Mixed Frequency Data Sampling Regression Models: the R Package midasr

Virmantas Kvedaras
Vilnius University

Vaidotas Zemlys
Vilnius University

Abstract

The implementation of MIDAS approach in the R package **midasr** is described.

Keywords: MIDAS, specification test.

1. Introduction

In econometric applications it is common to encounter time series which are sampled in different frequencies, e.g. quarterly vs yearly, etc. In order to use differently sampled series in regression analysis one of the series is usually aggregated. It is evident that some of the information is lost during such transformation. One of the solutions to this problem is the mixed data sampling (MIDAS) approach introduced in Ghysels and Valkanov (2004). It has gained popularity in financial and some macroeconomic applications (see e.g. Foroni, Marcellino, and Schumacher 2011, and Sinko, Sockin, and Ghysels 2012, for a recent overview of various contributions). In the most cases, it is used for the forecasting purposes.

The main idea of MIDAS approach is based on observation that aggregation of high frequency time series is actualy a specific embedding of the high frequency domain to the low frequency domain. Say we have yearly series Y_t and quaterly series x_τ and we want to estimate the model

$$Y_t = f(x_\tau) + \varepsilon_t$$

The usual approach is to aggregate x_τ to a yearly sampling frequency:

$$X_t = \frac{1}{4}(x_{4t} + x_{4t-1} + x_{4t-2} + x_{4t-3}),$$

where we assume that yearly time series are observed at the same time as the fourth quarter of the quarterly time series. Then we can rewrite the model in the following form:

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

If we substitute the aggregation equation we see that this model is a restricted form of the more general model:

$$Y_t = \alpha + w_1 x_{4t} + w_2 x_{4t-1} + w_3 x_{x4t-2} + w_4 x_{4t-3} + \varepsilon_t$$

We can extend this model for general frequency ratio m and including more lags resulting in so called U-MIDAS model:

$$Y_t = \alpha + \sum_{h=0}^{k} w_h x_{tm-h} + \varepsilon_t.$$

If frequency ratio m is high, such model might not be feasible to estimate, due to lack of degrees of freedom. To solve this problem MIDAS approach suggests restricting the weights:

$$w_h = g(h, \lambda), \quad , h = 0, 1, ..., k$$

where g is some function and λ is a vector of hyper-parameters. In MIDAS literature function g is usually chosen from a fixed set of functions. The important question is then whether this chosen function is the correct one. Recently Kvedaras and Zemlys Kvedaras and Zemlys (2012) proposed a test which lets to test hypothesis whether the chosen weight function is appropriate.

Package **midasr** is aimed at applied researcher. It allows to estimate the MIDAS regression model and test its feasibility.

2. Theory

2.1. Simple MIDAS model

Consider a situation where we observe two processes $y = \{y_t \in \mathbb{R},\ t = 0, \pm1, \pm2 \dots\}$ and $x = \{x_\tau \in \mathbb{R},\ \tau = 0, \pm1, \pm2 \dots\}$ at different frequencies. For each single low frequency observation m high-frequency observations are available.

The MIDAS regression of y on x has the following representation

$$y_t = \sum_{h=0}^{d} \beta_h x_{tm-h} + \varepsilon_t. \tag{1}$$

The coefficients β_h are usually constrained by a functional constraint:

$$\beta_h = g(\boldsymbol{\lambda}, h), h = 0, ..., d$$

where $\boldsymbol{\lambda}$ is a vector of hyper-parameters.

3. Implementation in midasr package

3.1. Data formats

In order to estimate MIDAS regression high frequency regressor is embedded into low frequency domain. This is done via `fmls` function:

```
> library(midasr)
> x <- 1:16
> fmls(x,3,4)
```

```
     X.0/m X.1/m X.2/m X.3/m
[1,]     4     3     2     1
[2,]     8     7     6     5
[3,]    12    11    10     9
[4,]    16    15    14    13
```

It is assumed that for each low-frequency observation there are exactly m high-frequency observations.

3.2. Model specification

The specification of MIDAS regression is done via usual `formula` interface with the help of the function `fmls`. For MIDAS regression we need to specify the frequency ratio m, the number of lags and the weight function. Since optimisation is used to estimate the coefficients, the starting values must be supplied.

```
midas_r(y~fmls(x,11,12,nealmon),start=list(x=c(0,0,0)))
```

The first argument of weight function must be the vector of hyper-parameters. The second argument must be the number of the coefficients the weight function returns. The number of the coefficients is the number of lags plus one so it is already defined in formula specification. Accordingly it is passed to weight function during optimisation. The additional arguments if there are any must be listed into call to `fmls`.

Set up the data and load all necessary packages

```
> library(midasr)
> data("USunempr")
> data("USrealgdp")
> y <- diff(log(USrealgdp))
> x <- window(diff(USunempr), start = 1949)
> midas_r(y~fmls(x,11,12,nealmon),start=list(x=c(0,0,0)))

MIDAS regression model
 model: y ~ fmls(x, 11, 12, nealmon)
(Intercept)          x1          x2          x3
    0.03243    -0.19137    15.62062     1.76868

Function optim was used for fitting
```

References

Foroni C, Marcellino M, Schumacher C (2011). "U-MIDAS: MIDAS regressions with unrestricted lag polynomials." (2011,35). URL http://ideas.repec.org/p/zbw/bubdp1/201135.html.

Ghysels E, Valkanov PSCR (2004). "The MIDAS touch: Mixed data sampling regression models." URL http://http://rady.ucsd.edu/faculty/directory/valkanov/pub/docs/midas-touch.pdf.

Kvedaras V, Zemlys V (2012). "Testing the functional constraints on parameters in regressions with variables of different frequency." *Economics Letters*, **116**(2), 250 – 254. ISSN 0165-1765. doi:10.1016/j.econlet.2012.03.009. URL http://www.sciencedirect.com/science/article/pii/S0165176512000961.

Sinko A, Sockin M, Ghysels E (2012). "Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regression Models." URL http://www.unc.edu/%7Eeghysels/papers/MIDAS_Usersguide_Version3.pdf.

Affiliation:

Vaidotas Zemlys
Department of Econometric Analysis
Faculty of Mathematics and Informatics
Vilnius University
Naugarduko g. 24, Vilnius, Lithuania
E-mail:Vaidotas.Zemlys@mif.vu.lt
URL: http://vzemlys.wordpress.com