附录 1 混频动态因子模型

混频数据是指各变量的频率不完全相同，例如，月度数据和季度数据的混合。混频动态因子模型将不同频率的数据结合分析，不同于经典因子模型只能处理少量的变量，混频动态因子模型可以从一系列时间序列中提取公共因子，这些从高维数据中提取的因子可以充分解释经济波动。本文采用生产者物价指数、社会消费品零售总额、工业增加值等一系列月度指标和季度GDP指标，基于混频动态因子模型对月度GDP进行预测。借鉴郑挺国和王霞(2013)关于经济周期测度的方法，将季度 GDP 同比增长率和影响 GDP 变动的一系列月度经济指标结合预测出我国月度 GDP 同比增长率。

为将月度和季度变量有效整合进同一分析模型，首先要探讨月度与季度变量同比增长率之间的内在逻辑关联。将季度序列视为只包含特定月份观测值的月度序列，若 Y_t^Q 是季度流量变量，则月份 $t=3,9,12$ 时，Y_t^Q 有观测值，而在其他月份为缺失值。如果 Y_t^M 是对应 Y_t^Q 的不可观测的月度变量，那么：$Y_t^Q = Y_t^M + Y_{t-1}^M + Y_{t-2}^M = 3[\frac{1}{3}(Y_t^M + Y_{t-1}^M + Y_{t-2}^M)]$ 。即可以将每季度 3 个月度变量的算术平均值视为季度变量，借鉴 Mariano 和 Murasawa (2003)的处理方式，可得：

$$Y_t^Q = 3(Y_t^M + Y_{t-1}^M + Y_{t-2}^M)^{1/3} \tag{1}$$

进一步，对式(1) 两边取对数后求 12 阶差分可得:

$$\ln Y_t^Q - \ln Y_{t-12}^Q = \frac{1}{3}[(\ln Y_t^M - \ln Y_{t-12}^M) + (\ln Y_{t-1}^M - \ln Y_{t-13}^M)(\ln Y_{t-2}^M - \ln Y_{t-14}^M)]$$

设 $y_t^Q = \ln Y_t^Q - \ln Y_{t-12}^Q$ 和 $y_t^M = \ln Y_t^M - \ln Y_{t-12}^M$ 分别为季度变量和 t 月份月度变量的同比增长率，相应月份的月度变量增长率进行加权平均：

$$y_t^Q = \frac{1}{3}(y_t^M + y_{t-1}^M + y_{t-2}^M) \tag{2}$$

$\boldsymbol{X}_t$ 是 n 个可观测月度同比增长率构成的向量， $\boldsymbol{y}_t^M$ 是对应于 $\boldsymbol{y}_t^Q$ 不可观测的月度同比增长率向量，进一步利用同比增长率 $\boldsymbol{y}_t^M$ 构造动态因子模型：

$$\begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{y}_t^M \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\beta}^M \end{pmatrix} \boldsymbol{f}_t + \begin{pmatrix} \boldsymbol{u}_t \\ \boldsymbol{u}_t^M \end{pmatrix}$$

$$\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{f}_t = \boldsymbol{v}_t$$

由于月度 GDP 的同比增长率 $\boldsymbol{y}_t^M$ 不能被直接观测到，通常使用季度 GDP 同比增长率 $\boldsymbol{y}_t^Q$ 代替 $\boldsymbol{y}_t^M$，最终月度混频动态因子模型表达式为：

$$\begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{y}_t^Q \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} & \boldsymbol{0} & \boldsymbol{0} \\ \frac{1}{3}\boldsymbol{\beta}^M & \frac{1}{3}\boldsymbol{\beta}^M & \frac{1}{3}\boldsymbol{\beta}^M \end{pmatrix} \begin{pmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \end{pmatrix} + \begin{pmatrix} \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_t \\ \boldsymbol{u}_t^M \\ \boldsymbol{u}_t^M \\ \boldsymbol{u}_t^M \end{pmatrix}$$

本文借鉴 Stock 和 Watson(2002)的估计方法，基于 EM 算法对月度 GDP 同比增长率进行估计。

附表 1 样本数据的统计特征

	R	LD	M	HA	TS	EX	GDP_M	CPI
均值	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
中位数	0.033	0.070	0.084	0.074	-0.016	0.079	0.013	-0.145
最大值	2.063	2.061	2.141	2.590	4.496	2.284	4.723	3.280

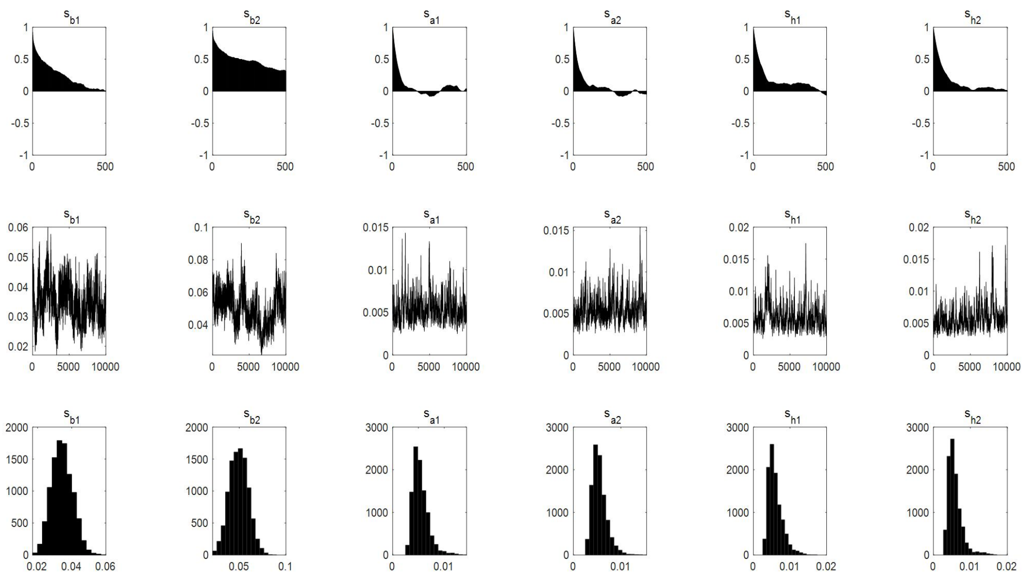
最小值	-4.267	-4.371	-3.281	-6.265	-2.070	-5.321	-5.724	-2.133
标准差	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
偏度	-0.812	-0.793	-0.703	-1.538	1.438	-0.802	-0.986	0.686
峰度	1.678	1.666	0.653	7.518	4.042	2.619	12.568	0.933
JB 统计量	54.264***	52.637***	24.190***	656.552***	245.626***	93.458***	1607.748***	27.510***
样本量	249	249	249	249	249	249	249	249

注: ***表示在 1%的水平下显著。 JB 统计量为 Jarque-Bera 正态性检验统计量，用来检验序列是否服从正态分布。

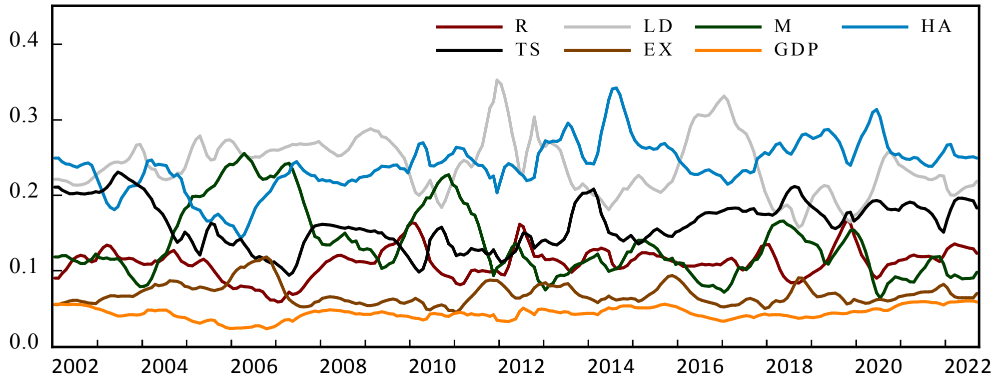
附表 2　　单位根检验

	ADF 检验				PP 检验			
变量	(C T P)	T 统计量	P 值	结论	(C T)	T 统计量	P 值	结论
R	(C T 0)	-6.017	0.000***	平稳	(C T)	-4.913	0.000***	平稳
LD	(C T 0)	-3.212	0.085*	平稳	(C T)	-3.357	0.0597*	平稳
M	(C T 0)	-3.594	0.032**	平稳	(C T)	-4.333	0.003***	平稳
HA	(C T 0)	-3.316	0.066*	平稳	(C T)	-3.685	0.025**	平稳
TS	(C T 0)	-4.348	0.003***	平稳	(C T)	-3.711	0.023**	平稳
EX	(C T 0)	-4.350	0.003***	平稳	(C T)	-4.525	0.001***	平稳
MONGDP	(C T 0)	-5.587	0.000***	平稳	(C T)	-5.653	0.000***	平稳
CPI	(C T 0)	-3.577	0.034**	平稳	(C T)	-3.439	0.048**	平稳

注: ***，**，*分别表示在 1%，5%和 10%的水平上显著性。(C T N)指 ADF、PP 检验中的常数项、时间趋势项和滞后阶数。下同。



附图 1　自相关(第一行)、样本路径(第二行)及后验密度图(第三行)



附图 2　各变量在 HDFCI 中的动态权重

附表 3　不同周期的 Granger 因果检验

	原假设	阶数	F-Statistic	Prob.
短周期	HDFCI 不是 CPI 的格兰杰原因	6	1.0662	0.3837
	CPI 不是 HDFCI 的格兰杰原因	6	0.5931	0.7357
	HDFCI 不是 GDP 的格兰杰原因	6	1.2037	0.3052
	GDP 不是 HDFCI 的格兰杰原因	6	2.9753	0.008***
中周期	HDFCI 不是 CPI 的格兰杰原因	6	2.1407	0.0497**
	CPI 不是 HDFCI 的格兰杰原因	6	0.5033	0.8056
	HDFCI 不是 GDP 的格兰杰原因	6	2.5975	0.0187**
	GDP 不是 HDFCI 的格兰杰原因	6	2.2099	0.043**
长周期	HDFCI 不是 CPI 的格兰杰原因	6	3.6427	0.0018***
	CPI 不是 HDFCI 的格兰杰原因	6	0.9639	0.4504
	HDFCI 不是 GDP 的格兰杰原因	6	0.4631	0.8352
	GDP 不是 HDFCI 的格兰杰原因	6	2.5202	0.0221**

注: ***，**，*分别表示在 1%，5%和 10%的水平上显著性。下同。

附录 2 谱分析

谱分析主要是为了描述两个信号之间的统计相关程度而提出的基于随机过程的理论，假设两时间序列$\{x_t\}$和$\{y_t\}$是平稳的，互谱和互相关函数存在一定的联系，它们之间互为傅里叶变换，已知两个时间序列的协方差函数为$\gamma_{xy}(\mathrm{k}),\mathrm{k}=0,\pm1,\pm2,\cdots$，对应的自协方差生成函数为：

$$r_{xy}(B)=\sum_{k=-\infty}^{\infty}\gamma_{xy}(\mathrm{k})B^k$$

为了计算互谱密度，对互相关函数进行傅里叶变换得到的互谱密度函数为：

$$f_{xy}(w)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_{xy}(\mathrm{k})e^{-iwk}=\frac{1}{2\pi}r_{xy}(e^{-iw})$$

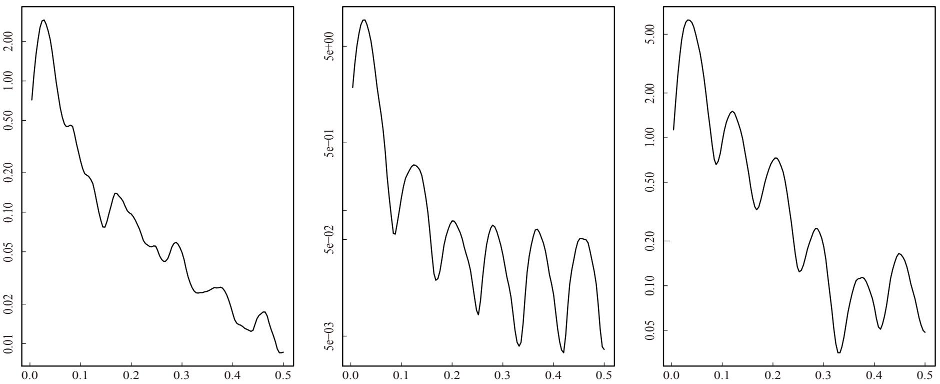
为了更好地作互谱分析，时域信号需要转换为频域信号，为了进一步了解信号中包含的幅度信息和相位信息，需要做出假设$f_{xy}(w)=c_{xy}(w)-iq_{xy}(w)$，其中$c_{xy}(w)$和$q_{xy}(w)$分别称为共谱和积分谱，进一步，$f_{xy}(w)=A_{xy}(w)e^{i\phi_{xy}(w)}$可得平方相干函数如下：

$$K_{xy}^{2}(w)=\frac{\left|f_{xy}(w)\right|^{2}}{f_{x}(w)f_{y}(w)}$$

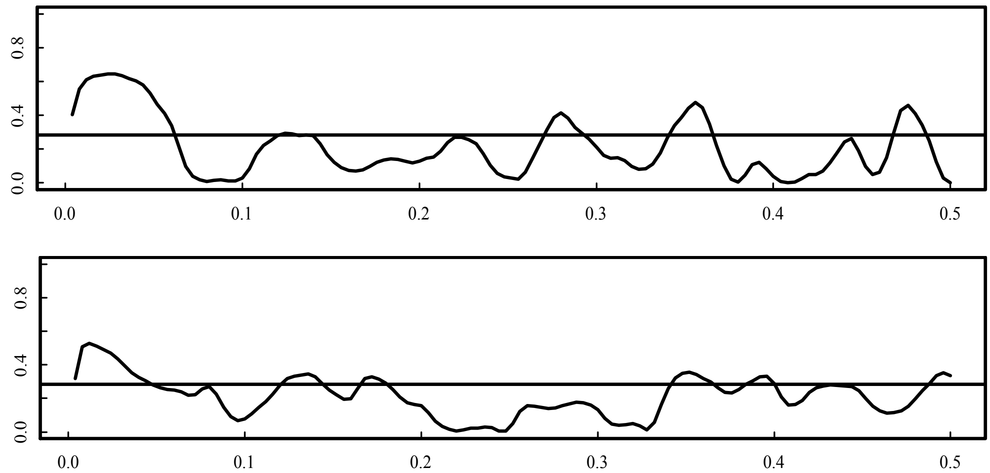
它表示$\{x_t\}$和$\{y_t\}$的线性相关程度，对应的相谱为：

$$\phi_{xy}(w)=\tan^{-1}\left(\frac{-q_{xy}(w)}{c_{xy}(w)}\right)$$

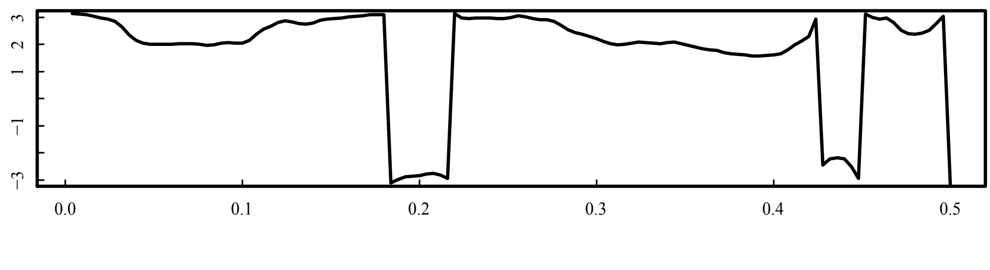
上式的正负反映了向量之间是否有领先关系，表现为所有频率的延时特性，而且相谱与对应频率商的绝对值表示领先或者滞后的期数，互谱分析模型在构建 HDFCI 过程中不可或缺。

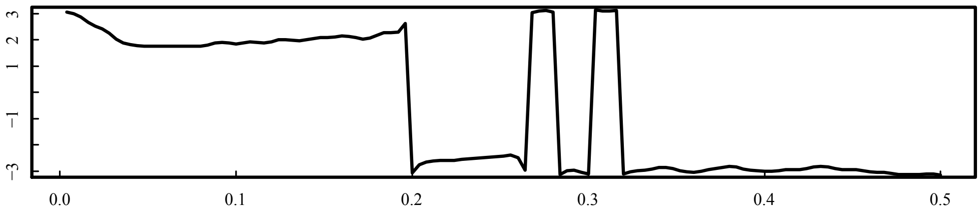


附图 3　HDFCI、CPI、GDP 的样本周期图

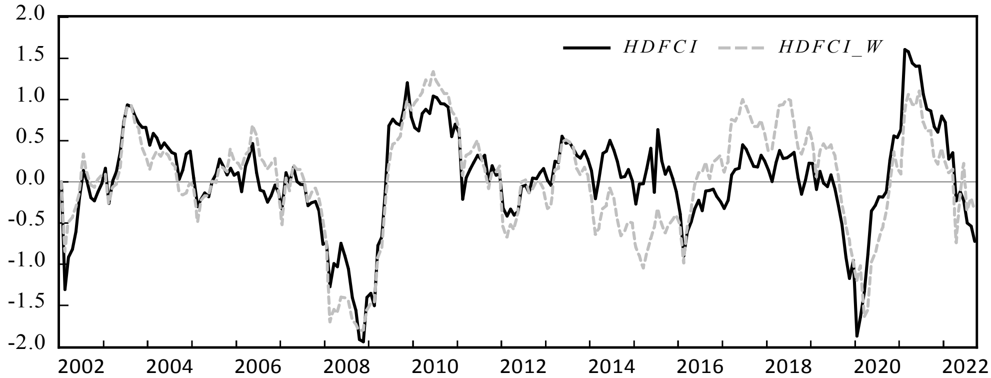


附图 4　HDFCI 对 CPI、GDP 的平方相干谱图





附图 5　HDFCI 对 CPI、GDP 的相谱图



附图 6　HDFCI、HDFCI_W 的趋势图

附表 4　HDFCI_W 预警能力稳定性检验

	滞后阶数					
	1		2		3	
原假设:	F 值	P 值	F 值	P 值	F 值	P 值
HDFCI_W 不是 CPI 的 Granger 原因	20.179	0.000***	8.019	0.000***	8.994	0.000***
CPI 不是 HDFCI_W 的 Granger 原因	29.728	0.000***	14.143	0.000***	7.780	0.000***
HDFCI_W 不是 GDP 的 Granger 原因	18.920	0.000***	14.074	0.000***	11.647	0.000***
GDP 不是 HDFCI_W 的 Granger 原因	6.047	0.015**	4.974	0.008 ***	4.274	0.006 ***
	4		5		6	
原假设:	F 值	P 值	F 值	P 值	F 值	P 值
HDFCI_W 不是 CPI 的 Granger 原因	6.522	0.000***	5.423	0.000***	4.360	0.000***
CPI 不是 HDFCI_W 的 Granger 原因	5.119	0.000***	3.815	0.002 ***	3.112	0.006 ***
HDFCI_W 不是 GDP 的 Granger 原因	9.040	0.000***	7.618	0.000***	6.098	0.000***
GDP 不是 HDFCI_W 的 Granger 原因	2.658	0.034**	2.133	0.0624*	1.778	0.1041