Change detection exploration

Lisa Hopcroft

14/09/2021

Investigating the {gets} package

This document has been generating in an attempt to understand the output from the {gets} package as used in the Change Detection repo, as at this commit.

An `arguments` list is used here to mimic the input from the command line.

It uses an example file (OUTPUT.Rdat) which was contains dummy data generated via OpenSAFELY for PINCER.

Results calibration

There are some variables that control which results are extracted from the models. These are:

- `known.t` (0): time of known intervention in the sample (e.g., medication became available as generic at observation)
- `break.t.lim` (0.8): proportion offset after a break
- `slope.lim` (0.5): proportion of slope drop for construction of slope measure

Definition of offset: to complete

Capturing the relevant results

There will be a row for each of the variables that have been analysed. The results outputs are stored in a dataframe (`results`) with the following columns:

- `is.nbreak`: number of breaks
- `is.tfirst`: first break (up/down/both as requested)
- `is.tfirst.pknown`: First break after a known intervention date (up/down/both as requested)
- `is.tfirst.pknown.offs`: First break after a known intervention date not offset by a XX% increase (up/down/both as requested)
- `is.tfirst.offs`: First break not offset by a XX% increase (up/down/both as requested)
- `is.tfirst.big`: steepest break as identified by is.slope.ma

Specific information regarding the steepest segment are recorded:

- `is.slope.ma`: Average slope over steepest segment contributing at least XX% of total drop
- `is.slope.ma.prop`: Average slope as proportion to prior level
- `is.slope.ma.prop.lev`: Percentage of the total drop the segment used to evaluate the slope makes up

Specific information regarding the level measures are recorded in these variables:

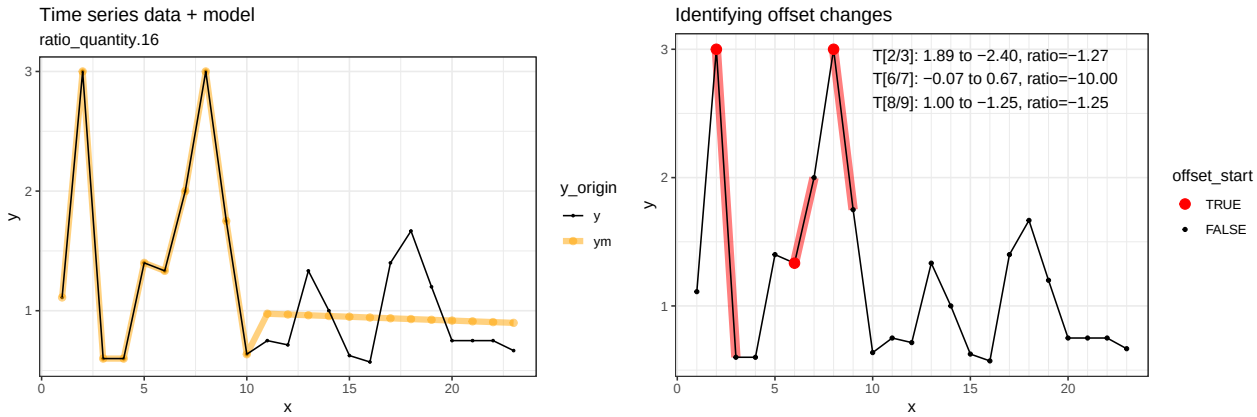
- `is.intlev.initlev`: Pre-drop level
- `is.intlev.finallev`: End level
- `is.intlev.levd`: Difference between pre and end level
- `is.intlev.levdprop`: Proportion of drop

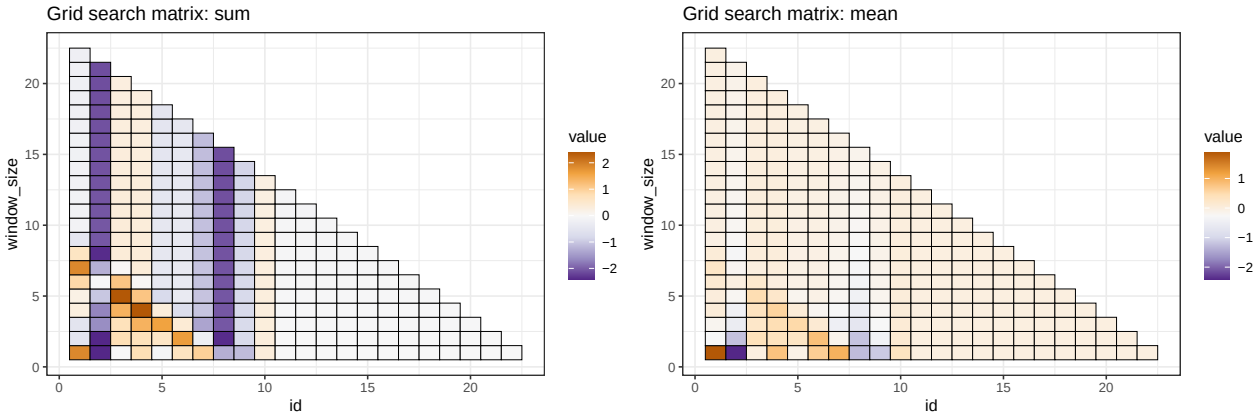
Looking at the results

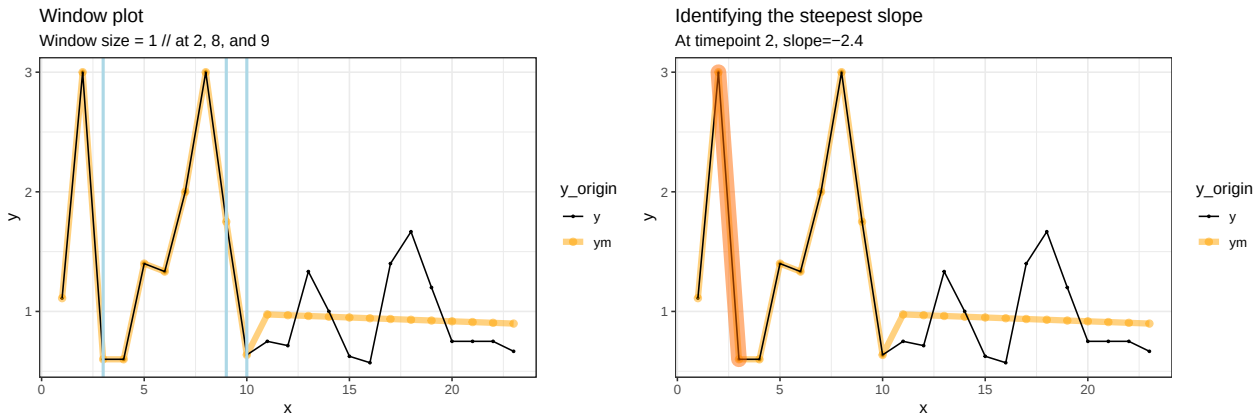
There are 7 variables in the results file that we want to look at: ratio_quantity.16, ratio_quantity.19, ratio_quantity.22, ratio_quantity.25, ratio_quantity.20, ratio_quantity.29, and ratio_quantity.34.

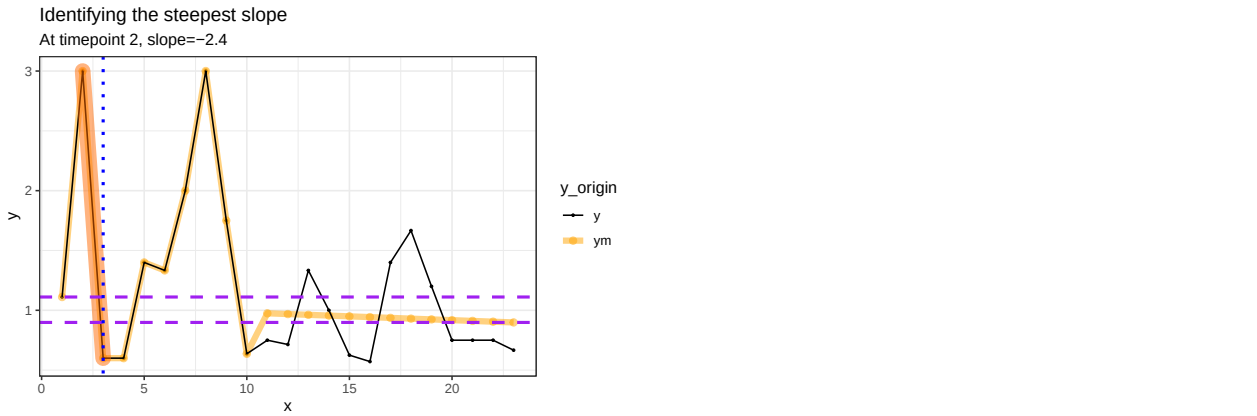
```
## [1] ratio_quantity.16
```

```
## Warning: The ‘x‘ argument of ‘as_tibble.matrix()‘ must have unique column names if ‘.name_repair‘ is
## Using compatibility ‘.name_repair‘.
```

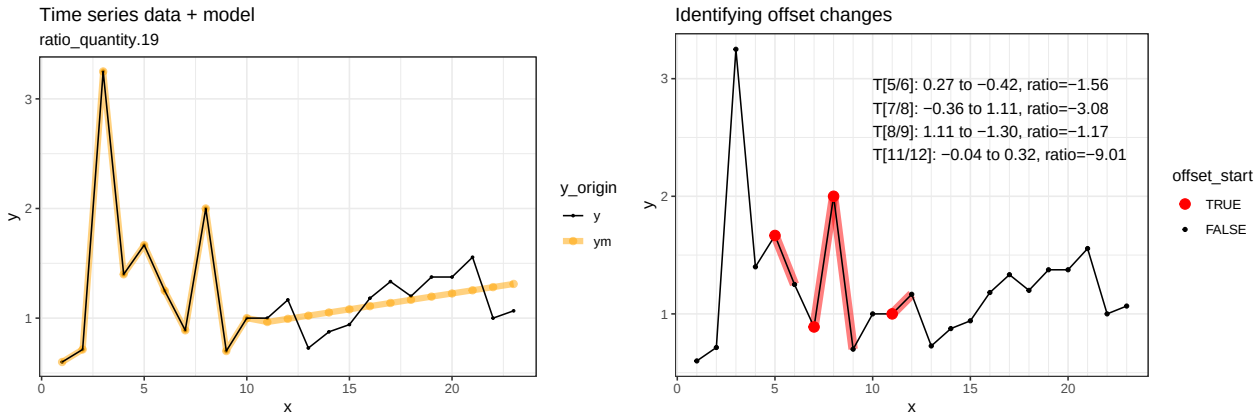


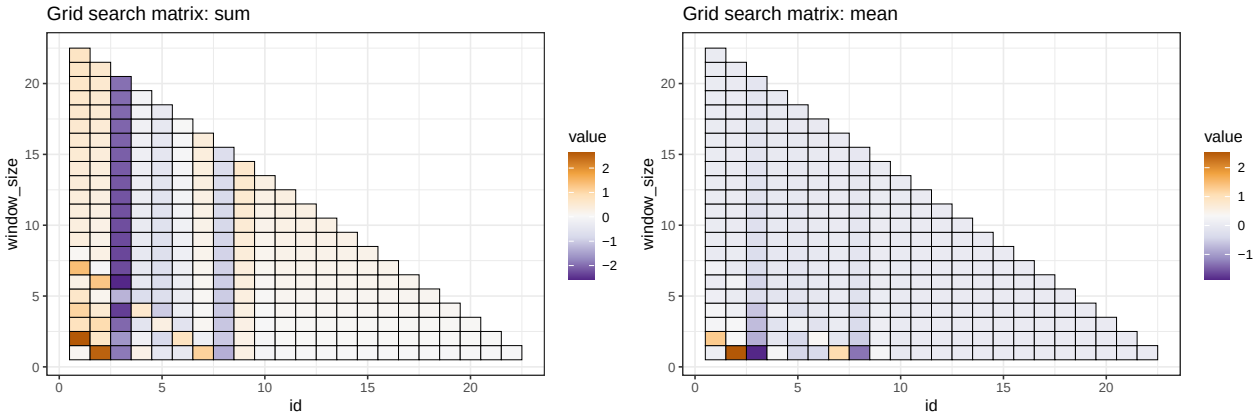


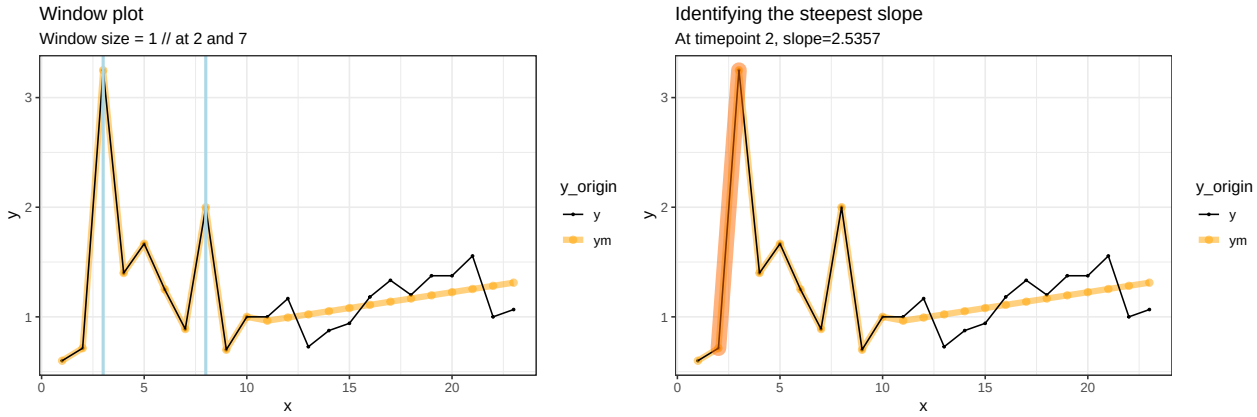


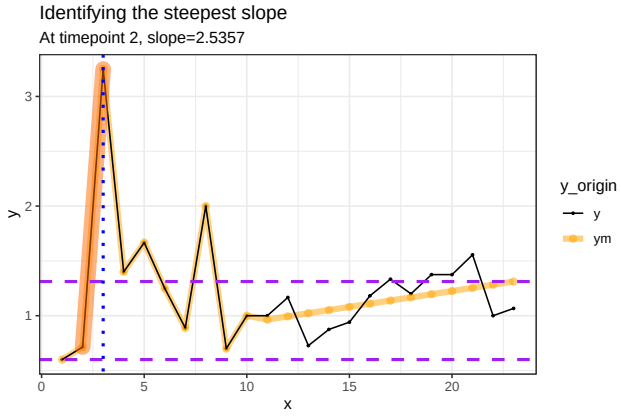



```
## [2] ratio_quantity.19
```

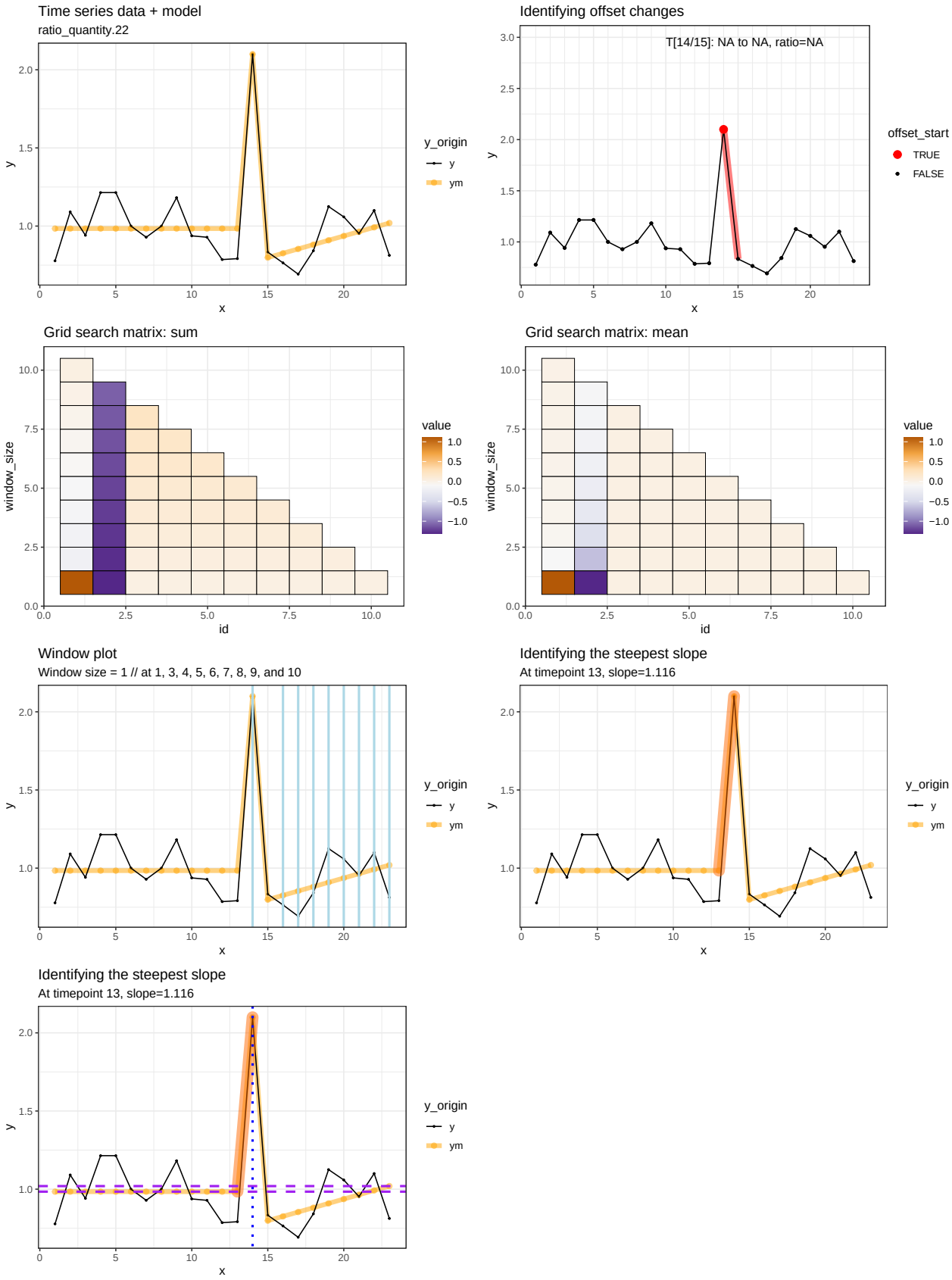




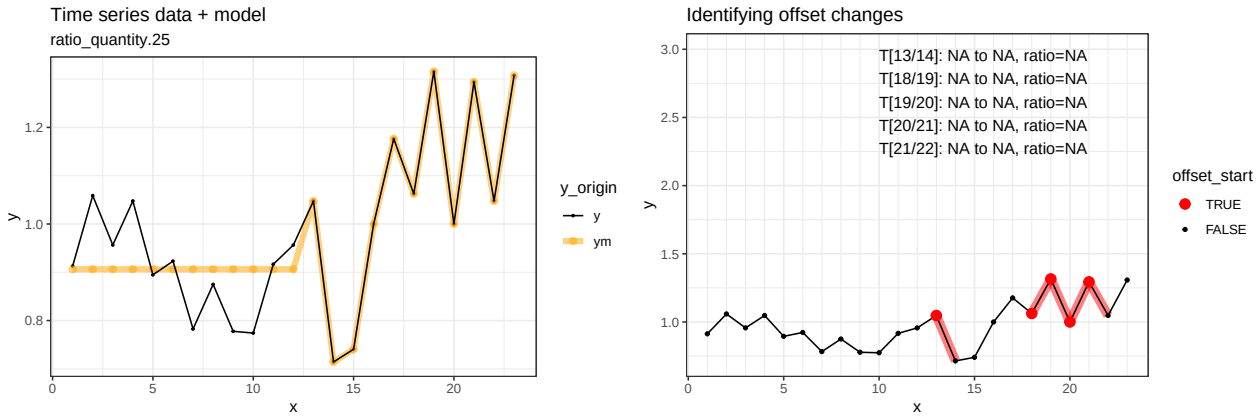


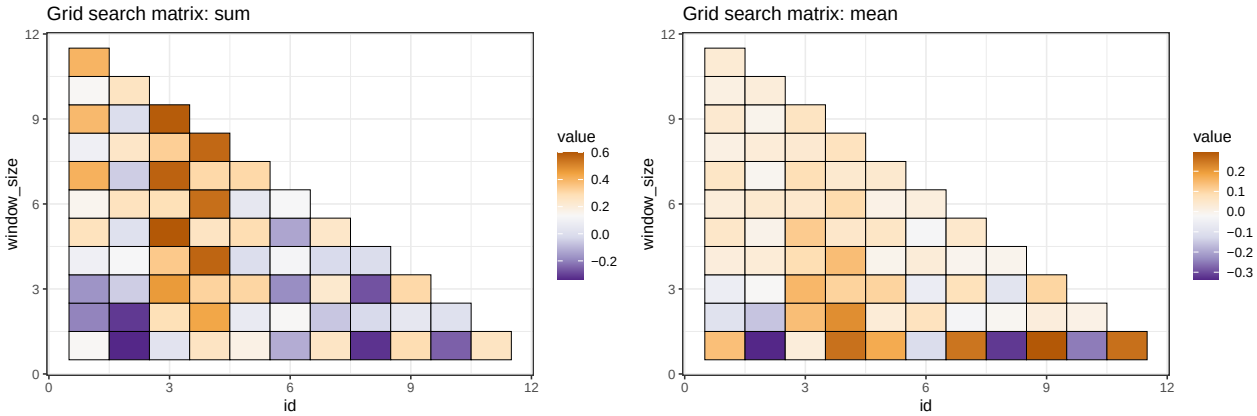


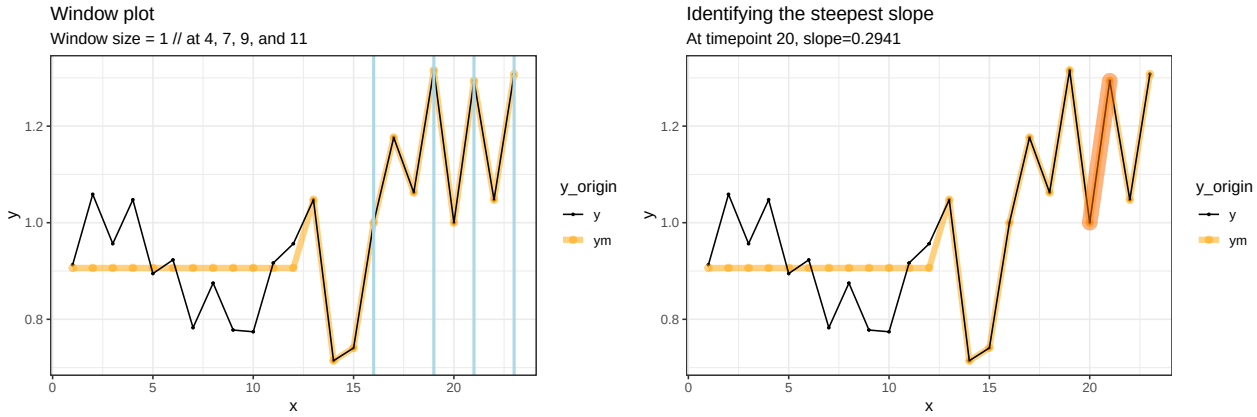

```
## [3] ratio_quantity.22
```

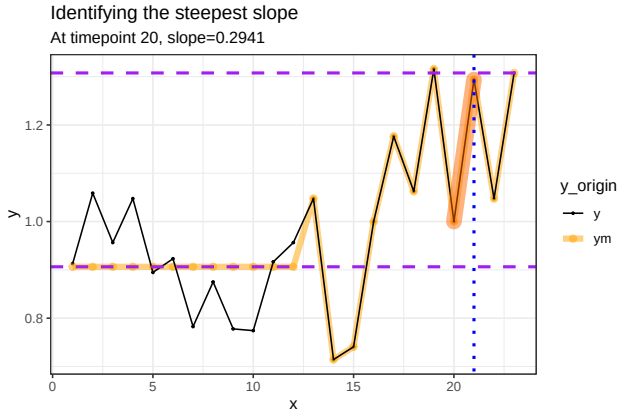


[4] ratio_quantity.25

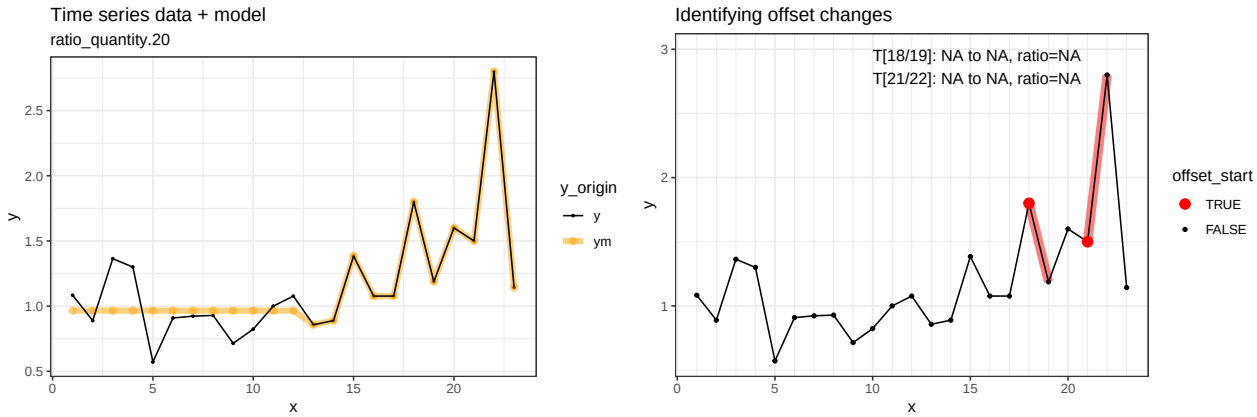


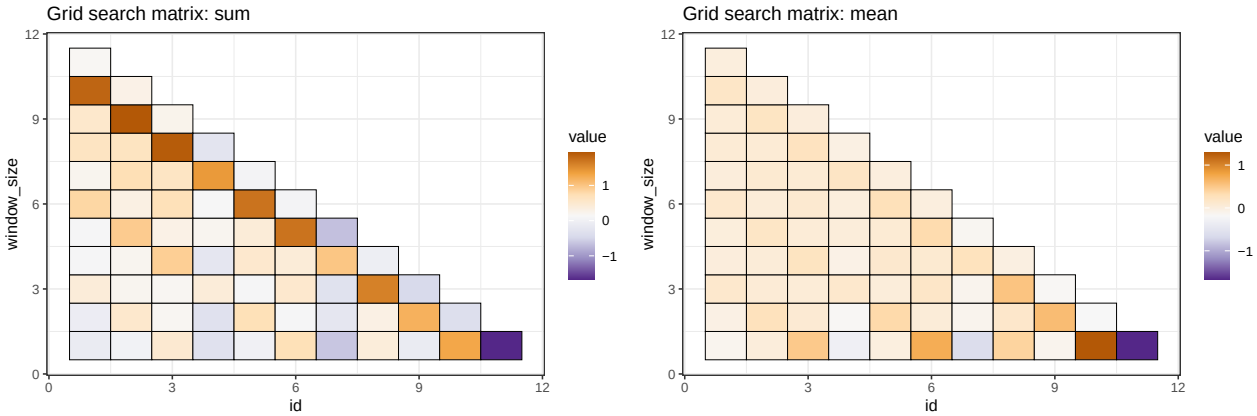


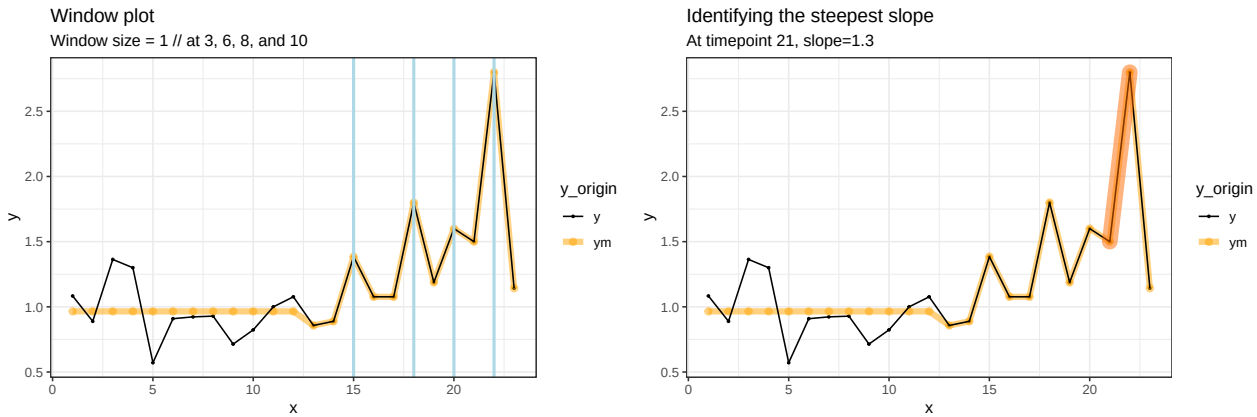


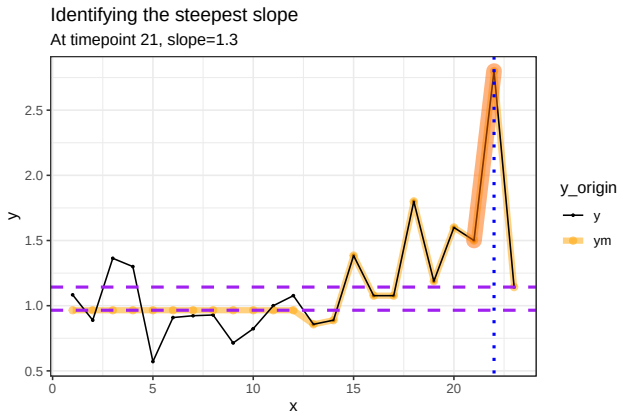



```
## [5] ratio_quantity.20
```

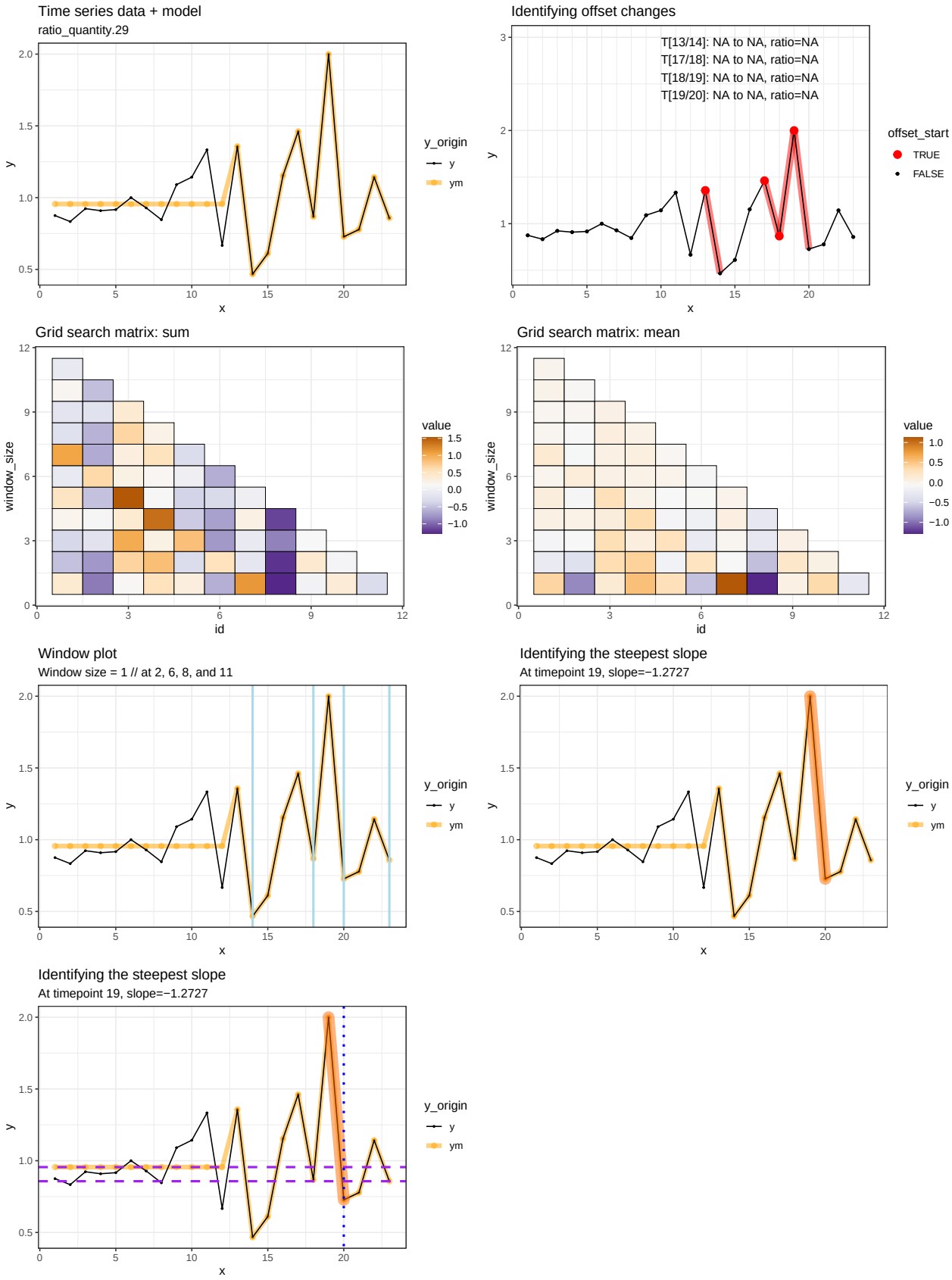




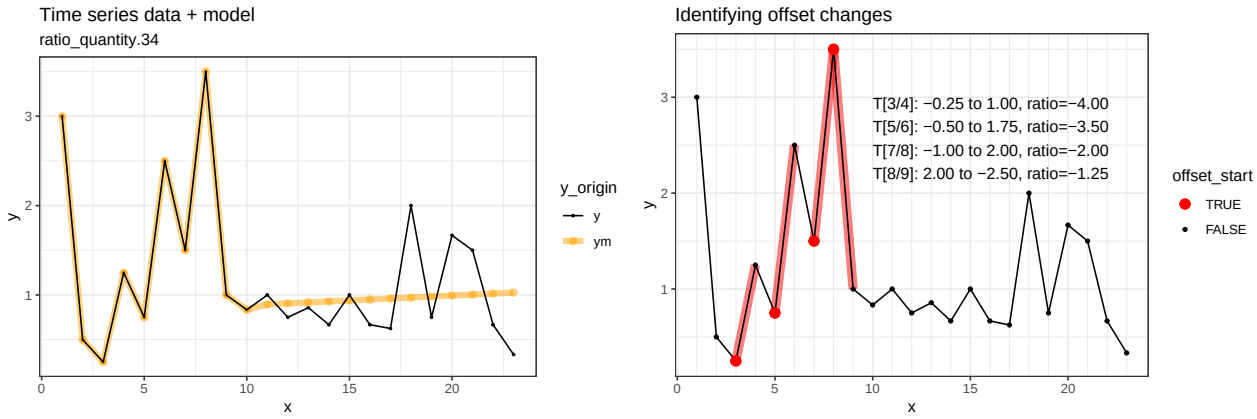




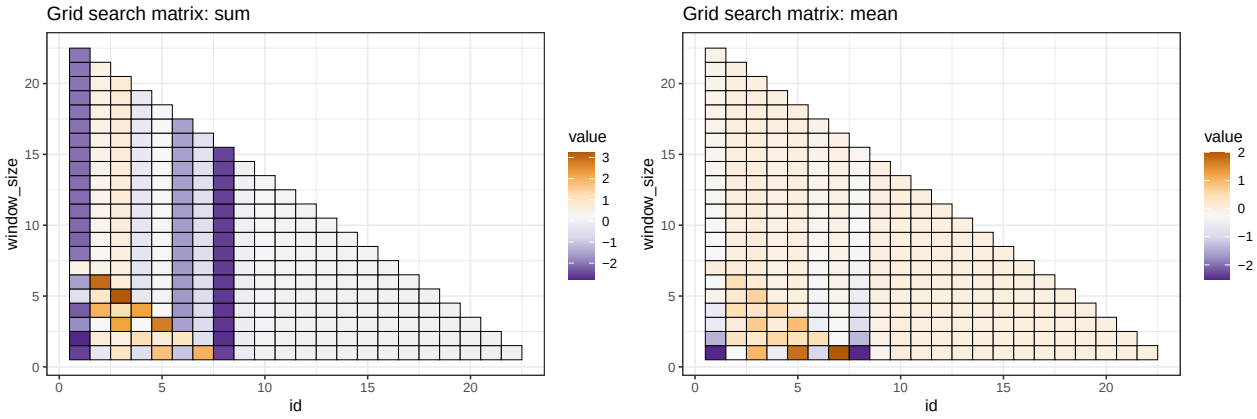

```
## [6] ratio_quantity.29
```



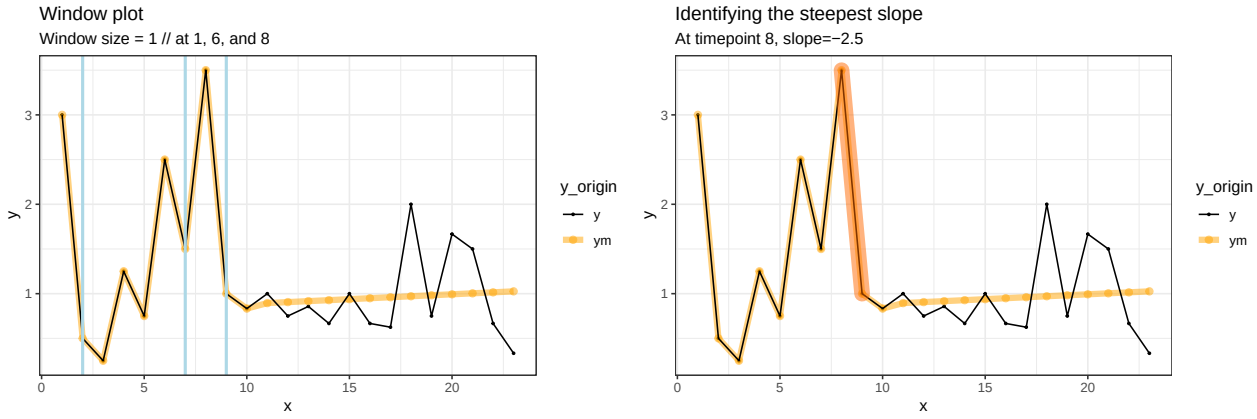

```
## [7] ratio_quantity.34
```


Time series data + model
ratio_quantity.34
y
x
y_origin
y
ym

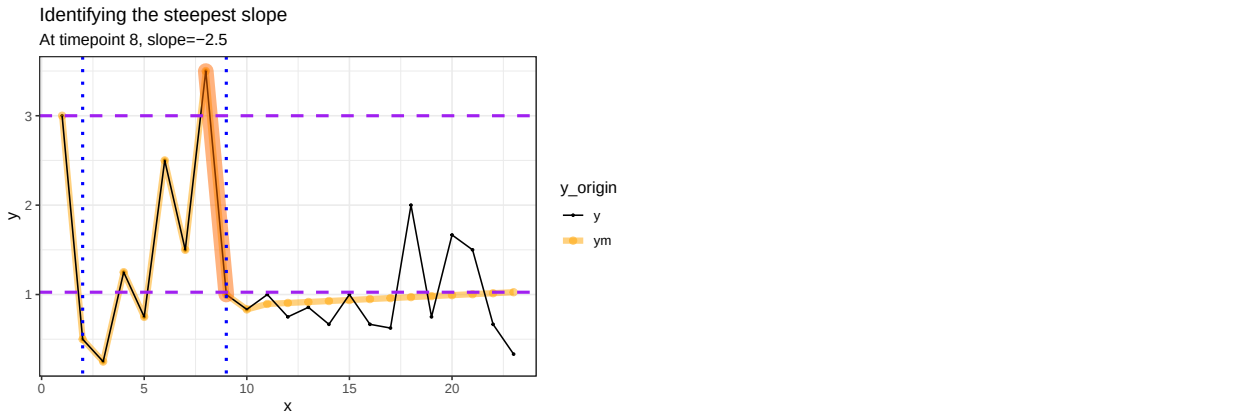
Identifying offset changes
T[3/4]: −0.25 to 1.00, ratio=−4.00
T[5/6]: −0.50 to 1.75, ratio=−3.50
T[7/8]: −1.00 to 2.00, ratio=−2.00
T[8/9]: 2.00 to −2.50, ratio=−1.25
y
x
offset_start
TRUE
FALSE


Grid search matrix: sum
window_size
id
value

Grid search matrix: mean
window_size
id
value


Window plot
Window size = 1 // at 1, 6, and 8
y
x
y_origin
y
ym

Identifying the steepest slope
At timepoint 8, slope=−2.5
y
x
y_origin
y
ym


Identifying the steepest slope
At timepoint 8, slope=−2.5
y
x
y_origin
y
ym