

主要时间在制作mixed frequency data 的整合PPT

以下为部分成果，更多成果请参见[混频数据处理-共享文件夹/lcx/new/混频模型Presentation](#)

MIDAS模型是在分布式滞后模型的基础上进行改造的

如果被解释变量的 y_t 不仅受同期解释变量 x_t 的影响，而且还明显依赖于 x 的滞后值 $x_t, x_{t-1}, x_{t-2} \dots$ 。这样的模型就是**分布滞后模型**

一般情况下分布式滞后模型(DL)表示为

$$y_t = \alpha + B(L)x_t + \varepsilon_t$$

其中 $B(L)$ 是某个有限或无限滞后多项式算子。

或写作：

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_k x_{t-k} + \mu_t \quad (1)$$

或

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \mu_t \quad (2)$$

式(1)为有限滞后模型，式(2)为无限滞后模型。

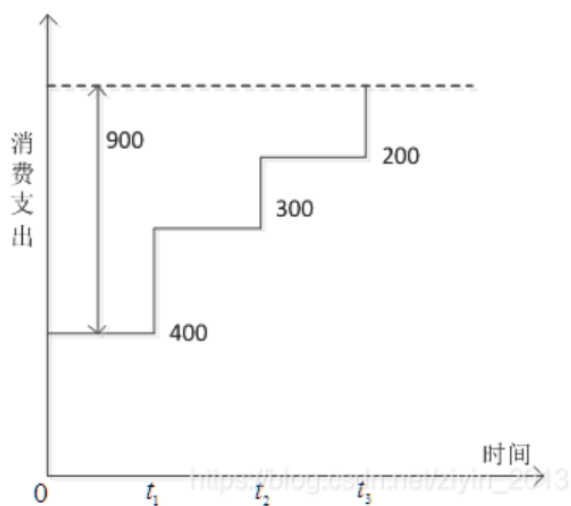
其中

- k 为滞后解释变量的滞后期长度或者说滞后时间间隔。
- 滞后算子 β_0 表示本期 x 变化一单位对 y 平均值的影响程度
- 滞后算子 $\beta_i, (i = 1, 2, \dots, k)$ 表示各滞后期 x 的变动对 y 平均值影响程度

例：消费滞后。按照一般规律，人们的消费不仅依赖于当期收入，还依赖于前期收入，即人们会分期消费。假定某人收入每年增加1000元，那么这种收入的增加对该消费者的年消费支出会产生什么影响？它也许会在收入增加后的第一年增加400元的消费支出，第二年增加300元，第三年增加200元，把余下的100元作为储蓄。到第三年末此人的消费将增加900元。

则消费分布滞后函数为 $y_t = \alpha + 0.4x_{t-2} + 0.3x_{t-1} + 0.2x_t + \mu_t$

其中 α 代表基础消费，这里 $x_t = x_{t-1} = x_{t-2} = 1000$ 代表每年增加的收入， μ_t 代表修正误差。



$$y_t = \alpha + B(L)x_t + \varepsilon_t$$

- 一旦所有的高频值被聚合为相应的低频值，这种分布式滞后模型就构成了基础的桥式模型。
- 考虑到混合频率数据，Ghysels等人(2004)引入了混合数据采样(MIDAS)方法。

MIDAS模型

- 假定 $\{y_t, t \in Z\}$ 为低频可观测的一元时间序列，其滞后算子用 B 表示，即 $B y_t = y_{t-1}$ 。
- ✓ • MIDAS 回归涉及到对随机过程 $\{x_\tau^{(i)}, \tau \in Z\}, i = 0, \dots, k$ 的线性投影。
其中 $x_\tau^{(i)}$ 为高频可观测序列，即在对应的每一个低频时期 $t = t_0$ ，我们观测到 $x_\tau^{(i)}$ 的时期，其中 $m_i \in N$ 为高频解释变量相对于低频解释变量抽样频率的倍差。
- 当 $m_i = 1$ 时，MIDAS 模型退化为同频模型。
- 高频解释变量 $x_\tau^{(i)}$ 滞后算子用 L 表示，即 $L x_\tau^{(i)} = x_{\tau-1}^{(i)}$ 。

基于上述设定，MIDAS 模型可表述为如下形式：

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

上述模型设定形式可采用通常的时间序列回归或贝叶斯方法进行估计。但随着模型中滞后阶数的增加，待估参数 $d = p + \sum_i^k l_i$ 也会快速增加。为了解决这一问题，Ghysels(2004)等建议采用一个充分灵活的函数形式对原参数进行约束，即

$$\beta_j^{(i)} = f_i(\gamma_i, j), j = 0, \dots, l_i, \gamma_i = (\gamma_i^{(1)}, \dots, \gamma_{q_i}^{(i)}), q_i \in N$$

通过函数约束可以大大减少待估参数个数，从 $d = p + \sum_i^k l_i$ 减少到 $q = \sum_{i=0}^k q_i$ 函数化约束虽然可以有效减少待估参数的个数，但也导致模型不再线性化，因此需要采用 NLS（非线性最小二乘法）或极大似然进行估计。

为了说明函数化约束的优势，下图分别给出了无约束、正确约束和不正确约束三种条件下均方误差和估计值与真实参数差异的结果。

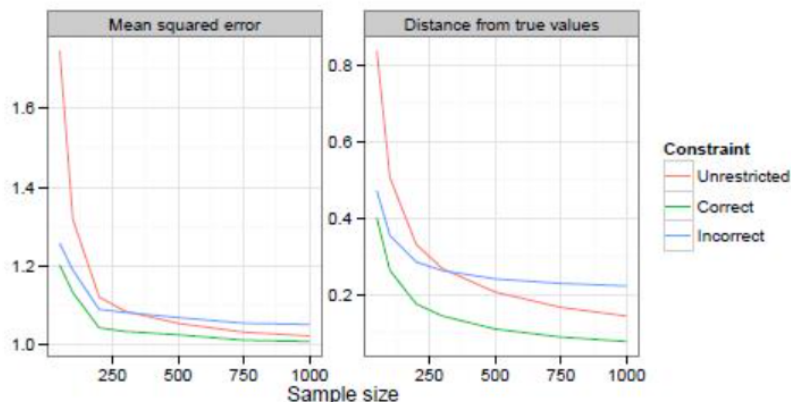


Figure 1: A plot depicting efficiency gains when the correct non-linear constraint is imposed. The left panel plots the average out-of-sample prediction accuracy against sample size. The right panel plots the average Euclidean distance of estimated model parameters to their true values.

图1:当施加正确的非线性约束时，描述效率增益的图。左面板绘制了平均样本外预测精度与样本大小的关系。右图绘制了估计模型参数到其真实值的平均欧氏距离。

由图 1 结果可知，随着样本容量的增加，即使不正确的约束也表现出了较为理想的效果。

更为常见的，上述 MIDAS 模型是以一个小巧的形式表述：

$$\alpha(B)y_t = \beta(L)^T x_{t,0} + \varepsilon_t$$

其中 $\alpha(z) = 1 - \sum_{j=1}^p \alpha_j z^j$ ，且

$$x_{t,0} := (x_{tm_0}^0, \dots, x_{tm_i}^i, \dots, x_{tm_l}^l)^T,$$

$$\beta(z) := \sum_{j=0}^l \beta_j z^j, \beta_j = (\beta_j^0, \dots, \beta_j^i, \dots, \beta_j^l)^T,$$

$$L^j x_{t,0} := x_{t,j} = (L^j x_{tm_0}^0, \dots, L^j x_{tm_i}^i, \dots, L^j x_{tm_l}^l)^T.$$

-
- 该方法与分布滞后模型密切相关，但在这种情况下，在较低频率采样的因变量 y_t 在较高频率采样的 T 的分布滞后上回归。在接下来的内容中，我们首先介绍模型的基本特征Ghysels等人(2004)，相应的无限制版本如Forni, Marcellino和Schumacher(2012)，然后是在文献中介绍的扩展。在符号方面， $tq = 1, \dots, T$ 表示基本时间单位(如季度)， m 表示在基本时间单位中数据被频率出现的次数。例如，将季度GDP增长和月度指标作为解