

# Mend - A Package for Mending Time Series with Missing Observations and Structural Breaks

by Ekkehart Schlicht

Version 1, May 2011

The package implements a method for mending time series with missing observations and structural breaks. The method is described in Ekkehart Schlicht: "Trend Extraction from Time Series with Structural Breaks and Missing Observations", *Journal of the Japan Statistical Society* Vol. 38 (2008), No. 2, pages 285-292, freely available at [http://www.jstage.jst.go.jp/article/jjss/38/2/285/\\_pdf](http://www.jstage.jst.go.jp/article/jjss/38/2/285/_pdf).

The package deals with a time series  $x = \{x_1, x_2, \dots, x_T\}$  of length  $T$ . The value  $x_t$  denotes the value of the time series at time  $t$ . A missing value is indicated by a blank, as for the forth element in  $\{1, 2, 1, , 3, 4, 6\}$ . Points where structural breaks occur are supplied by a vector  $bp = \{bp_1, bp_2, \dots\}$  where  $bp_i$  gives the point after which the structural break occurs. The method is designed for allowing trend extraction by means of the Leser filter (also known as the HP-Filter) to enable dealing with time series with gaps and structural breaks.

The package provides the following three function for mending a time series:

*Mend[x,  $\alpha$ ]* repairs a time series with missing observations  $x$  by filling the gaps with appropriate estimates. The parameter  $\alpha$  gives the smoothing parameter for the Leser filter (also known as the HP filter). If this parameter is not provided, the default  $\alpha=100$  is used..

*Mend[x,  $\alpha$ , bp]* estimates dummies for the break points and returns the time series  $x$  adjusted for the break points. It also fills any missing values. If the parameter  $\alpha$  is not provided, as in *Mend[x, bp]*, the default value  $\alpha=100$  is used.

*Mend[ ]* returns a list  $\{y, \text{breaks}, \text{gaps}\}$

- $y = \{y_1, y_2, \dots, y_T\}$  is the mended time series,
- $\text{breaks} = \{\text{break}_1, \text{break}_2, \dots\}$  gives the breaks, with  $\text{break}_i = \{bp_i, d_i\}$  denoting the position  $bp_i$  of break  $i$  and  $d_i$  denoting the dummy.
- $\text{gaps} = \{\text{gap}_1, \text{gap}_2, \dots\}$  gives the gaps, with  $\text{gap}_i = \{g_i, d_i\}$  denoting the position  $bp_i$  of gap  $i$  and  $d_i$  denoting the dummy that is used to fill gap  $i$ .

The elements of the mended time series are the elements of the original series, with values  $d_i$  substituted for the missing values at positions  $g_i$  and the sum of relevant dummies  $\sum_{i \text{ with } bp_i \geq t} d_i$  added to each element  $x_t$ .

This loads the package

```
In[1]:= << Mend`
```

Here is a series with 52 elements - the US unemployment rate from 1951 through 2002:

```
In[2]:= x = {0.033054286, 0.030252121, 0.029335036, 0.055997854, 0.043621372,
            0.041253782, 0.042983567, 0.068377089, 0.054576795, 0.055616434, 0.066794496,
            0.055476952, 0.056434263, 0.051669067, 0.045061629, 0.03784539, 0.038421758,
            0.035540616, 0.03506801, 0.049849379, 0.059515226, 0.056036976, 0.048756301,
            0.056246608, 0.084670976, 0.076943478, 0.070384152, 0.060519286, 0.058453689,
            0.071705897, 0.076156094, 0.097192669, 0.095894813, 0.075125038, 0.071997945,
            0.069964438, 0.061855025, 0.055038601, 0.05267365, 0.056103504, 0.068378853,
            0.07502916, 0.069099691, 0.060859999, 0.055979347, 0.053982948, 0.049366231,
            0.045059882, 0.042198238, 0.043358661, 0.047586757, 0.057901872};
```

We drop observations 20 through 25 to produce a time series with gaps at {20,21,22,23,24,25}:

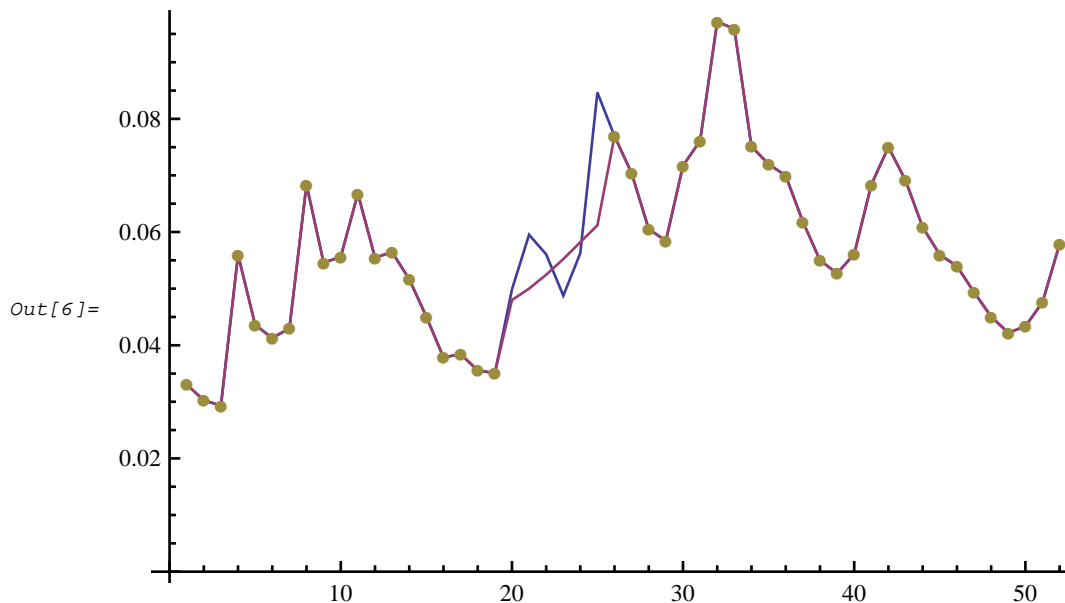
```
In[3]:= y = x;
        y[[20 ;; 25]] = { , , , , , };
```

The following code mends the time series y:

```
In[5]:= {z, {breaks, gaps}} = Mend[y, 100];
```

Here is a plot of the original time series  $x$  (blue), the incomplete time series  $y$  (dots) and the mended time series  $z$  (red).

```
In[6]:= ListPlot[{x, z, y}, Joined -> {True, True, False}, PlotMarkers -> { , { , { , 7}}}]
```

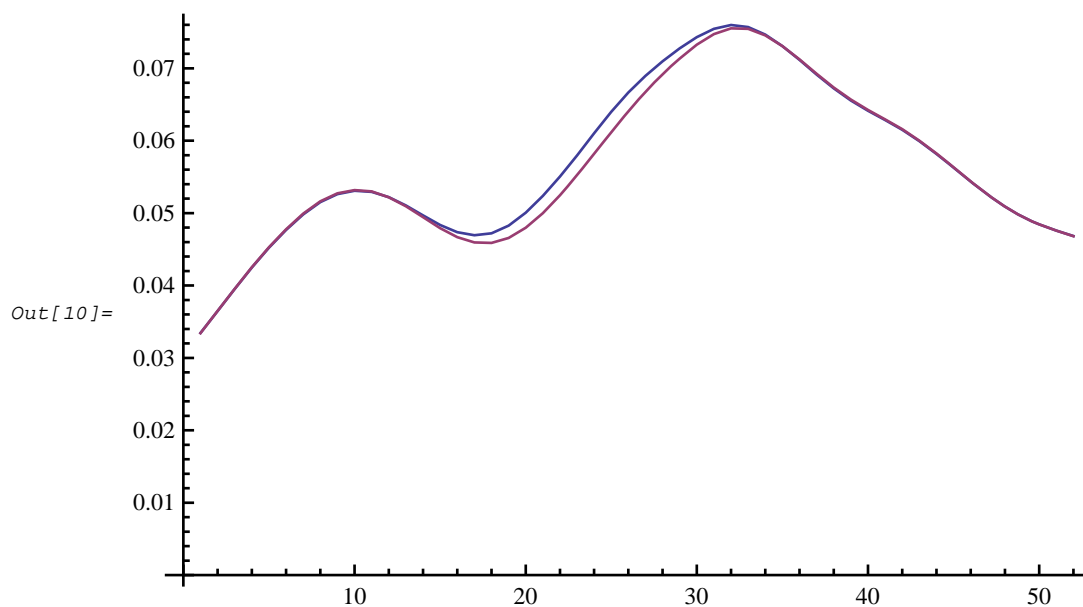


As compared to the trend of the complete series, the trend looks quite similar, though, even with such a considerable gap.

(This uses the package `HPFilter` from the Wolfram Library Archive.)

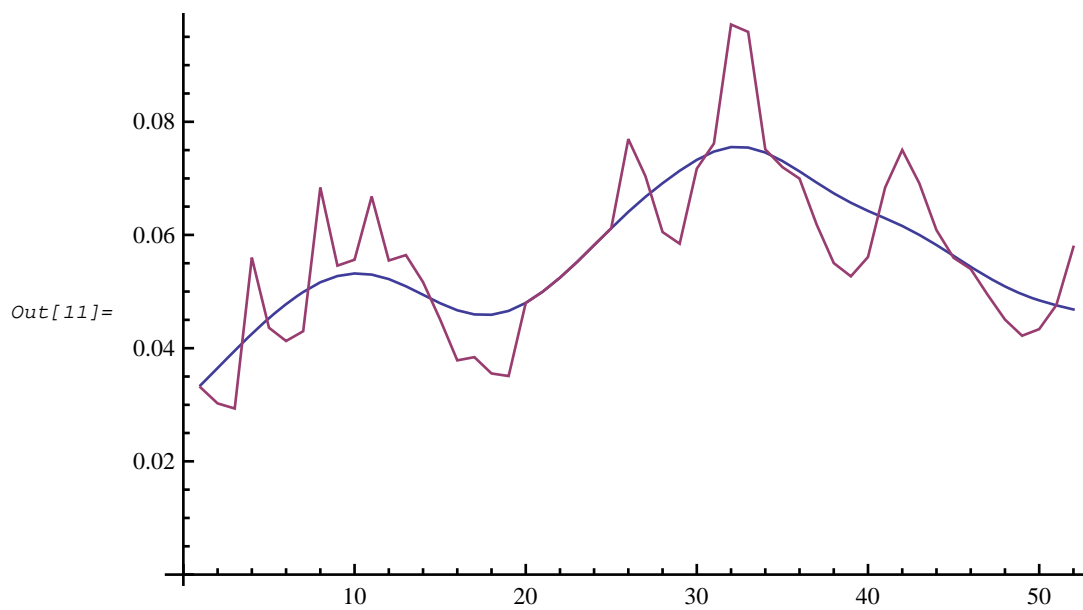
```
In[7]:= << HPFilter`
```

```
In[8]:= yf = HPFilter[x, 100];  
zf = HPFilter[Mend[z, 100][[1]], 100];  
ListLinePlot[{yf, zf}]
```



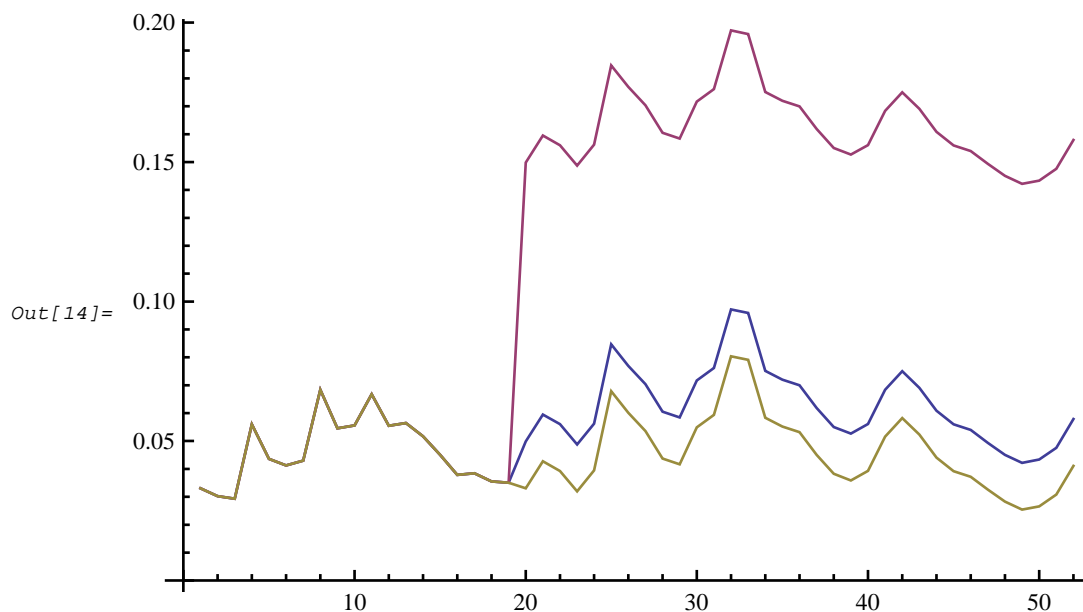
The mended values at positions 20 through 25 are given by the trend:

```
In[11]:= ListLinePlot[{zf, z}]
```



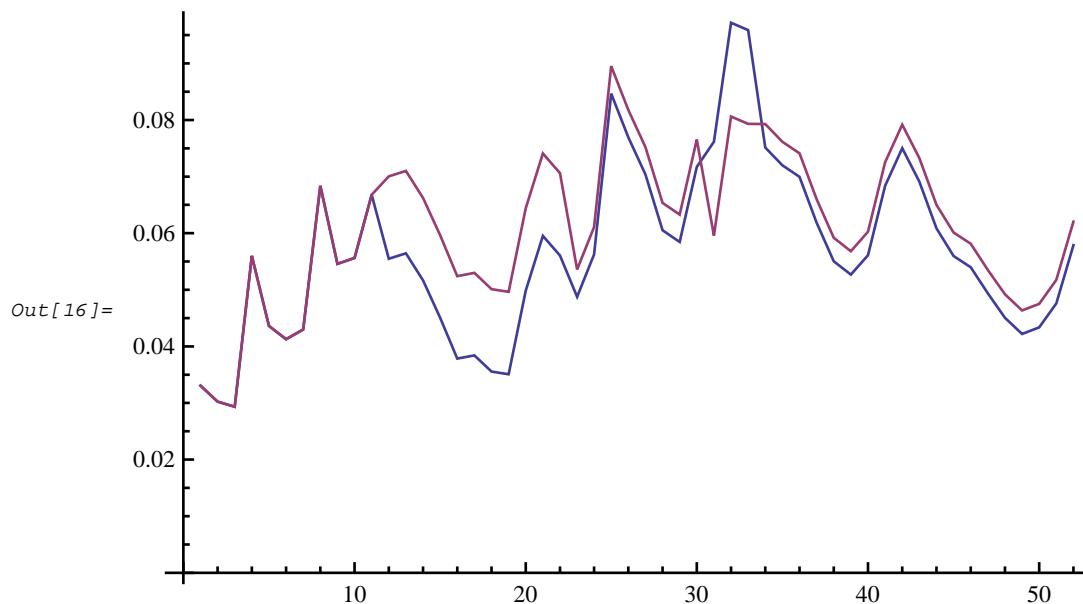
Here we have the original time series (blue). We introduce a structural break of  $+0.1$  at  $t=20$  (red) and correct for it (green). In this example, the structural break is over-corrected:

```
In[12]:= y = x;
y[[20 ;; Length[x]]] = x[[20 ;; Length[x]]] + .1;
{z, {breaks, gaps}} = Mend[y, 100, {20}];
ListLinePlot[{x, y, z}, AxesOrigin -> {0, 0}]
```



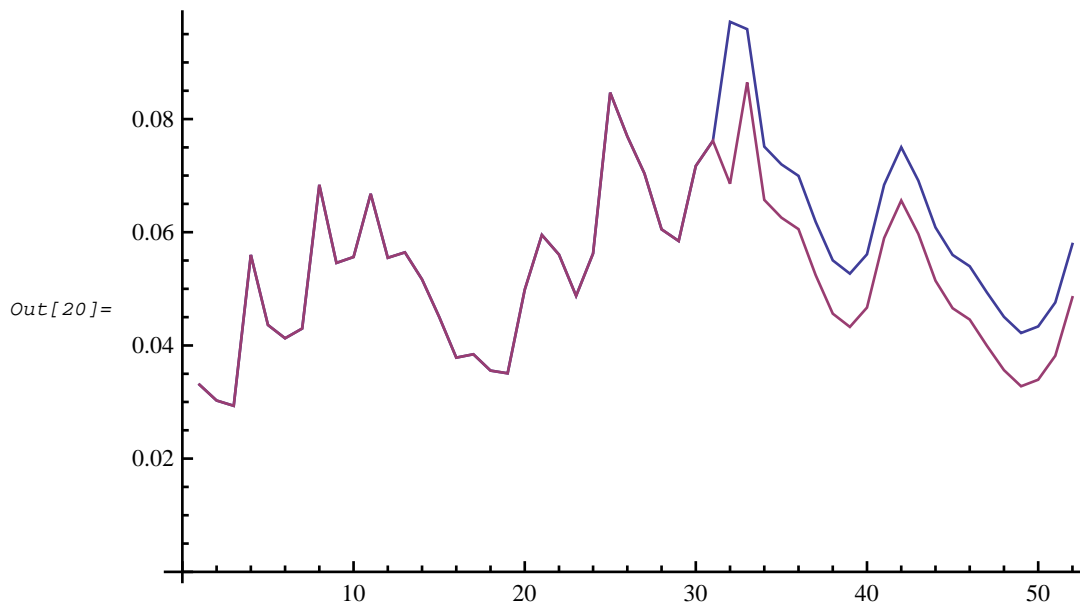
You can have several break points:

```
In[15]:= {y, {breaks, gaps}} = Mend[x, 100, {12, 23, 31, 34}];
ListLinePlot[{x, y}]
```



You can have a gap and a break at the same point in time:

```
In[17]:= y = x;
y[[32]] = Null;
{z, {breaks, gaps}} = Mend[y, 100, {32}];
ListLinePlot[{x, z}]
```



Too few data and too many gaps produce sundry error messages

```
In[21]:= y = {0.033054286, 0.030252121, 0.029335036, , 0.043621372};
Mend[y, 100, {4, 5}];
```

Inverse::sing: Matrix

```
{{0.298713, -0.00157827, 0.300291}, {-0.00157827, <<20>>, -0.398019}, {0.300291, -
0.398019, 0.69831}} is singular. >>
```

Inverse::sing: Matrix

```
{{0.298713, -0.00157827, 0.300291}, {-0.00157827, <<20>>, -0.398019}, {0.300291, -
0.398019, 0.69831}} is singular. >>
```

Mend::singular: Singularity encountered, probably due to  
an overlap or bad combination of undefined data points and breaks.

Bad combinations of breaks and gaps may lead to similar error messages:

```
In[23]:= y = x;
        y[[1 ;; 3]] = Null;
        Mend[y, 100, {1, 2, 3}];
```

Inverse::luc: Result for Inverse of badly conditioned matrix

$$\{\{-4.45247 \times 10^{-13}, -4.36217 \times 10^{-13}, -4.2595 \times 10^{-13}, -9.02987 \times 10^{-15}, -1.02663 \times 10^{-14}, -1.15534 \times 10^{-14}\}, \ll 4 \gg, \{\ll 23 \gg, \ll 4 \gg, \ll 19 \gg\}\}$$

may contain significant numerical errors. >>

Mend::singular: Singularity encountered, probably due to  
an overlap or bad combination of undefined data points and breaks.

*GuessBreak[x, α] returns a number that indicates the position of a possible structural break. It requires as an input the time series and the smoothing parameter α, as in Mend[ ] above.*

We introduce a break of +.01 at t=20 and guess it:

```
In[26]:= y = x;
        y[[20 ;; Length[x]]] = x[[20 ;; Length[x]]] + .1;
        GuessBreak[y]
```

```
Out[28]= 20
```

*Criterion[x, α] mends possible gaps in the time series x and returns the value of the criterion that underlies the mending procedure. It requires as an input the time series and the smoothing parameter α, as in Mend[ ] above.*

*For the definition and interpretation of the criterion, see Schlicht (2008), equation 2.6.*

Call the criterion:

```
In[29]:= Criterion[x, 100]
```

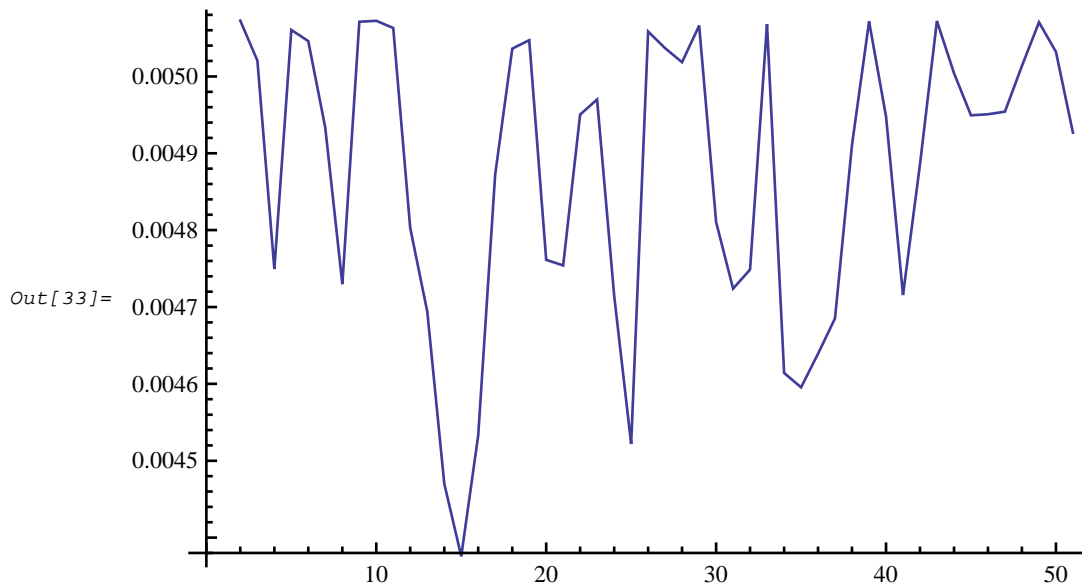
```
Out[29]= 0.0050727
```

You may use the criterion to evaluate different values for filling in gaps and breaks.

The following depicts how the criterion assumes different values for a break inserted at subsequent break points  $t=2, 3, \dots, T$ . It can be seen that the criterion is improved (reduced) most pronouncedly at  $t=15$ .

`In[30]:=`

```
T = Length[x]; crit = {Null};
Do[
  crit = Append[crit, Criterion[Mend[x, 100, {k}][[1]]],
  {k, 2, T - 1}];
crit = Append[crit, Null];
ListLinePlot[{crit}]
```



GuessBreak returns the position of the minimum in this series:

`In[34]:=`

```
GuessBreak[x]
```

`Out[34]= 15`

## References

Ludsteck, Johannes (2005).: "HPFilter, a *Mathematica* Package",  
freely available at  
<http://library.wolfram.com/infocenter/MathSource/5161/>

Schlicht, Ekkehart (2005): "Estimating the Smoothing Parameter in the So-called Hodrick-Prescott Filter",  
*Journal of the Japan Statistical Society*  
Vol. 35 (2005), No. 1, pages 99-119,  
freely available at

<http://www.scipress.org/journals/jjss/pdf/3501/35010099.pdf>

Schlicht, Ekkehart (2008): "Trend Extraction from Time Series with Structural Breaks and Missing Observations",

*Journal of the Japan Statistical Society*

Vol. 38 (2008), No. 2, pages 285-292,

freely available at

[http://www.jstage.jst.go.jp/article/jjss/38/2/285/\\_pdf](http://www.jstage.jst.go.jp/article/jjss/38/2/285/_pdf).

### Acknowledgement

I thank Johannes Ludsteck for helping me with expert advice in preparing this package.

Ekkehart Schlicht,  
Department of Economics  
University of Munich,  
Germany  
[www.lrz.de/~ekkehart](http://www.lrz.de/~ekkehart)

May 2011