Jak połączyć dane o różnej częstotliwości: Mixed-data sampling (MIDAS) regression models

22 January, 2020

Outline

- MIDAS regression model - introduction;
- ANN-MIDAS
- Application examples;
  - GDP prediction;
  - Financial volatility prediction;
- Next steps: intrinsic time;

MIDAS

Assume:

- $y_t \in Y$ is observed once in a period $[t-1; t]$;
- $x_t^{(m)} \in X$ is observed m times in a period $[t-1; t]$;

$$y_t = \beta_0 + \beta_1 B(L^{1/m}, \Theta) x_t^{(m)} + \epsilon_t^{(m)},$$

where:

- $B(L^{1/m}, \Theta) = \sum_{k=0}^{K} B(k, \Theta) L^{k/m}$
- $L^{1/m}$ - lag operator such that $L^{1/m} x_t^{(m)} = x_{(t-1)/m}^{(m)}$

MIDAS

$$y_t = \beta_0 + \beta_1 (\sum_{k=0}^{K} B(k, \Theta) L^{k/m} x_t^{(m)}) + \epsilon_t^{(m)} =$$

$$= \beta_0 + \beta_1 \sum_{k=0}^{K} B(k, \Theta) x_{\frac{t-k}{m}}^{(m)} + \epsilon_t^{(m)}$$

- y_t is observed once in a period $[t-1; t]$;
- $x_t^{(m)}$ is observed m times in a period $[t-1; t]$;

$B(k, \Theta)$ parametrisation

- Exponential Almon Lag

$$B(k, \Theta) = \frac{e^{\Theta_1 k + \ldots + \Theta_Q k^Q}}{\sum_{k=1}^{K} e^{\Theta_1 k + \ldots + \Theta_Q k^Q}}$$

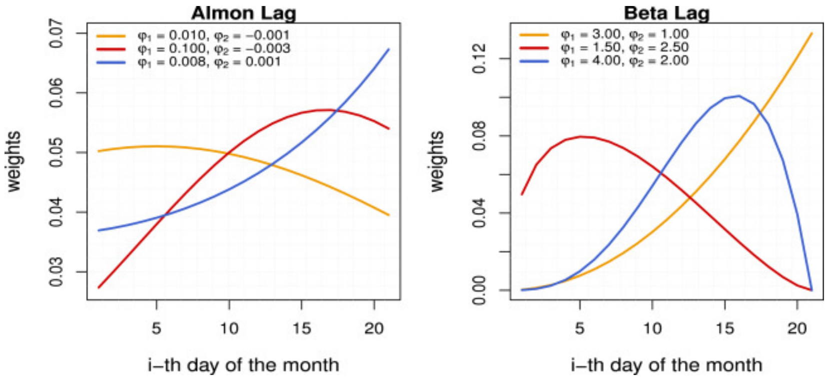
- Beta Lag

$$B(k, \Theta_1, \Theta_2) = \frac{f(\frac{k}{K}, \Theta_1, \Theta_2)}{\sum_{k=1}^{K} f(\frac{k}{K}, \Theta_1, \Theta_2)}$$

where $f(x, a, b) = \dfrac{x^{a-1}(1-x)^b \Gamma(a+b)}{\Gamma(a)\Gamma(b)}$

Ghysels, E., Santa-Clara, P., Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. *CIRANO Working Papers*.

$B(k, \Theta)$ parametrisation



Source: Gorgi, Paolo, Siem Jan Koopman, and Mengheng Li. "Forecasting economic time series using score-driven dynamic models with mixed-data sampling."*International Journal of Forecasting (2019)*.

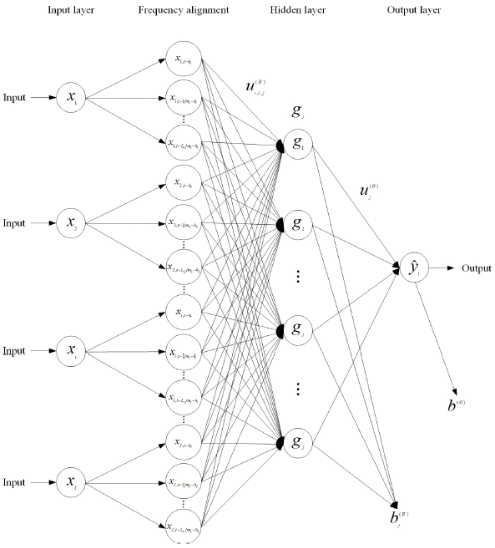
MIDAS

$$y_t = \beta_0 + \beta_1 (\sum_{k=0}^{K} B(k,\Theta) L^{k/m} x_t^{(m)}) + \epsilon_t^{(m)} =$$

$$= \beta_0 + \beta_1 \sum_{k=0}^{K} B(k,\Theta) x_{\frac{t-k}{m}}^{(m)} + \epsilon_t^{(m)}$$

- y_t is observed once in a period $[t-1;t]$;
- $x_t^{(m)}$ is observed m times in a period $[t-1;t]$;

ANN-MIDAS


Input layer
Frequency alignment
Hidden layer
Output layer
Input
Input
Input
Input
Output

ANN-MIDAS

Assumptions

- $\{x_{\tau_i}\}_{i=1}^{I}$ - original input variables observed in high frequency data;
- $\{y_t\}_{t=1}^{T}$ - target output observed in low frequency;
- $\{m_i\}_{i=1}^{I}$ - frequency mismatches between y_t and $\{x_{\tau_i}\}_{i=1}^{I}$

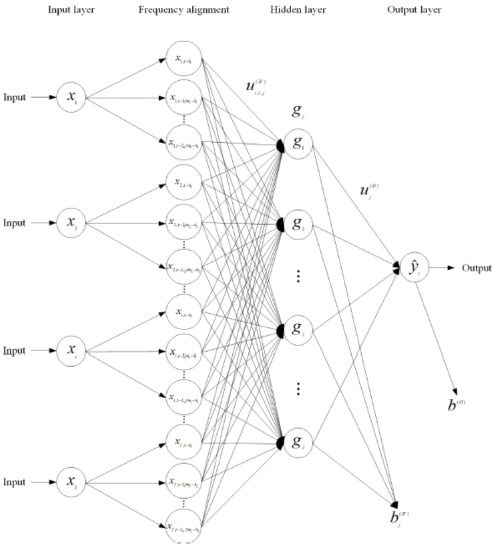
For example $m = 12$ for annual/monthly sampling, $m = 4$ for annual/quarterly sampling

ANN-MIDAS

Step 1 Frequency alignment on each predictor x_{τ_i} to get
$(x_{i,t-h_i}, x_{i,t-1/m_i-h_i}, ..., x_{i,t-L_i/m_i-h_i})$
of the same frequency as the output y_t
where

- $\{L_i\}_{i=1}^{I}$ - maximum lag order
- $h_i(i = 1, 2, ..., I)$ - high frequency forecast horizon related to the release of the high frequency variable x_{τ_i}

ANN-MIDAS


Input layer
Frequency alignment
Hidden layer
Output layer
Input
Input
Input
Input
Output

ANN-MIDAS

Step 2 Hidden layer calculation

$$g_j = f^{(H)}(x_1, x_2, ..., x_I) = f^{(H)}(\sum_{i=1}^{I} u_{i,j}^{(H)} \sum_{l=0}^{L_i} w_i(\delta; l) x_{i,t-l/m_i-h_i} + b_j^{(H)})$$

where w_i Almon lag polynominal is defined as

$$w_i(\delta; l) = \frac{exp(\delta_1 l + \delta_2 l^2)}{\sum_{l=0}^{L_i} exp(\delta_1 l + \delta_2 l^2)}$$

- $u^{(H)}$ - weight vector of the hidden layer;
- $b^{(H)}$ - bias vector of the hidden layer;
- $f^{(H)}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ - activation function;

ANN-MIDAS

Step 3 Output layer calculation and error function calculation

$$\hat{y}_t = f^{(O)}(g1, g2, ..., g_j) = f^{(O)}(\sum_{j=1}^{J} u_j^{(O)} g_j + b^{(O)})$$

- $u^{(O)}, b^{(O)}$ - input vector and bias of an output layer;
- $f^{(O)}$ - output layer activation function (here identity);

$$E = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2$$

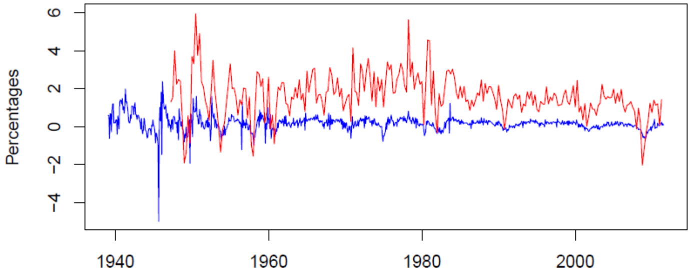
- T - number of low frequency observations

Xu, Q., Zhuo, X., Jiang, C., Liu, Y. (2019). An artificial neural network for mixed frequency data. *Expert Systems with Applications*, 118, 127-139.

Application example: MIDAS GDP prediction

$$y_t = \alpha + \rho y_{t-1} + \sum_{j=3}^{11} \Theta_j x_{3t-j} + \epsilon_t$$

- y_t - log difference of US GDP measured quarterly;
- x_{3t} - log difference of monthly total employment non-farms payroll



Application example: Data alignment

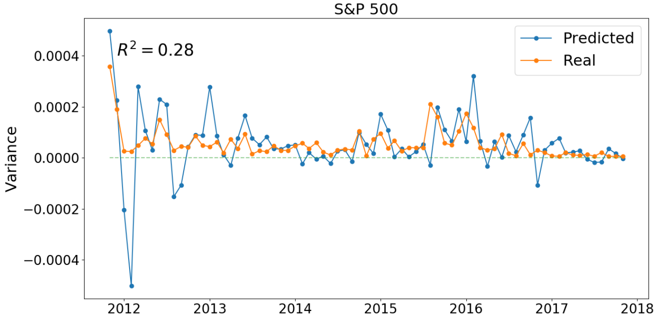
	Month	Year	Value	Index
0	1	1985	0.276393	x3t-11
1	2	1985	0.128584	x3t-10
2	3	1985	0.357919	x3t-9
3	4	1985	0.202185	x3t-8
4	5	1985	0.281963	x3t-7
5	6	1985	0.149919	x3t-6
6	7	1985	0.194764	x3t-5
7	8	1985	0.197452	x3t-4
8	9	1985	0.207262	x3t-3

	x3t-3	x3t-4	x3t-5	...	x3t-9	x3t-10	x3t-11
0	NaN	NaN	NaN	...	NaN	NaN	NaN
1	NaN	NaN	NaN	...	NaN	NaN	NaN
2	NaN	NaN	NaN	...	NaN	NaN	NaN
3	0.207262	0.197452	0.194764	...	0.357919	0.128584	0.276393
4	0.169499	0.212533	0.191565	...	0.149919	0.281963	0.202185
5	0.095057	0.108313	0.126683	...	0.207262	0.197452	0.194764
6	-0.094756	0.128043	0.188835	...	0.169499	0.212533	0.191565
7	0.347833	0.114538	0.320197	...	0.095057	0.108313	0.126683
8	0.204166	0.185605	0.185950	...	-0.094756	0.128043	0.188835

Ghysels, Eric, Virmantas Kvedaras, and Vaidotas Zemlys. "Mixed frequency data sampling regression models: the R package midasr."*Journal of statistical software* (2016): 1-35.

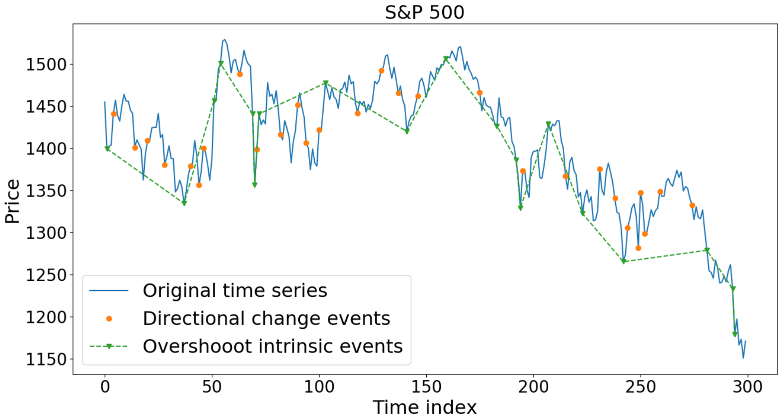
Example 2: Volatility prediction

$$\sigma^2_{t+1} = \sum_{d=0}^{252} r^2_{t+1-d}$$ r - daily return rate σ^2 - monthly return variance



Basing on: Ghysels, E., Santa-Clara, P., Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76(3), 509-548.

Future steps: intrinsic time



Idea: Petrov Vladimir, Anton Golub, and Richard B. Olsen. Agent-Based Model in Directional-Change Intrinsic Time. *Quantitative Finance* (2019): 1-20

Future steps: intrinsic time

	Date	Price	Int_event	Int_index
0	2000-01-03	1454.48	0	0
1	2000-01-04	1399.15	-2	1
2	2000-01-05	1401.87	0	None
3	2000-01-06	1403.60	0	None
4	2000-01-07	1440.45	1	2
5	2000-01-10	1456.74	0	None
6	2000-01-11	1438.51	0	None
7	2000-01-12	1432.00	0	None
8	2000-01-13	1450.01	0	None
9	2000-01-14	1463.72	0	None
10	2000-01-18	1455.70	0	None
11	2000-01-19	1455.53	0	None
12	2000-01-20	1444.98	0	None
13	2000-01-21	1440.86	0	None
14	2000-01-24	1400.89	-1	3
15	2000-01-25	1409.72	0	None
16	2000-01-26	1404.17	0	None
17	2000-01-27	1398.62	0	None
18	2000-01-28	1362.10	0	None
19	2000-01-31	1393.45	0	None
20	2000-02-01	1408.92	1	4
21	2000-02-02	1409.61	0	None

Data Source: Heber G, Lunde A, Shephard N, Sheppard K (2009). "Oxford-Man Institute's Realized Library." Oxford-Man Institute, University of Oxford. Library Version 0.3.

Thank you!

Bibliography

- Heber G, Lunde A, Shephard N, Sheppard K (2009). "Oxford-Man Institute's Realized Library." Oxford-Man Institute, University of Oxford. Library Version 0.3.
- Ghysels, E., Santa-Clara, P., Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. *CIRANO Working Papers*.
- Ghysels, E., Santa-Clara, P., Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76(3), 509-548.
- Ghysels, Eric, Virmantas Kvedaras, and Vaidotas Zemlys. "Mixed frequency data sampling regression models: the R package midasr."*Journal of statistical software* (2016): 1-35.

Bibliography

- Gorgi, Paolo, Siem Jan Koopman, and Mengheng Li. "Forecasting economic time series using score-driven dynamic models with mixed-data sampling."*International Journal of Forecasting (2019)*.
- Petrov Vladimir, Anton Golub, and Richard B. Olsen. Agent-Based Model in Directional-Change Intrinsic Time. *Quantitative Finance* (2019): 1-20
- Xu, Q., Zhuo, X., Jiang, C., Liu, Y. (2019). An artificial neural network for mixed frequency data. *Expert Systems with Applications*, 118, 127-139.