

基于 IGARCH 投影寻踪回归的国际油价走势拟合模型

王吉培¹ 肖宏伟²

(1. 西南财经大学统计学院, 成都, 610074, 2. 北京化工大学经济管理学院, 北京, 100029)

摘要:随着我国对石油进口的依赖程度不断增加,国际油价变化对我国经济的影响越来越大。本文充分考虑了金融时间序列的方差时变性的特点,建立基于三种不同分布的 IGARCH 模型对国际油价的走势进行刻画,引入非参数投影寻踪回归模型的理论方法,对多维数据进行投影降维分析,结果表明基于 IGARCH 的投影寻踪回归具有较强的模型拟合能力,我国应采取一定的措施把国际油价变化对我国经济的影响减到最小。

关键词: 国际油价; IGARCH 模型; 投影寻踪回归; 金融时序

中图分类号: F224

文献标识码: A

一、引言

2002年以来,WTI油价从每桶20美元左右一路飙升到90美元,2008年4月达到98美元;三位数的油价指日可待。1993年我国开始成为石油净进口国以来,我国原油进口总量呈逐年增长趋势,尤其是近年来这种趋势更为明显。尽管我国人均石油消费量不足美国的1/20、日本与欧洲的1/15,但在消费总量上却是仅次于美国的第二大石油消费国。2003年我国每天的原油消费量增长到44万桶,占全球原油需求增长的35%。在今后一个时期,随着工业化和城镇化步伐加快,我国的石油进口将持续增长,供求缺口可能会进一步扩大。据预测,到2010年和2020年,我国石油供需缺口将分别为1.55-1.57亿吨和2.4-2.95亿吨,对海外资源的依存度分别为46.3%-52.3%和55.8%-62.1%,与目前美国58%的对外依存度大体相当。随着我国石油对外依存度的提高,国际油价持续上涨的趋势会越来越大。

本文通过1986年1月至2008年4月的月度数据,对国际国际油价的走势进行拟合,建立了基于三种不同分布的 IGARCH (1, 1) 模型,用这三种模型对国际油价的走势进行刻画,然后引入非参数投影寻踪回归模型的理论方法,将基于三种不同分布的 IGARCH (1, 1) 模型的拟合数据作为输入,对高维数据进行投影降维分析,从而提高了提高了模型的拟合能力。以廓清国际国际油价上涨的程度,为我国成品油定价,和制定能源工业发展规划提供理论依据。

二、GARCH 模型和投影寻踪回归介绍

(一) GARCH 模型

1. GARCH 模型

GARCH 模型一般由两个方程组成,一个是条件均值方程,另一个是条件方差方程。模型的一般表达式可写成^[1]:

$$\begin{aligned} r_t &= x_t' \beta + \varepsilon_t \\ h_t &= \omega + \sum_{j=1}^p \beta_j h_{t-j} + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \\ \varepsilon_t &= v_t \sqrt{h_t} \end{aligned} \quad (1)$$

其中, h_t 为条件方差, v_t 为独立同分布的随机变量, h_t 与 v_t 互相独立,参数满足条件

$\omega > 0$ ，一般常假定 v_t 为标准正态分布，另外，以上模型中 $\sum_{j=1}^p \beta_j + \sum_{i=1}^q \alpha_i < 1$ ，蕴涵 GARCH

过程为宽平稳。如果将此限制放宽即是积分型 GARCH 模型，即 IGARCH 模型。对于 IGARCH 模型而言，无条件方差并不存在。

2. 关于 GARCH 模型的扰动项的分布的假设

关于 GARCH 的扰动项的分布，一般会有 3 个假设：Normal (Gaussian) 分布，Student-t 分布，Generalized Error(GED)分布。在给定的分布假设下，GARCH 模型常用极大似然估计法进行估计，似然函数可通过对偶牛顿算法或信赖域算法极大化得，以 GARCH(1,1)模型为例介绍这三种分布的对数似然函数^[2]。

(1). 对于扰动项服从 Normal (Gaussian) 分布的 GARCH(1,1)模型，它的对数似然函数为：

$$\ln L(\theta) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_1^T \ln h_t^2 - \frac{1}{2h_t^2} \sum_1^T (y_t - x_t \gamma)^2 \quad (2)$$

这里的 h_t^2 是 ε_t 的条件方差。

(2). 对于扰动项服从 Student-t 分布的 GARCH(1,1)模型，它的对数似然函数为：

$$\ln L(\theta) = -\frac{T}{2} \ln \left\{ \frac{\pi(k-2)\Gamma(\frac{k}{2})^2}{\Gamma(\frac{k+1}{2})^2} \right\} - \frac{1}{2} \sum_1^T \ln h_t^2 - \frac{k+1}{2} \ln \left\{ 1 + \frac{(y_t - x_t \gamma)^2}{h_t^2(k-2)} \right\} \quad (3)$$

参数的估计变成了在自由度 $k > 2$ 的约束下是对数似然函数最大化的问题。当 $k \rightarrow \infty$ 时，Student-t 分布接近于正态分布。

(3). 对于扰动项服从 Generalized Error(GED)分布的 GARCH(1,1)模型，它的对数似然函数为：

$$\ln L(\theta) = -\frac{T}{2} \ln \left\{ \frac{\Gamma(\frac{1}{r})^3}{\Gamma(\frac{3}{r})(\frac{r}{2})^2} \right\} - \frac{1}{2} \sum_1^T \ln h_t^2 - \sum_1^T \ln \left\{ \frac{\Gamma(\frac{3}{r})(y_t - x_t \gamma)^2}{h_t^2 \Gamma(\frac{1}{r})} \right\}^{r/2} \quad (4)$$

这里的 $r > 0$ 。 r 为 GED 的分布参数，或自由度。它控制着分布尾部的薄厚程度， $r < 2$ 表示尾部比正态分布更厚， $r = 2$ 表示 GED 分布退化为标准正态分布； $r > 2$ 则说明尾部比正态分布更薄。文中关于基于 GDE-GARCH 类建模，发现 $r < 2$ ，即具有厚尾的性质。

(二) 投影寻踪回归模型

投影寻踪回归模型 (Projection Pursuit Regression) 是由 Peter Hall 提出的，它是用于分析和处理非正态、非线性数据的一种新方法，其基本思想是：利用计算机技术，把高维数据通过某种组合投影到低维子空间上，并通过极小化某个投影指标，寻找出能反映原数据结构或特征的投影，以达到研究和分析高维数据的目的^[3]。

投影寻踪回归模型如下： $Y = G(\theta \cdot x) + \varepsilon$ ，作出 $u = \theta \cdot x$ 的密度核估计 $f(u)$ ，其中 Y 是一元的， x 是 k 元， G 是待定的函数形式， θ 为待定的投影方向。然后用权函数作出 G 的非

参数估计 g ，最后用交叉核实的方法求 $\min S(\theta) = \sum (Y - g)^2$ ， θ 的确定可用单纯形法求出，从而求出 G 的估计^[4]。

已知数据 $\{x_k, Y_k\}_{k=1}^n$ ， x_k 是 p 元的， Y_k 是一元的，则有非参数方程：

$$Y_k = G(x_k) + \varepsilon_k, 1 \leq k \leq n. \quad (5)$$

估计模型中 p 元函数 G ， $G(x) = E\{Y_k | x_k = x\}$ 。 G 是先将 p 元变量投影成一元变量，即取 $u = \theta'x_k$ ，再建立一元实数 u 和一元函数 G 的映射。其映像的投影方向为 $\theta = (\theta_1, \dots, \theta_p)$ ， $\{x_k, 1 \leq k \leq n\}$ 在这里，该投影方向应该使估计的误差平方和最小。

若选择解释变量的组合 $\{x_k, 1 \leq k \leq n\}$ 是来自密度函数 f 的 p 元随机样本，对每一个 p 元样本 x_k ，有一元的观察值 Y_k 与之相对应，且 $E\{Y_k | x_k = x\} = G(x)$ 。 G 是回归函数或是目标函数。从观测值 $\{x_k, Y_k\}_{k=1}^n$ 做出 p 元函数 $G(x)$ 的估计，为了解决数据的位数过高的问题，作沿着 θ 方向的一元函数：

$$g_\theta(u) = E\{G(x) | \theta'X = u\}, \theta \in \Omega. \quad (6)$$

在区域 $A \in R^p$ 内对 G 的第一次投影逼近函数 $G_1(x) = g_{\theta_1}(\theta_1'x)$ 。 θ_1 是极小化 $S(\theta) = E\{[G(x) - g_\theta(\theta'X)]^2 I(X \in A)\}$ 的结果。由于 G 是未知的，要作出 $S(\theta)$ 和 $g_\theta(u)$ 的估计，才能得到 $G_1(x)$ 的估计。设这里设 $\theta'x$ 的密度函数为 f_θ ，称为沿方向 θ 的边沿密度函数 X ，用样本 x_f 但不包括 x_k 来构造 f_θ 的核估计：

$$\hat{f}_{\theta(k)}(u) = \frac{1}{(n-1)h} \sum_{j=k} K\left(\frac{u - \theta \cdot x_j}{h}\right). \quad (7)$$

K 是核函数， h 是核框，排除 x_k 在外的 g_θ 估计为：

$$\hat{g}_{\theta(k)}(u) = \left[\frac{1}{(n-1)h} \sum_{j=k} Y_j K\left(\frac{u - \theta \cdot x_j}{h}\right) \right] / \hat{f}_{\theta(k)}(u). \quad (8)$$

借助交叉核实的思想， $\min \hat{S}(\theta) = \frac{1}{n} \sum_{k=1}^n [Y_k - \hat{g}_{\theta(k)}(\theta \cdot x_k)]^2 I(x_k \in A)$ 。 $\hat{\theta}_1$ 是 θ 的估计， $G_{1(k)}(x) = \hat{g}_{\hat{\theta}_1(k)}(\hat{\theta}_1 \cdot x)$ 是回归函数 G 在区域 A 的第一次投影逼近，即投影寻踪回归。

三、实证分析

(一)样本数据说明

本文选择西得克萨斯轻质原油（WTI）现货价格作为国际油价，国际油价数据来自美国能源部能源信息署（EIA）；国际油价为月度数据，样本区间选取为1986年1月至2008年4月。

(二)GARCH模型的建立和预测

1. GARCH模型的建立

首先，为了消除数据的短期线性相依性，我们引入ARMA(p,q)模型，根据国际油价序列的自相关系数ACF和偏相关系数PCF的特征，以及AIC和SC最小化原则，我们认为国际油价序列服从一个ARMA(1,1)过程。然后利用LM检验对残差 ε_t 是否存在ARCH或GARCH效应进行检验，结果显示样本序列在1%的显著性水平下，残差序列存在ARCH效应，说明样本序列具有异方差性。根据GARCH模型系数要求和赤池信息最小准则，我们比较GARCH(p,q)模型，发现 $\sum_{j=1}^p \beta_j + \sum_{i=1}^q \alpha_i$ 的值很接近于1，故选择IGARCH(1,1)模型建模，

以下是基于三种不同分布的IGARCH模型的估计结果：

表 1 基于三种不同分布的IGARCH模型的参数估计

IGARCH-GED		IGARCH-Guass			IGARCH-Student	
均值方程						
ω	21.9251 (0.0214)	ω	21.4861 (0.0004)		ω	22.3254 (0.1063)
ψ	0.9860 (0.0000)	ψ	0.9789 (0.0000)		ψ	0.9901 (0.0000)
β	0.2819 (0.0005)	β	0.3105 (0.0020)		β	0.2535 (0.0029)
方差方程						
ω	0.1120 (0.0794)	ω	0.1095 (0.0908)		ω	0.1059 (0.0581)
α	0.2344 (0.0000)	α	0.2429 (0.0000)		α	0.2221 (0.0000)
β	0.7655	β	0.7570		β	0.7778
GED Para	1.6459 (0.0000)				Student (DF)	10.8262 (0.0924)
Akaike	4.2464	Akaike	4.2510		Akaike	4.2311
Schwarz	4.3268	Schwarz	4.3179		Schwarz	4.3115

对三种模型的残差进行ARCH-LM检验，得到的残差序列在多项滞后时的统计结果接受原假设。说明基于三种不同分布的IGARCH模型能够很好的消除原序列的异方差。

2. Pearson吻合度(Pearson goodness-of-fit) 检验

Pearson吻合度(Pearson goodness-of-fit) 检验能够比较真分布和理论分布的接近程度(Palm等，1997)。将标准化残差 $\hat{e}_t$ 按大小分成g个单元， n_i 是第i个单元的观测数，在理

论分布是真分布的原假设下，检验统计量 $p(g) = \sum_{i=1}^g \frac{(n_i - En_i)^2}{En_i}$ 检验统计量的渐进分布介

于 $\chi^2(g-1)$ 和 $\chi^2(g-k-1)$ 之间，k是参数个数^[5]。Palm等(1997年)指出的选择不可观测，对T=2252，可设8=50。Konig(1982年)提出以的变化率的0.4次幂的速率变化。对于本文选择的样本总量T=1748。我们大约取g=40。Pearson吻合度检验结果如下列表2.

表2

Pearson吻合度检验

单元数(g)	IGARCH-GED		IGARCH-Guass		IGARCH-Student	
	P 值(g-1)	P 值(g-k-1)	P 值(g-1)	P 值 (g-k-1)	P 值(g-1)	P 值(g-k-1)
40	0.861153	0.636582	0.612942	0.380345	0.449199	0.203348
50	0.692326	0.447183	0.648507	0.442805	0.174643	0.061663
60	0.807312	0.612696	0.354745	0.201184	0.369646	0.185927

注：表中第三列和第四列分别是统计量P(g)服从 $\chi^2(g-1)$ 和 $\chi^2(g-k-1)$ 分布的概率P值。k表示参数个数。

Pearson吻合度检验表明在0.05的显著水平下接受原假设，即三种模型的分布地选择是正确的。

（三）基于 GARCH 模型的投影寻踪回归模型

前面由基于三种不同分布的 IGARCH 模型得到的和有价格的拟合值作为同应寻踪模型的输入变量，即 x_k ， $k=1,2,3$ ，国际油价的观测值作为 Y。利用交叉核实的思想，对 $Y = G(\theta \cdot x) + \varepsilon$ 进行核估计，得到各个投影的权重系数，该系数从大到小反映了投影向量对因变量的影响大小，投影寻踪模型的统计量 GCV 得分为 7.83，及对数似然估计值为 -583.78，投影矩阵为：

$$P=\begin{bmatrix} 0.5768,0.5772,0.5887 \\ 0.5607,0.5785,0.5767 \\ 0.5940\ 0.5763\ 0.5663 \end{bmatrix}$$

利用 bootstrap 抽样技术对模型进行进一步诊断，由以下 bootstrap 抽样估计出的投影值可以看出，每次抽样所求得的投影值相差不大，基本可以认为用该交叉核实的方法所得到的投影值是稳定合理的。表 3 是前三次 bootstrap 抽样的投影值。

表 3

基于 bootstrap 抽样的投影值

抽样	第一次 bootstrap 抽样			第二次 bootstrap 抽样			第三次 bootstrap 抽样		
变量	投影 1	投影 2	投影 3	投影 1	投影 2	投影 3	投影 1	投影 2	投影 3
x1	0.5747	0.5771	0.5785	0.5783	0.5769	0.5768	0.5770	0.5772	0.5771
x2	0.5534	0.5548	0.5524	0.5721	0.5916	0.6027	0.5677	0.5688	0.5635
x3	0.6029	0.5993	0.6001	0.5816	0.5632	0.5514	0.5871	0.5860	0.5912

（四）拟合模型的评价

1.拟合模型的评价标准^[6]

(1) MSE 准则

均方误差（Mean Square Error）是实际值与预测值之差取平方后的算术平均：

$$MSE = \frac{1}{T} \sum_{i=1}^T (y_i - \hat{y}_i)^2 \quad (9)$$

(2) MAE 准则

平均绝对误差(Mean Absolute Error) 考虑了误差的绝对距离：

$$MAE = \frac{1}{T} \sum_{i=1}^T |y_i - \hat{y}_i| \quad (10)$$

(3) MSPE 准则

均方比例误差（Mean Square Percentage Error）是指预测误差占实际值比例取平方后的算术平均，它考虑了误差相对于实际值的比例，不会赋予相同的比重。

$$MSPE = \frac{1}{T} \sum_{i=1}^T \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2 \quad (11)$$

(4) MAPE 准则

平均绝对比例误差（Mean Absolute Percentage Error）类似于 MSPE，是误差占实际值的比例取绝对值后的算术平均。

$$MAPE = \frac{1}{T} \sum_{i=1}^T \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (12)$$

(5) AMAPE 准则

不对称平均绝对比例误差（Asymmetric Mean Absolute Percentage Error）更进一步的考虑了使用 MAPE 法时，在相同的误差下由于不同实际值所产生的不对称现象。

$$AMAPE = \frac{1}{T} \sum_{i=1}^T \left| \frac{y_i - \hat{y}_i}{y_i + \hat{y}_i} \right| \quad (13)$$

表 4 四种模型拟合精度比较

	IGARCH-GED	IGARCH-Guass	IGARCH-Student	IGARCH-PPR
MSE	6.438547721	6.57487307	6.360068326	4.566513
MAE	1.770818101	1.7809142	1.765351402	1.523484
MSPE	0.006547673	0.006552961	0.006551258	0.005676
MAPE	0.061276971	0.061345164	0.061294283	0.05659
AMAPE	0.030657239	0.030709855	0.030654735	0.028084

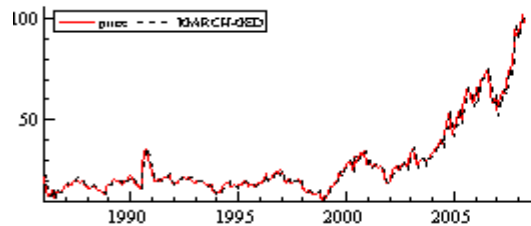


图1 IGARCH-GED模型的拟合值和观测值

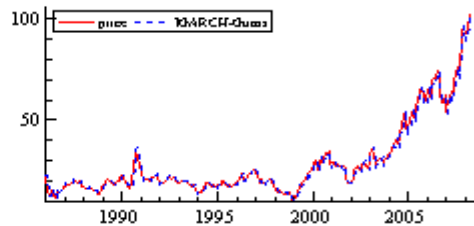


图2 IGARCH-Guass模型的拟合值和观测值

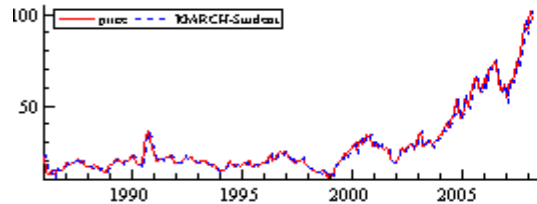


图3 IGARCH-Student模型的拟合值和观测值

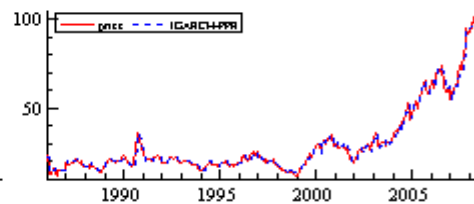


图4 IGARCH-投影寻踪回归模型的拟合值和观测值

四、结论

1. 本文以国际油价的走势为例, 详尽的比较了三种不同分布的IGARCH模型, 以及在此基础之上的投影寻踪, 由于石油价格序列存在ARCH效应, 即方差具有时变性, 不能用一般的最小二乘回归模型去拟合, 所以引入了刻画时变方差的GARCH模型。

2. 本文引入了三种不同分布的IGARCH模型对动态序列进行拟合, 并对建模后的残差进行对三种模型的残差进行了ARCH-LM检验, 得到的残差序列在多项滞后时的不存在ARCH效应, 说明基于三种不同分布的IGARCH模型能够很好的消除原序列的异方差, 然后又对残差分布进行了Pearson吻合度检验, 结果表明在0.05的显著水平下接受原假设, 即三种模型的分布地选择是正确的。

3. 从模型拟合效果评价来看, 基于Student的IGARCH模型的拟合结果最优, 其次是基于GED的IGARCH模型, 但两种分布下的拟合效果差异并不大, 拟合最差的是基于Guass分布的IGARCH。这是因为, 在IGARCH-GED模型中, r 为GED的分布参数, 或自由度, 它控制着分布尾部的薄厚程度, $r < 2$ 表示尾部比正态分布更厚, 这里 $r = 1.6459$, 所以石油价格序列是厚尾分布, 相对于正态Guass分布, 显然GED和Student分布更合适一些, 所以拟合误差会小一些。

4. 鉴于投影寻踪具有数据降维的特性, 在IGARCH模型拟合的基础之上, 引入了人投影寻踪回归的概念, 将基于三种不同分布的IGARCH模型的拟合值作为解释变量, 石油价格的实际观测值作为被解释变量, 利用交叉核实的方法, 对动态数据进行非参数拟合, 并利用bootstrap抽样的技术对模型进行进一步诊断, 拟合结果显示, 基于三种不同分布的IGARCH模型的投影寻踪回归拟合效果明显优于其他三种模型, 为动态时序的拟合和预测提供了一种新的思路。

5. 国际油价变化是多种因素互动造成的, 应对的措施也应该是综合性的, 既要着力解决影响当前经济运行的突出矛盾, 更要放眼长远, 建立健全多方面、系统性的石油供应保障和风险规避体系, 维护国家石油安全, 保障经济持续快速协调健康发展。建立完善石油战略储备, 积极参与期货交易, 规避国际油价风险, 把国际油价变化对我国经济的影响减到最小。

参考文献:

- [1]Bollerslev, T.Autoregressive Conditional Heteroskedasticity[J] . Journal of Econometrics, 1986, 31: 307~ 327。
- [2]高铁梅. 计量经济学分析方法与建模[M] . 清华大学出版社, 2005: 180-183。
- [3]Friedman J H, Stuetzle W. Projection pursuit regression[J].Journal American Statistics Association, 1981, 76: 817—823。
- [4]唐启义, 冯明光. DPS 数据处理系统[M] . 科学出版社, 2006: 940-941。
- [5]Palm,F,and Vlaar,P.Simple Diagnostics Procedures for Modelling Financial Time Series[J] . Allgemeines Statistisches Archiv, 81.1997: 85—101。
- [6]陈卓雷, 蒋寒迪. 基于 GARCH—BP 模型的股指预测及实证分析[J] . 当代财经, NO.6.2006。

**The fitted model of the international oil prices trend based on the
IGARCH projection pursuit regression**

Wang Jipei Xiao Hongwei

Abstract: As the dependence on oil imports increasing, the influence of international oil prices on China's economy is increasing. this paper take full account of the variability characteristics of the financial time series variance, establish three IGARCH models based on different distribution, and describe the trend of international oil prices, with the theory of the introduction of non-projection pursuit regression model parameters. Projection drop-dimensional analysis on the multi-dimensional data, the results show that the projection pursuit regression models based on IGARCH have strong capacity of fitted, China should take certain measures to reduce oil prices on China's economy and minimize the impact.

Keywords: international oil prices, IGARCH model, projection pursuit regression, financial time series

收稿日期: 2008-6-29

作者简介: 王吉培, 男, 1984. 05. 26, 汉族, 籍贯河北省邢台, 西南财经大学统计学院在读硕士研究生, 研究方向: 计量经济理论与应用研究, 通讯地址: 四川省成都市光华村西 55 号西南财经大学统计学院; 肖宏伟, 北京化工大学经济管理学院, 北京, 100029