模型理论和简单实现

WHH
jinlin@zuel.edu.cn

中南财经政法大学统计与数学学院

2020 年 3 月

1 模型理论

2 模型实现

todo

- 滞后阶数如何选择，根据现有文献有的直接给定，有的根据 AIC,BIC 确定
- 模型估计再详细一点，写出式 (14) 中具体的待估参数有哪些
- RV 的滚动窗口，参考[1]

1 模型理论

- GARCH-MIDAS
- DCC-MIDAS 模型

2 模型实现

1 模型理论

- GARCH-MIDAS
- DCC-MIDAS 模型

在有关金融市场波动率的研究中，若想要探究经济因素对金融市场的影响，则传统的 GARCH 类模型略显不足，它只能处理同频数据。Ghysels 等引入了混频抽样模型 (Mixed Data Sampling), 简称 MIDAS, 后由 Engel 等对其进行扩展，将其引入 GARCH, 形成 GARCH-MIDAS 模型, 该模型将波动率分为短期波动率和长期波动率，长期波动率的设定中可以引入与研究对象不同频率的外生变量，这极大地促进了金融市场混频数据的研究[1]。为了方便 GARCH-MIDAS 模型的表述，下面首先对传统的基于同频数据的 GARCH(1,1) 作简单介绍。

$$
\begin{cases}
r_t = \mu_t + \varepsilon_t \\
\varepsilon_t = \sqrt{h_t} e_t \\
h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}
\end{cases}
\tag{1}
$$

式 (1) 中，r_t 表示时刻 t 的收益率，h_t 表示条件方差，e_t 是均值为 0，方差相等的独立同分布随机变量序列。第一个方程为均值方程，用来描述收益率序列条件均值的变化过程，第三个方程为方差方程，用来描述收益率序列条件方差的变化过程。

现对 GARCH-MIDAS 模型进行介绍。

$$r_{i,t} = \mu_t + \varepsilon_{i,t} \tag{2}$$

$$\frac{\varepsilon_{i,t}}{\sqrt{\tau_t}} = \sqrt{g_{i,t}} e_{i,t} \tag{3}$$

$$\sigma_{i,t}^2 = \tau_t g_{i,t} \tag{4}$$

其中，r_{it} 表示第 t 个周期 (周, 月或者季度等低频周期) 第 i 天 (天, 分钟等高频) 的收益率，μ_t 表示收益率的条件均值，$\Phi_{i-1,t}$ 表示收益率在第 t 周期第 i 天之前的所有信息集，g_{it} 和 τ_t 分别表示短期收益率和长期收益率，式 (4) 给出了收益率的条件方差：$\sigma_{i,t}^2 = Var(\varepsilon_{i,t}|\Phi_{i-1,t})$，即波动率的设定形式，表示波动率由短期波动成分和长期波动成分乘积而得。尽管 $g_{i,t}$ 每天都在变化，但是 τ_t 在时段 t 内的所有天都是恒定的，因此仅以较低的频率变化。

短期波动成分

短期波动率成分用来描述众所周知的波动率聚集性，并假定遵循 GARCH(1,1) 的过程：

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{\varepsilon_{i-1,t}^2}{\tau_t} + \beta g_{i-1,t} \tag{5}$$

可见，相比于传统的同频 GARCH(1,1) 过程，GARCH-MIDAS 中 GARCH(1,1) 的特别之处就在于 $\frac{\varepsilon_{i-1,t}^2}{\tau_t}$。另外，短期波动率成分的设定还可能是 GARCH 族的其他形式，如用来描述非对称效应的 GJR-GARCH（1,1）过程

$$g_{i,t} = (1 - \alpha - \gamma/2 - \beta) + \left(\alpha + \gamma \mathbb{1}_{\{\varepsilon_{i-1,t}<0\}}\right) \frac{\varepsilon_{i-1,t}^2}{\tau_t} + \beta g_{i-1,t} \tag{6}$$

其中 $\mathbb{1}_{\{\varepsilon_{i-1,t}<0\}}$ 是示性函数，当 () 内容成立时，取值为 1，否则为 0. γ 是非对称杠杆系数。下面对 GJR-GARCH 中涉及 $e_{i,t}$ 和短期波动成分中参数的假设作出介绍。

假设一: e_{it} 满足独立同分布条件，且 $E(e_{it})=0, E(e_{it}^2)=1$, 对于 $\kappa=E(e_{it}^4)$ 时，有 $1<\kappa<\infty$, 即标准残差序列的四阶矩存在。

假设二: $\alpha>0$，$\alpha+\gamma>0$，$\beta\geq 0$，$\alpha+\gamma/2+\beta<1$，同时这些参数满足 $(\alpha+\gamma/2)^2\kappa+2(\alpha+\gamma/2)\beta+\beta^2<1$. GIR-GARCH 中 $g_{i,t}$ 的一阶矩和二阶矩分别是 $E(g_{i,t}=1)$,

$$E(g_{i,t}^2)=\frac{1-(\alpha+\gamma/2+\beta)^2}{1-((\alpha+\gamma/2)^2\kappa+2(\alpha+\gamma/2)\beta+\beta^2)}$$

.

长期波动成分

长期波动成分 τ_t 被设定为解释变量及其滞后 K$(K \geq 1)$ 阶的函数，且 $f(\cdot) > 0$：

$$\tau_t = f(X_T, X_{T-1}, \cdots, X_{t-K}) \tag{7}$$

式 (7) 中，X_t 被假设为满足严格平稳条件，并且独立于模型的 GARCH 部分，因此，τ_t 也是平稳序列且独立于模型的 GARCH $g_{i,t-j}e^2_{i,t-j}$ 部分[2]。在实证研究中，$f(\cdot) > 0$ 通常被设定为解释变量 X_t 的线性形式:

$$\tau_t = m + \pi_1 X_{T-1} + \cdots, \pi_K X_{t-K} \tag{8}$$

式 (8) 中, 为了保证 $f(\cdot) > 0$, 需满足对于所有的 $l = 1, \cdots, K$,$m > 0, \pi_l \geq 0$, 且 X_t 也是非负变量。如果 X_t 不是非负变量，那么可以选择指数形式:

$$\tau_t = exp(m + \pi_1 X_{T-1} + \cdots, \pi_K X_{t-K}) \tag{9}$$

同样对于所有的 t, i, j, 要求 X_t 独立于 $e_{i,t-j}$ ，但对 $f(\cdot)$ 没有限制。

下面给出常用来衡量长期波动成分贡献度的指标 VR，通过方差比的形式给出:

$$VR = \frac{Var(log(\tau_t))}{Var(log(\tau_t g_t))} \tag{10}$$

其中，$g_t = \sum_{i=1}^{N_t} g_{i,t}$. 这个比率表述了条件方差的总方差中被长期波动成分的方差解释的比值。

至此, 式 (2),(3),(5),(7) 构成 GARCH-MIDAS 模型。其中，短期波动成分和长期波动成分都有多种设定。

根据现有文献，对长期波动成分的设定主要有以下两种形式。(要列出具体的参考文献) 一种是参考 Merton(1980),Schewert(1989) 等人的做法，用一定时间区间 (一个月或一个季度) 的可实现波动率来衡量长期波动，如月度已实现波动率 RV_t。大多学者通过 MIDAS 回归来平滑 RV_t。

$$\tau_t = m + \theta \sum_{k=1}^{K} \varphi_k\left(\omega_1, \omega_2\right) RV_{t-k}$$

$$RV_t = \sum_{i=1}^{N_t} r_{i,t}^2 \tag{11}$$

τ_t 也满足如下形式:

$$E_{t-1}\left[\left(r_{i,t}-\mu\right)^2\right] = \tau_t E_{t-1}\left(g_{i,t}\right) = \tau_t \tag{12}$$

式 (12) 中短期波动成分的均值 $E_{t-1}(g_{i,t})$ 等于其无条件期望，即 $E_{t-1}(g_{i,t}) = 1$.

为了完成此模型设定，还需对长期波动成分，式 (11) 的权重系数进行说明:

$$\varphi_k(\omega)=\begin{cases}\frac{(k/K)^{\omega_1-1}(1-k/K)^{\omega_2-1}}{\sum_{j=1}^{K}(j/K)^{\omega_1-1}(1-j/K)^{\omega_2-1}} & \text{Beta}\\ \omega^k/\left(\sum_{j=1}^{K}\omega^j\right) & \text{Exp. Weighted}\end{cases} \tag{13}$$

所有权重系数的和为 1。式中，基于 Beta 函数的 Beta 滞后非常灵活，可以用于多种滞后结构，既能表示权重系数的单调增或单调减，也能表示多峰型的权重方案[1]。

另外一种是在长期波动成分中包含有多个解释变量。常以已实现波动率和其他同频变量的线性组合为主，如夏婷用同时包含已实现波动率和工业增长率或通货膨胀率等宏观经济变量的长期波动成分来刻画我国股市的波动率[3]，具体如下:

$$\tau_t = m + \theta_1 \sum_{k=1}^{K} \varphi_{1k}\left(\omega_{11}, \omega_{12}\right) RV_{t-k} + \theta_2 \sum_{k=1}^{K} \varphi_{2k}\left(\omega_{21}, \omega_{22}\right) X_{t-k}$$

最后根据收益率的分布形式和模型设定，用极大似然法进行估计，极大似然函数为：

$$LLF = -\frac{1}{2} \sum_{t=1}^{T} \left[\log g_t(\Phi) \tau_t(\Phi) + \frac{\left(r_t - \mu_t\right)^2}{g_t(\Phi) \tau_t(\Phi)} \right] \tag{14}$$

1 模型理论

- GARCH-MIDAS
- DCC-MIDAS 模型

单变量 DCC-MIDAS

DCC-MIDAS 模型由 Colacito,Engle 和 Ghysel 于 2011 年提出，它将 GARCH-MIDAS 模型和 DCC 结合起来, 用于通过混频数据提取长期相关性成分[4]。下面对该模型作出介绍。

假设

$$\mathbf{r}_t \sim_{\text{i.i.d.}} N(\mu, H_t)$$
$$H_t = D_t R_t D_t$$

其中，$\mathbf{r}_t$ 是一个收益率向量，$\mathbf{r}_t = [r_{1,t}, \cdots, r_{n,t}]$，$\mu$ 是无条件方差向量，H_t 是条件方差矩阵，D_t 是对角矩阵，其对角线上元素是每个收益率的标准差，R_t 是条件相关矩阵，$R_t = E_{t-1}[\xi_t \xi_t']$，$\xi_t = D_t^{-1}(\mathbf{r}_t - \mu)$. 因此 $\mathbf{r}_t = \mu + H_t^{1/2}\xi_t$，$\xi_t \sim_{\text{i.i.d.}} N(0, I_n)$

具体来说，对于每个资产 $i=1,\cdots,n$，单变量收益率的 GARCH-MIDAS 过程为:

$$r_{i,t}=\mu_i+\sqrt{m_{i,\tau}\cdot g_{i,t}}\xi_{i,t},\forall t=\tau N_v^i,\ldots,(\tau+1)N_v^i$$

其中，\$$\tau$ 表示第 τ 个周期，因此 t 从 τN_v^i 开始，$g_{i,t}$ 遵循 GARCH(1,1) 过程:

$$g_{i,t}=(1-\alpha_i-\beta_i)+\alpha_i\frac{(r_{i,t-1}-\mu_i)^2}{m_{i,\tau}}+\beta_i g_{i,t-1}$$

$m_{i,\tau}$ 是一定时间段内已实现波动率 K_v^i 阶滞后变量的加权和

$$m_{i,\tau}=\bar{m}_i+\theta_i\sum_{l=1}^{K_\nu^i}\varphi_l\left(\omega_v^i\right)RV_{i,\tau-l}$$

其中，$RV_{i,\tau}=\sum_{j=(\tau-1)N_v^i+1}^{\tau N_v^i}\left(r_{i,j}\right)^2$，$N_v^i$ 可以是一个月或一个季度。权重系数 φ 通过 Beta 函数给定:

$$\varphi_l\left(\omega_v^i\right)=\frac{\left(1-\frac{1}{K_i^i}\right)^{\omega_v^i-1}}{\sum_{j=1}^{K_v^i}\left(1-\frac{j}{K_v^i}\right)^{\omega_v^i-1}} \tag{15}$$

式 (15) 的权重函数形式与式 (13) 中的 Beta 函数是同一类型，当 (13) 中的 Beta 函数满足 $\omega_1=\omega_2$ 时, 式 (13) 就变为式 (15)。

在长期波动成分，$m_{i,\tau}$ 既可以保持局部不变，又可以基于局部移动窗口来表示。但 Engle(2006) 等人发现这两者之间的差异可以忽略不计。考虑到局部移动窗口有更大的适用性，于是采用局部移动窗口的形式来描述长期波动成分。也就是说，使用标准残差 $\xi_{i,t}$ ，可以得到一个矩阵 Q_t ，他的元素是:

$$q_{i,j,t} = \bar{\rho}_{i,j,t}(1-a-b) + a\xi_{i,t-1}\xi_{j,t-1} + bq_{i,j,t-1} \tag{16}$$

$$\bar{\rho}_{i,j,t} = \sum_{l=1}^{K_c^{ij}} \varphi_l\left(\omega_r^{ij}\right) c_{i,j,t-l}$$

$$c_{i,j,t} = \frac{\sum_{k=t-N_c^{ij}}^{t} \xi_{i,k}\xi_{j,k}}{\sqrt{\sum_{k=t-N_c^{ij}}^{t} \xi_{i,k}^2}\sqrt{\sum_{k=t-N_c^{ij}}^{t} \xi_{j,k}^2}}$$

$$\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}}\sqrt{q_{j,j,t}}} \tag{17}$$

其中，$q_{i,j,t}$ 是资产 i 和资产 j 的短期相关系数,$\rho_{\bar{i},j,t}$ 是长期相关系数。可以看出，短期相关系数和长期相关系数都是 GARCH-MIDAS 部分标准残差序列的函数。式 (16) 可以重写成:

$$q_{i,j,t} - \bar{\rho}_{i,j,t} = a\left(\xi_{i,t-1}\xi_{j,t-1} - \bar{\rho}_{i,j,t}\right) + b\left(q_{i,j,t-1} - \bar{\rho}_{i,j,t}\right)$$

这表示短期相关系数围绕着长期相关系数波动。DCC-MIDAS 模型的思想与 GARCH-MIDAS 模型类似。在 GARCH-MIDAS 中，提取了波动的两个组成成分，一个涉及短期波动，一个涉及长期波动。短期波动成分基于每日的回报率，它围绕由一定时间段内（一个月或一个季度）的已实现波动率驱动的长期波动成分上下浮动。并且同样可以将长期相关系数与宏观变量联系起来。

$$
\begin{aligned}
QL(\Phi,\Xi) &= QL_1(\Phi) + QL_2(\Phi,\Xi) \\
&\equiv -\sum_{t=1}^{T}\left(n\log(2\pi) + 2\log|D_t| + r_t' D_t^{-2} r_t\right) - \sum_{t=1}^{T}\left(\log|R_t| + \xi_t' R_t^{-1}\xi_t + \xi_t'\xi_t\right)
\end{aligned}
$$

DCC-MIDAS 的估计采用两步法，Φ 中包含了 GARCH–MIDAS 部分的待估参数 $(\mu,\alpha,\beta,\omega,m,\theta)$,Ξ 包含了 DCC-MIDAS 中的待估参数 (a,b,ω_r) 。具体计算时，第一步估计 Φ ，然后再利用所估计的系数计算标准残差序列，得到标准残差序列 ξ ，估计参数 Ξ 。重点是 $\omega_{i,j}$ ，$k_{i,j}$ 以及 $N_c^{i,j}$ 如何选择。

多变量 DCC-MIDAS

与 GARCH-MIDAS 模型类似，在 DCC-MIDAS 的长期相关系数中可以同时引入多个变量，如张宗新在研究金融市场的流动性时，同时考虑了已实现波动率和经济政策不确定性两个因素的影响[5]。首先在 GARCH-MIDAS 的长期波动成分中引入两个变量，

$$m_{i,\tau}=\bar{m}_i+\theta_{i,1}\sum_{l=1}^{K_\nu^i}\varphi_l\left(\omega_{1v}^i\right)RV_{i,\tau-l}+\theta_{i,2}\sum_{l=1}^{K_\nu^i}\varphi_l\left(\omega_{2v}^i\right)X_{\tau-l}$$

然后对 $\bar{\rho}_{i,j,t}$ 作 Fisher's z 变换，即:

$$\bar{\rho}_{i,j,t} = \frac{\exp\left(2z_{i,j,t}\right) - 1}{\exp\left(2z_{i,j,t}\right) + 1}$$

其中:

$$z_{i,j,t} = m_c + \theta_{c,1} \sum_{l=1}^{K_c^{ij}} \varphi_l\left(\omega_{r1}^{ij}\right) c_{i,j,t-l} + \theta_{c,2} \sum_{l=1}^{K_c^{ij}} \varphi_l\left(\omega_{r2}^{ij}\right) X_{\tau-l}$$

这样就有多个变量来解释长期相关系数的动态变化。

1 模型理论

2 模型实现

- GARCH-MIDAS
- DCC-MIDAS

2 模型实现

- GARCH-MIDAS
- DCC-MIDAS

```r
library(mfGARCH)
```

```
## Warning: package 'mfGARCH' was built under R version 3.5.3
```

```r
#head(df_financial)
#rv是5分钟已实现波动率，return是日对数收益率*100，nfic频率为周，其余为日
head(df_mfgarch)
```

```
##          date      return  open_close rv vix  year_week dhousing   dindpro
## 1 1971-01-04 -1.09111828 -1.09111828 NA  NA 1971-01-03 -3.49404 0.7663291
## 2 1971-01-05  0.71058072  0.71058072 NA  NA 1971-01-03 -3.49404 0.7663291
## 3 1971-01-06  0.59733547  0.59733547 NA  NA 1971-01-03 -3.49404 0.7663291
## 4 1971-01-07  0.03247875  0.03247875 NA  NA 1971-01-03 -3.49404 0.7663291
## 5 1971-01-08 -0.20587860 -0.20587860 NA  NA 1971-01-03 -3.49404 0.7663291
## 6 1971-01-11 -0.22804918 -0.22804918 NA  NA 1971-01-10 -3.49404 0.7663291
##    nai nfci year_month
## 1 0.72 0.42 1971-01-01
## 2 0.72 0.42 1971-01-01
## 3 0.72 0.42 1971-01-01
## 4 0.72 0.42 1971-01-01
## 5 0.72 0.42 1971-01-01
## 6 0.72 0.46 1971-01-01
```

```r
#vix频率是日，NAI,dhousing，dinpro是月
```

```
model1=fit_mfgarch(data=df_financial,y='return',x='nfci',low.freq = ”week”, K = 52)
#默认短期波动是GJR-GARCH(1,1),权重系数是$\omega_1=\omega_2$
plot_weighting_scheme(model1)
#权重系数呈指数下降
str(model1)
plot(model1$g,type='l',ylim=c(0,40))
plot(model1$tau,type='p')
plot(model1$est.weighting)
var(log(model1$tau))
summary(log(model1$tau))
var(log(model1$tau*model1$g)[254:11306])/var(log(model1$tau)[254:11306])
var(log(model1$tau)[254:11306])/var(log(model1$tau*model1$g)[254:11306])
str(model1)
model2=fit_mfgarch(data=df_financial,y='return',x='nfci',low.freq = ”week”, K = 52,weighting ='beta.unrestricted')
model3=fit_mfgarch(data = df_mfgarch, y = ”return”, x = ”nfci”, low.freq = ”year_week”, K = 52,
x.two = ”dindpro”, K.two = 12, low.freq.two = ”year_month”, weighting.two =
”beta.restricted”)
#具有两个不同频率的变量来描述长期波动
```

2 模型实现

- GARCH-MIDAS
- DCC-MIDAS

midas 理论介绍

这个包解决这样一类回归问题

$$y_t - \alpha_1 y_{t-1} - \cdots - \alpha_p y_{t-p} = \sum_{i=0}^{k} \sum_{j=0}^{l_i} \beta_j^{(i)} x_{tm_i-j}^{(i)} + \varepsilon_t \tag{18}$$

其中，要求:

$$\mathrm{E}\left(\varepsilon_t | y_{t-1}, \ldots, y_{t-p}, x_{tm0}^{(0)}, \ldots, x_{tm_0-l_i}^{(0)}, \ldots, x_{tm_k}^{(k)}, \ldots, x_{tm_k-l_k}^{(k)}\right) = 0$$

以及参数限制

$$\beta_j^{(i)} = f_i\left(\gamma_i, j\right), j = 0, \ldots, l_i, \gamma_i = \left(\gamma_1^{(i)}, \ldots, \gamma_{q_i}^{(i)}\right), q_i \in \mathbb{N}$$

这样可以减少待估参数，原来有 $d = p + \sum_{i=0}^{k} l_i$, 现在待估参数有 $q = \sum_{i=0}^{h_i} q_i$

假设 y_t 是季度数据，现用月度数据 x_t 来解释 y_t，对于每个季度有 3 个月，并且当前和上一季度的月度数据具有解释力。也就是说在当前季度 t，用 t 季度观察到的 $x_{3t}, x_{3t-1}, x_{3t-2}$，以及变量 y_{t-1} 和 $x_{3t-1}, x_{3t-1-1}, x_{3t-1-2}$ 来对 y_t 建模。公式 (18) 可以写成矩阵形式

$$\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} \alpha_1 + \begin{bmatrix} x_6 & \dots & x_1 \\ \vdots & \vdots & \vdots \\ x_{3n} & \dots & x_{3n-5} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

通过用矩阵表示法编写模型，将高频变量 x_t 转换为低频矢量 $((x_{3t}, \dots, x_{3t-5}))^T$. 这种变换叫频率对其，要求 $$ 的观测数正好是 $3n$

如果还有一个以周为频率的变量，假设一个月有 4 周，1 个季度有 12 周，则有

$$\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} \alpha_1 + \begin{bmatrix} x_6 & \cdots & x_1 \\ \vdots & \vdots & \vdots \\ x_{3n} & \cdots & x_{3n-5} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_5 \end{bmatrix} + \begin{bmatrix} z_{24} & \cdots & z_1 \\ \vdots & \vdots & \vdots \\ z_{12n} & \cdots & z_{12n-23} \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \vdots \\ \gamma_{23} \end{bmatrix} + \begin{bmatrix} \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

也就是说，通过频率转换，将高频数据 x_t 转换为低频矢量 $\left(x_{tm_i}^{(i)}, x_{tm_i-1}^{(i)}, \ldots, x_{tm_i-l}^{(i)}\right)^\top$

式 (18) 的矩阵表达为:

$$
\begin{bmatrix} y_l \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{l-1} & \cdots & y_{l-p} \\ \vdots & \vdots & \vdots \\ y_{n-1} & \cdots & y_{n-p} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \sum_{i=0}^{k} \boldsymbol{X}^{(i)} \begin{bmatrix} \beta_0^{(i)} \\ \vdots \\ \beta_l^{(i)} \end{bmatrix} + \begin{bmatrix} \varepsilon_l \\ \vdots \\ \varepsilon_n \end{bmatrix}
$$

其中

$$\boldsymbol{X}^{(i)} := \begin{bmatrix} x^{(i)}_{um_i} & x^{(i)}_{um_i-1} & \cdots & x^{(i)}_{um_i-l} \\ x^{(i)}_{(u+1)m_i} & x^{(i)}_{(u+1)m_i-1} & \cdots & x^{(i)}_{(u+1)m_i-l} \\ \vdots & \vdots & \cdots & \vdots \\ x^{(i)}_{tm_i} & x^{(i)}_{tm_i-1} & \cdots & x^{(i)}_{tm_i-l} \\ \vdots & \vdots & \cdots & \vdots \\ x^{(i)}_{(n-1)m_i} & x^{(i)}_{(n-1)m_i-1} & \cdots & x^{(i)}_{(n-1)m_i-l} \\ x^{(i)}_{nm_i} & x^{(i)}_{nm_i-1} & \cdots & x^{(i)}_{nm_i-l} \end{bmatrix} \tag{19}$$

注意 u 是最小的整数，要求 $um_i - l > 0$ 且 $u > p$

对于 MIDAS 模型，需要注意两点，一点是函数约束的合理选择，另一点是最大滞后阶数的选择。解决这两个问题的一种方法是使用某种信息准则，使用样本内或样本外的精度度量，根据参数限制和滞后阶数选择最佳模型。

函数 fmls (x , k , m) 完全执行公式 (19) 中定义的变换，将给定（潜在）高频序列的观测向量 x 转换为 (k + 1) 低频序列的观测矩阵，由最大滞后阶数 k 和频率比 m 定义的（与 k 个滞后同时）。

midas 实现

```r
library(midasr)
```

```
## Loading required package: sandwich
```

```
## Loading required package: optimx
```

```
## Loading required package: quantreg
```

```
## Loading required package: SparseM
```

```
##
## Attaching package: 'SparseM'
```

```
## The following object is masked from 'package:base':
##
##     backsolve
```

```r
x=1:12
fmls(x,k=2,m=3)#数据处理
```

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

```r
set.seed(1001)
n=250
trend=1:n
x=rnorm(4*n)
z=rnorm(12*n)
fn_x <- nealmon(p = c(1, -0.5), d = 8)
fn_z <- nealmon(p = c(2, 0.5, -0.1), d = 17)
y <- 2 + 0.1 * trend + mls(x, 0:7, 4) %*% fn_x + mls(z, 0:16, 12) %*% fn_z + rnorm(n)#数据模拟
```

```r
#对于参数无限制的情况，用OLS估计，
eq_u <- lm(y ~ trend + mls(x, k = 0:7, m = 4) + mls(z, k = 0:16, m = 12))
eq_u <- midas_r(y ~ trend + mls(x, 0:7, 4) + mls(z, 0:16, 12),start = NULL)
#使用指数Almon滞后多项式约束参数，则用NLS估计
eq_r <- midas_r(y ~ trend + mls(x, 0:7, 4, nealmon) + mls(z, 0:16, 12,nealmon), start = list(x = c(1, -0.5), z = c(2, 0.5, -0.1)))
#summary(eq_r)
```

$$y_t = c + \sum_{j=1}^{6} \alpha_j y_{t-j} + \sum_{j=0}^{7} \beta_j x_{4j-1} + \varepsilon_t$$

```r
#含有y滞后项的情况
eq_xy=midas_r(y~mls(y,1:6,1,nealmon)+mls(x,0:7,4,nealmon),start=list(y=c(1,-0.5),x=c(1,-0.5)
    ))
summary(eq_xy)
```

```
##
## MIDAS regression model with "numeric" data:
## Start = 8, End = 250
##
##  Formula y ~ mls(y, 1:6, 1, nealmon) + mls(x, 0:7, 4, nealmon)
##
##  Parameters:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.382272   0.218654   1.748 0.081702 .
## y1           0.993992   0.013264  74.936  < 2e-16 ***
## y2          -0.005845   0.085168  -0.069 0.945341
## x1           1.030485   0.197130   5.227 3.76e-07 ***
## x2          -0.694800   0.208140  -3.338 0.000979 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Residual standard error: 1.383 on 238 degrees of freedom
```

[1] Engle R F 等. STOCK MARKET VOLATILITY AND MACROECONOMIC FUNDAMENTALS[J]. REVIEW OF ECONOMICS AND STATISTICS, ONE ROGERS ST, CAMBRIDGE, MA 02142-1209 USA: MIT PRESS, 95(3): 776–797.

[2] Conrad C, Kleen O. Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models[J]. Journal of Applied Econometrics, Wiley, 2020, 35(1): 19–45.

[3] 夏婷, 闻岳春. 经济不确定性是股市波动的因子吗?——基于 GARCH-MIDAS 模型的分析 [J]. 中国管理科学, 2018, 26(12): 1–11.

[4] Colacito R 等. A component model for dynamic correlations[J]. Journal of Econometrics, Elsevier Science Sa, 2011, 164(1): 45–59.

[5] 张宗新等. 经济政策不确定性如何影响金融市场间的流动性协同运动?——基于中国金融周期的视角 [J]. 统计研究, 2020, 37(02): 37–51.