仔细阅读《A SURVEY OF ECONOMETRIC METHODS FOR MIXED-FREQUENCY DATA》
理解桥式方程、混合抽样分布、MIDAS基础模型和AR-MIDAS模型的思想和数学公式,明天应该会将文章中模型整理至表格中:

> ## 2.2 桥式方程
>
> 在混合频率数据存在的情况下，早期计量经济学方法之一依赖于桥式方程的使用，参见例如Baffigi、Golinelli、Parigi(2004)和Diron(2008)。
> 桥式方程是将("桥")***高频变量(如工业生产或零售销售)与低频变量(如季度实际GDP增长)连接起来***的线性回归方程，***在数据发布之前提供当前和短期发展的一些估计***。"桥模型"技术允许通过***使用高频指标计算低频变量的早期估计***。它们不是标准的宏观计量模型，因为具体指标的纳入不是基于因果关系，而是基于统计事实，即它们包含及时更新的信息。原则上，桥模型要求在投影期间应该知道**整套回归量**，**只允许对当前期间进行估计**。无论如何，在实践中，情况并非如此，尽管**桥梁模型的预测期限相当短，最多提前一两个季度。**
> 以GDP预测为例，由于月度指标在预测期内通常只有部分可用，因此季度GDP增长预测分两步进行。首先，对季度剩余时间的月度指标进行预测，通常基于*单变量时间序列模型*(在某些情况下，为了对月度指标进行更好的预测，已经实施了VAR)，然后*聚合以获得其季度对应值*。
> 其次，在桥式方程中使用聚合值作为回归量，从而获得GDP增长的预测。因此，待估桥梁模型为:
>
> $$y_{t_q} = \alpha + \sum_{i=1}^{j} \beta_i(L) x_{it_q} + u_{t_q} \qquad (1)$$
>
> $\beta_i(L)$为长度为k的*滞后多项式*，x_{itq}为按*季度频率汇总的选定月度指标*。
> 桥模型中包含的*月度指标*的选择通常基于*从一般到特定*的方法，并*依赖于不同的样本内或样本外标准，如信息标准或RMSE性能*。Bencivelli, Marcellino和Moretti(2012)提出了一种基于贝叶斯模型平均(BMA)的替代过程，该过程在实践上表现相当好。为了预测每月指标中缺失的观察值，然后将其相加得到季度的x_{itq}值，通常使用自回归法模型，其中滞后长度基于信息标准。

> ## 2.3混合数据抽样分布
>
> 滞后(DL)模型在文献中通常用于*描述解释变量变化的滞后效应随时间的分布*。一般情况下，给出了程式化的分布式滞后模型

$$y_{t_q} = \alpha + B(L)\, x_{t_q} + \varepsilon_{t_q} \tag{2}$$

其中$B(L)$是某个有限或无限滞后多项式算子
一旦*所有的高频值被聚合为相应的低频值*，这种模型就构成了桥方程的基础。
为了考虑混合频率数据，Ghysels等人(2004)引入了混合数据采样(MIDAS)方法，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$y_t q$在较高频率采样的x_{tm}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Foroni, Marcellino和Schumacher(2012)，然后是在文献中介绍的扩展。
在符号方面，$t_q = 1, \ldots, T_q$表示基本时间单位(如季度)，m表示在同一基本时间单位中较高采样频率出现的倍数。例如，将季度GDP增长和月度指标作为解释变量，$m = 3$。
w 是比待估计的低频变量更早获得的指标的月数值的数量。
低频变量在高频处可表示为$y_{tm} = y_{tq}, \forall t_m = m t_q$，其中$t_m$为高频处的时间下标(时间点）。

2.3.1 MIDAS基础模型

MIDAS回归基本上是紧密参数化的，简化形式的回归，涉及在不同频率采样的过程。
对高频解释变量的响应，使用***高度精简的分布式滞后多项式建模***，以防止可能导致的参数扩散，以及与滞后顺序选择相关的问题。单解释变量和hq-step-ahead的基本MIDAS模型，$hq = hm/m$由:

$$y_{t_q+mh_q} = y_{t_m+h_m} = \beta_0 + \beta_1 b(L_m;\theta)\, x^{(m)}_{t_m+w} + \varepsilon_{t_m+h_m} \tag{3}$$

当 $b\left(L^{1/m};\theta\right) = \sum_{k=0}^{K} c(k;\theta)\, L_m^k$,并且 $L_m^x x_{t_m}^{(m)} = x_{t_m-x}^{(m)}.\ x_{t_m+w}^{(m)}$ 是从高频的指标$x_t m$跳跃采样的（skip-sampled)。
$c(k,\theta)$以简约的方式是MIDAS的关键特性之一。
最常用的参数化之一是被称为”指数阿尔蒙滞后(Exponential Almon Lag)”的参数化，因为它与用于减少分布式滞后文献中多重共线性的平滑多项式阿尔蒙滞后函数密切相关。它常被表达为

$$c(k;\theta) = \frac{\exp\left(\theta_1 k + \ldots + \theta_Q k^Q\right)}{\sum_{k=1}^{K} \exp\left(\theta_1 k + \ldots + \theta_Q k^Q\right)} \tag{4}$$

众所周知，这个函数相当灵活时，只需几个参数就可以得到各种形状。
其中包括减少、增加或hump-shaped patterns(驼峰形状)。Ghysels, Santa-Clara和Valkanov(2005)使用带两个参数的函数形式，这允许一个巨大的灵活性，并决定了在回归中包含多少滞后。

注意，从*月指标计算季度级数的桥式方程的标准做法*对应于*对这个参数化函数施加限制*。具体来说，在季度-月的例子中，取季度中的最后一个月来生成季度系列相当于设置
$c(2;\theta)=c(3;\theta)=c(5;\theta)=c(6;\theta)=\ldots=c(11;\theta)=c(12;\theta)=0$

另一种可能的参数化，也只有两个参数，是所谓的“Beta Lag”，因为它是基于Beta函数:

$$c\left(k;\theta_1,\theta_2\right)=\frac{f\left(\frac{k}{K},\theta_1;\theta_2\right)}{\sum\limits_{k=1}^{K}f\left(\frac{k}{K},\theta_1;\theta_2\right)} \tag{5}$$

where $c\left(x,a,b\right)=\frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ and $\Gamma\left(a\right)=\int_0^\infty e^{-x}x^{a-1}dx$.

Ghysels, Rubia和Valkanov(2009)还提出了其他三种不同的滞后系数参数化:

Ghysels, Rubia and Valkanov (2009) propose also three other different parameterizations of the lag coefficients: a linear scheme, with $c\left(k;\theta\right)=\frac{1}{K}$, where there are no parameters to estimate in the lagged weight function; an hyperbolic scheme, with $c\left(k;\theta\right)=\frac{g\left(\frac{k}{K},\theta\right)}{\sum\limits_{k=1}^{K}g\left(\frac{k}{K},\theta\right)}$, $g\left(k,\theta\right)=\frac{\Gamma(k+\theta)}{\Gamma(k+1)\Gamma(\theta)}$ where the gamma function has only one parameter to estimate, but it's not as flexible as the Beta specification; a geometric scheme, with $c\left(k;\theta\right)=\frac{\theta^k}{\sum\limits_{k=1}^{\infty}\theta^k}$, $|\theta|\leq 1$ and $c\left(k;\theta\right)$ are normalized so that they sum up to one.

上面描述的参数化都非常灵活。对于不同的参数值，它们可以采用不同的形状:附加在不同滞后上的权重可以缓慢或快速下降，甚至具有驼峰形状。因此，从数据中估计参数会自动确定权重的形状，并相应地确定回归中要包含的滞后数。
MIDAS模型可以在y_t到$x_{t-h}^{(m)}$的回归中使用非线性最小二乘(NLS)进行估计。Ghysels, Santa-Clara和Valkanov(2004)表明MIDAS回归与*典型的集合所有序列到最低频率抽样的方法*相比，总是导致更有效的估计。此外，他们还表明，MIDAS和分布式滞后模型的离散化偏差是相同的，当回归量采样更频繁时，离散化偏差就消失了。
预测是根据：

$$\widehat{y}_{T_m^y+h_m|T_m^x}=\widehat{\beta}_0+\widehat{\beta}_1 b\left(L_m;\widehat{\theta}\right)x_{T_m^x}^{(m)}. \tag{6}$$

注意，MIDAS是依赖于h的，因此需要对每个预测水平进行重新估计。

2.3.2 AR-MIDAS模型

由于自回归模型通常提供与那些包含解释变量的模型相比具有竞争性的预测，在MIDAS模型中引入自回归术语是一种可取的扩展，尽管并不直接。Ghysels, SantaClara和Valkanov(2004)指出，引入滞后因变量会造成效率损失。此外，这将导致*解释变量中季节性模式的产生*。
考虑在$m=3$ (r为月，y为季度)的基础模型中增加一个较低的频率滞后 $ytm, ytm-3$:

$$y_{t_m}=\beta_0+\lambda y_{t_m-3}+\beta_1 b\left(L_m;\theta\right) x_{t_m+w-3}^{(3)}+\varepsilon_{t_m}. \tag{7}$$

正如Clements和Galvao(2009)所强调的，这种策略通常是不合适的。当我们将模型写成如下形式时，原因就很清楚了:

$$y_{t_m}=\beta_0(1-\lambda)^{-1}+\beta_1\left(1-\lambda L_m^3\right)^{-1} B\left(L_m;\theta\right) x_{t_m+w-3}^{(3)}+\left(1-\lambda L_m^3\right)^{-1} \varepsilon_{t_m}. \tag{8}$$

关于r的多项式是关于$L^{1/3}$的多项式和关于L的多项式的乘积。这个乘积产生y到$x^{(3)}$是否显示季节性模式。
为了避免这种不方便，作者建议引入AR动态作为一个公共因素:

$$y_{t_m}=\beta_0+\lambda y_{t_m-h_m}+\beta_1 b\left(L_m;\theta\right)\left(1-\lambda L_m^{h_m}\right) x_{t+w-h_m}^{(3)}+\varepsilon_{t_m}. \tag{10}$$

MIDAS (the basic model), take the residuals $\widehat{\varepsilon}_{t_m}$ and estimate an initial value for λ, say λ_0, where $\widehat{\lambda}_0=\left(\sum \widehat{\varepsilon}_{t_m+w-h_m}^2\right)^{-1} \sum \widehat{\varepsilon}_{t_m} \widehat{\varepsilon}_{t_m+w-h_m}$. Then construct $y_{t_m}^*=y_{t_m}-\widehat{\lambda}_0 y_{t_m-h_m}$ and $x_{t_m+w-h_m}^{*(3)}=x_{t_m+w-h_m}^{(3)}-\widehat{\lambda}_0 x_{t_m-(h_m-w)-h_m}^{(3)}$. The estimator $\widehat{\theta}_1$ is obtained by applying nonlinear least squares to:

$$y_{t_m}^*=\beta_0+\beta_1 b\left(L_m;\theta\right) x_{t_m+w-h_m}^{*(3)}+\varepsilon_{t_m}. \tag{11}$$

A new value of λ, $\widehat{\lambda}_1$, is obtained from the residuals of this regression. Then a new step is run, using $\widehat{\lambda}_1$ and $\widehat{\theta}_1$ as the initial values. In this way, the procedure gets the estimates and $\widehat{\lambda}$ and $\widehat{\theta}$ that minimize the sum of squared residuals.

该程序得到了使残差平方和最小的估计值

2.3.3 无限制MIDAS模型

Foroni, Marcellino和Schumacher(2012)研究了MIDAS的一种变体的性能，它不求助于函数分布式滞后多项式。在本文中，作者讨论了如何在一般线性动态框架中推导不受限制的MIDAS (U-MIDAS)回归，以及在这种情况下，可以识别底层高频模型的参数。
这个无限制MIDAS模型基于线性滞后多项式像:

$$\begin{aligned} c(L^m)\omega(L)y_{t_m} &= \delta_1(L)x_{1t_m-1}+\ldots+\delta_N(L)x_{Nt_m-1}+\epsilon_{t_m}, \\ t &= 1,2,3,\ldots \end{aligned} \tag{12}$$

where $c(L^m)=(1-c_1L^m-\ldots-c_cL^{mc})$, $\delta_j(L)=(\delta_{j,0}+\delta_{j,1}L+\ldots+\delta_{j,v}L^v)$, $j=1,\ldots,N$.

2.5.4 Factor-MIDAS

可以用从大型数据集中提取的因子来增强MIDAS回归，从而获得更丰富的模型族，利用大型高频数据集来预测低频变量。
虽然基本的MIDAS框架由一组高频指标上的一个低频变量的回归组成，但Factor-MIDAS方法利用估计的因素，而不是单一或小组经济指标作为回归量。Marcellino和Schumacher(2010)提出了另一种MIDAS回归。在标准MIDAS情况下，他们遵循Clements和Galvao(2008)的方法，而作为一种修正，他们评估了一种更一般的回归方法，称为不受限制的Factor0MIDAS，其中低频变量和高频指标之间的动态关系是不受限制的，与Ghysels等人(2007)提出的分布式滞后函数相反。作为第三种选择，他们考虑了Altissimo等人(2010)提出的回归方案，该方案只考虑在高频和低频采样的变量之间在特定频率的相关性。这种方法被称为平滑的MIDAS，因为回归基本上消除了高频相关性。
该信息集由大量固定的月度指标X_{tm}组成。最后一次观测是在时间$T_m + w, w > 0$，允许比待估计的低频变量更早获得的指标的月值最多为$w > 0$。X_{tm}使用因子表示建模，其中r个因子F_{tm}被估计，以便总结X_{tm}中的信息。估计因子F_{tm}用于季度频率变量的预测。

卡尔曼滤波技术可以处理由于发布滞后和时间序列的低频特性而导致的边缘数据和缺失值。