ Saturday \\ Sunday \end{bmatrix} = \begin{bmatrix} Contrast\ week \\ Contrast\ saturday \\ Length\ of\ period \end{bmatrix}$$

Those transformations have several advantages. They suppress from the contrast variables the mean and the seasonal effects, which are concentrated in the last variable. So, they lead to less correlated variables, which are better estimated.

HANDLING OF SPECIFIC HOLIDAYS

We consider in JDemetra+ the following type of holidays:

- Fixed day, corresponding to a fixed date in the year (for instance July, 21 for Belgium, New Year, Christmas).
- Easter related days, corresponding to days that are defined in relation to Easter (Easter +- n days; example: Ascension...)
- Fixed week days, corresponding to a fixed day in a give week of a give month (Labour Day...)

From a conceptual point of view, specific holidays are handled exactly the same way as the other days. We just have to decide in which group they are put. Usually - and it is the convention used in JDemetra+, they are handled has Sundays. Of course, except if the holiday falls on a Sunday, we have to correct both the group that should have contained the holiday and the group that contains the Sundays.

The trickiest aspect of specific holidays is the way they impact on the mean and the seasonal effects of CVs. We consider below the seasonal corrections that should be applied on the Mondays...Sundays (which are usually the basis for the definition of other calendar variables), following the kind of holidays.

The corrections are applied to the period(s) that may contain the holiday.

Fixed day

The probability that the holiday falls on a Sunday is 1/7 (and the same for the other days). The probability to have 1 Sunday more is 6/7. The effect on the means for the period that contains the date is (the correction on the calendar effect has the opposite sign):

Sundays	Others days	Contrast days
+ 6/7	- 1/7	-1

Easter related days

Easter related days always fall the same week day (X). However, they can fall during different periods (months or quarters). Suppose that, taking into account the distribution of the dates for Easter, the probability that the holiday falls during the period m (m+1) is p (1-p). Then, the effects on the seasonal means are:

Period	Sundays	Days X	Others days	Contrast X	Other contrasts
m	+ p	- p	0	- 2*p	- p
m+1	+ (1-p)	- (1-p)	0	- 2*(1-p)	- (1-p)

The distribution of the dates for Easter may be approximated in different ways.

A first solution consists in using some well known algorithms for computing Easter on a very long period. JDemetra+ provides the Meeus/Jones/Butcher's and the Ron Mallen's algorithms (they are identical till 4100, but they slightly differ after that date).
Another approach consists in deriving a raw theoretical distribution based on the definition of Easter. It is the solution used for Easter related days. It is shortly explained below.

"Easter is the first Sunday after the full moon (the Paschal Full Moon) following the northern hemisphere's vernal equinox. Ecclesiastically, the equinox is reckoned to be on 21 March (even though the equinox occurs, astronomically speaking, on 20 March in most years), and the "Full Moon" is not necessarily the astronomically correct date. Easter is delayed by 1 week if the full moon is on Sunday. The date of Easter therefore varies between 22 March and 25 April" (Wikipedia). Taking into account that an average lunar month is 29.53059 days, we can derive raw formulae for the approximated distribution of Easter (they don't take into account the actual ecclesiastical moon calendar).

22/3	1/7 * 1/29.53059
23/3	1/7 * 2/29.53059
24/3	1/7 * 3/29.53059
25/3	1/7 * 4/29.53059
26/3	1/7 * 5/29.53059
27/3	1/7 * 6/29.53059
28/3	1/29.53059
29/3	1/29.53059
...	...
18/4	1/29.53059
19/4	1/7 * (6 + 1.53059)/29.53059
20/4	1/7 * (5 + 1.53059)/29.53059

21/4	1/7 * (4 + 1.53059)/29.53059
22/4	1/7 * (3 + 1.53059)/29.53059
23/4	1/7 * (2 + 1.53059)/29.53059
24/4	1/7 * (1 + 1.53059)/29.53059
25/4	1/7 * 1.53059/29.53059

For example,

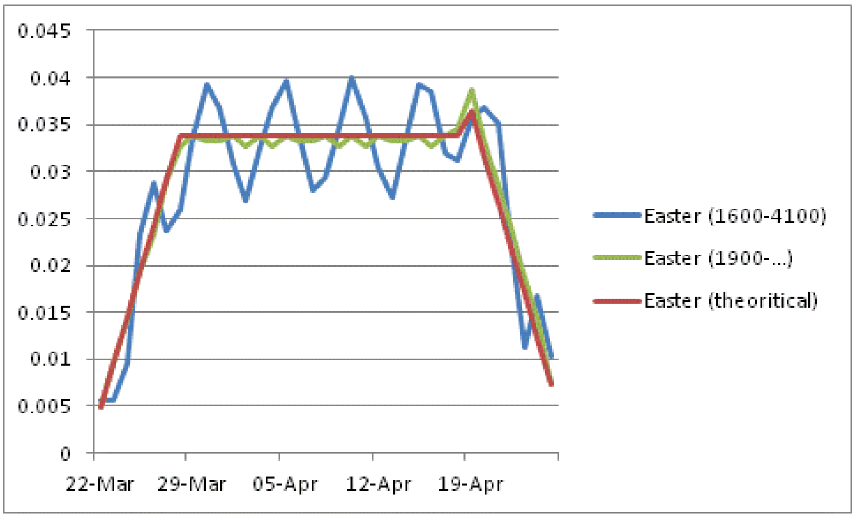
The probability that Easter falls on 25/3 is computed as follows:

- Probability that 25/3 is a Sunday: 1/7
- Probability that the full moon is on 21/3... 24/3 is 5/29.53059

The probability that Easter falls on 23/4 is computed as follows:

- Probability that 23/4 is a Sunday: 1/7
- Probability that the full moon is on 16/4, 17/4: 2/29.53059
- Probability that the fool moon is on 18/4 1.53059/29.53059[1]

The different solutions are presented below.



Fixed week days

Fixed week days always fall the same week day (X) and the same period. Their effect on the seasonal means is:

Sundays	Days X	Others days
+ 1	- 1	0

The impact of fixed week days on the regression variables is 0: the effect itself is compensated by the correction for the mean effect.

Remark

The long term corrections don't take into account the fact that some moving holidays could fall on the same day (for instance the May Day and the Ascension). Those events are exceptional, so that they impact on the final result is not significant.

[1] 18/4 is the last acceptable date for the full Moon.

IMPACT OF THE MEAN EFFECTS

When the Arima model contains a seasonal differencing - something that should always happen with CV - the mean effects contained in the CVs are automatically eliminated, so that they don't modify the estimation. The model is indeed estimated on the series/regression variables after differencing.

However, they lead to a different linearized series (y_{lin}). We will consider below the impact of other corrections (mean and/or fixed seasonal) on the decomposition. Such corrections could be obtained for instance by applying other solutions for the long term corrections or by computing them on the time span of the series.

Consider first a model with "correct" calendar effects (C), i.e. effects without mean and fixed seasonal effects. To simplify the problem, the model has no other regression effects.
We can write:

$$y_{lin} = y - C$$
$$T = F_T(y_{lin})$$
$$S = F_S(y_{lin}) + C$$
$$I = F_I(y_{lin})$$

where $F_X$is the linear filter for the component X.

Consider next other calendar effects ($\tilde{C}$) that contain some mean (cm, integrated to the final trend) and fixed seasonal effects (cs, integrated to the final seasonal).
We have now:

$$\tilde{C} = C + cm + cs$$
$$\tilde{y}_{lin} = y - \tilde{C} = y_{lin} - cm - cs$$
$$\tilde{T} = F_T(\tilde{y}_{lin}) + cm$$
$$\tilde{S} = F_S(\tilde{y}_{lin}) + C + cs$$
$$\tilde{I} = F_I(\tilde{y}_{lin})$$

Taking into account that $F_X$is a linear transformation and that[2]

$$F_T(cm) = cm, \qquad F_T(cs) = 0$$
$$F_S(cm) = 0, \qquad F_S(cs) = cs$$
$$F_I(cm) = 0, \qquad F_I(cs) = 0$$

We have:

$$\tilde{T} = F_T(\tilde{y}_{lin}) + cm = F_T(y_{lin}) - cm + cm = T$$
$$\tilde{S} = F_S(\tilde{y}_{lin}) + C + cs = F_S(y_{lin}) - cs + C + cs = S$$
$$\tilde{I} = I$$

If we don't take into account the effects and apply the same approach as in the "correct" calendar effects, we will get:

$$\check{T} = F_T(\tilde{y}_{lin}) = T - cm$$
$$\check{S} = F_S(\tilde{y}_{lin}) + \tilde{C} = S + cm$$
$$\check{I} = F_I(\tilde{y}_{lin}) = I$$

[2] In the case of Seats, the properties can be trivially derived from the matrix formulation of signal extraction. They are also valid for X11 (additive)

The trend, the seasonal and the seasonally adjusted series will only differ by a (usually small) constant.

So, the decomposition doesn't depend on the mean and fixed seasonal effects used for the CVs, provided that those effects are integrated in the corresponding final components. If we don't take into account those corrections, the main series of the decomposition will only differ by a constant.

SPECIAL CASE: HOLIDAYS WITH A VALIDITY PERIOD

When a holiday is valid only for a given time span, JDemetra+ applies the long term mean corrections only on the corresponding period. However, those corrections are computed the same way as in the general case[3].

It is important to note that using or not mean corrections will impact in that case the estimation of the RegArima model. Indeed, the mean corrections don't disappear after differencing. The differences between the SA series computed with or without mean corrections will no longer be constant.

SUMMARY OF THE METHOD USED IN JDEMETRA+ TO COMPUTE TD AND WD EFFECTS

1. Compute for all the periods the number of Mondays...Sundays
2. [Correct those variables with specific holidays (Sundays: +1, Days of the holiday: -1; no correction if the holiday falls on a Sunday), taking into account the validity period of the holiday]
3. Compute the usual contrast variables for trading days and working days
4. [Correct the variables for long term mean effects, taking into account the validity period of the holiday; see below for the different cases]

Remark:

All the different corrections may receive a weight corresponding to the part of the holiday considered as a Sunday.

EXAMPLE

We consider below a small example, with three holidays:

New Year	Fixed day: 01-01
Shrove Tuesday	47 days before Easter, till end 2012
Freedom day Portugal	Fixed day: 25-04

The monthly calendar is computed for 2012 and 2013

[3] The first versions of JDemetra+ applied the corrections on the complete time span of the time series.

Step1. Number of Mondays... Sundays

	M	T	W	T	F	Sat	S
Jan-12	5	5	4	4	4	4	5
Feb-12	4	4	5	4	4	4	4
Mar-12	4	4	4	5	5	5	4
Apr-12	5	4	4	4	4	4	5
May-12	4	5	5	5	4	4	4
Jun-12	4	4	4	4	5	5	4
Jul-12	5	5	4	4	4	4	5
Aug-12	4	4	5	5	5	4	4
Sep-12	4	4	4	4	4	5	5
Oct-12	5	5	5	4	4	4	4
Nov-12	4	4	4	5	5	4	4
Dec-12	5	4	4	4	4	5	5
Jan-13	4	5	5	5	4	4	4
Feb-13	4	4	4	4	4	4	4
Mar-13	4	4	4	4	5	5	5
Apr-13	5	5	4	4	4	4	4
May-13	4	4	5	5	5	4	4
Jun-13	4	4	4	4	4	5	5
Jul-13	5	5	5	4	4	4	4
Aug-13	4	4	4	5	5	5	4
Sep-13	5	4	4	4	4	4	5
Oct-13	4	5	5	5	4	4	4
Nov-13	4	4	4	4	5	5	4
Dec-13	5	5	4	4	4	4	5

Step 2. Contrast variables (series corrected for mean effects)

	M	T	W	T	F	Sat	Length
Jan-12	0	0	-1	-1	-1	-1	0
Feb-12	0	0	1	0	0	0	0.75
Mar-12	0	0	0	1	1	1	0
Apr-12	0	-1	-1	-1	-1	-1	0
May-12	0	1	1	1	0	0	0
Jun-12	0	0	0	0	1	1	0
Jul-12	0	0	-1	-1	-1	-1	0
Aug-12	0	0	1	1	1	0	0
Sep-12	-1	-1	-1	-1	-1	0	0
Oct-12	1	1	1	0	0	0	0
Nov-12	0	0	0	1	1	0	0
Dec-12	0	-1	-1	-1	-1	0	0
Jan-13	0	1	1	1	0	0	0
Feb-13	0	0	0	0	0	0	-0.25
Mar-13	-1	-1	-1	-1	0	0	0
Apr-13	1	1	0	0	0	0	0
May-13	0	0	1	1	1	0	0
Jun-13	-1	-1	-1	-1	-1	0	0
Jul-13	1	1	1	0	0	0	0
Aug-13	0	0	0	1	1	1	0
Sep-13	0	-1	-1	-1	-1	-1	0
Oct-13	0	1	1	1	0	0	0
Nov-13	0	0	0	0	1	1	0

Dec-13	5	5	4	4	4	4	5

Step 3. Corrections for holidays

New Year
2012: on Sunday: nothing to do
2013: on Tuesday: Tuesday --1, Sunday +1, to be applied on the contrasts in Jan (-2 for T, -1 for the others)

Shrove Tuesday
2012 only. Tuesday --1, Sunday +1, to be applied on the contrasts in Feb (-2 for T, -1 for the others)

Freedom Day
2012: on Wednesday: Wednesday --1, Sunday +1, to be applied on the contrasts in April (-2 for W, -1 for the others)
2013: on Thursday: Thursday --1, Sunday +1, to be applied on the contrasts in April (-2 for T, -1 for the others)

	M	T	W	T	F	Sat	Length
Jan-12	0	0	-1	-1	-1	-1	0
Feb-12	-1	-2	0	-1	-1	-1	0.75
Mar-12	0	0	0	1	1	1	0
Apr-12	-1	-2	-3	-2	-2	-2	0
May-12	0	1	1	1	0	0	0
Jun-12	0	0	0	0	1	1	0
Jul-12	0	0	-1	-1	-1	-1	0
Aug-12	0	0	1	1	1	0	0
Sep-12	-1	-1	-1	-1	-1	0	0
Oct-12	1	1	1	0	0	0	0
Nov-12	0	0	0	1	1	0	0
Dec-12	0	-1	-1	-1	-1	0	0
Jan-13	-1	-1	0	0	-1	-1	0
Feb-13	0	0	0	0	0	0	-0.25
Mar-13	-1	-1	-1	-1	0	0	0
Apr-13	0	0	-1	-2	-1	-1	0
May-13	0	0	1	1	1	0	0
Jun-13	-1	-1	-1	-1	-1	0	0
Jul-13	1	1	1	0	0	0	0
Aug-13	0	0	0	1	1	1	0
Sep-13	0	-1	-1	-1	-1	-1	0
Oct-13	0	1	1	1	0	0	0
Nov-13	0	0	0	0	1	1	0
Dec-13	0	0	-1	-1	-1	-1	0

Step 4. Long term corrections

Remark: long term corrections are applied on each year of the validity period of the holiday

New Year:
Correction on the contrasts: +1, to be applied on January of 2012 and 2013

Shrove Tuesday

Shrove Tuesday may fall either in January or in March

It will fall in March if Easter is on or after 17 April.
Taking into account the theoretical distribution of Easter, it gives:
prob(March) = .22147, prob(February)=.77853

The correction of the contrasts will be +1.55707 for T in Feb 2012, +.77853. for the other contrasts
The correction of the contrasts will be +.44293 for T in March 2012, +.22147 for the other contrasts

Freedom Day:
Correction on the contrasts: +1, to be applied on April of 2012 and 2013

	M	T	W	T	F	Sat	Length
Jan-12	1	1	0	0	0	0	0
Feb-12	-0.22115	-0.44229	0.778853	-0.22115	-0.22115	-0.22115	0.75
Mar-12	0.221147	0.442293	0.221147	1.221147	1.221147	1.221147	0
Apr-12	0	-1	-2	-1	-1	-1	0
May-12	0	1	1	1	0	0	0
Jun-12	0	0	0	0	1	1	0
Jul-12	0	0	-1	-1	-1	-1	0
Aug-12	0	0	1	1	1	0	0
Sep-12	-1	-1	-1	-1	-1	0	0
Oct-12	1	1	1	0	0	0	0
Nov-12	0	0	0	1	1	0	0
Dec-12	0	-1	-1	-1	-1	0	0
Jan-13	0	0	1	1	0	0	0
Feb-13	0	0	0	0	0	0	-0.25
Mar-13	-1	-1	-1	-1	0	0	0
Apr-13	1	1	0	-1	0	0	0
May-13	0	0	1	1	1	0	0
Jun-13	-1	-1	-1	-1	-1	0	0
Jul-13	1	1	1	0	0	0	0
Aug-13	0	0	0	1	1	1	0
Sep-13	0	-1	-1	-1	-1	-1	0
Oct-13	0	1	1	1	0	0	0
Nov-13	0	0	0	0	1	1	0
Dec-13	0	0	-1	-1	-1	-1	0