### 4.4.2时间序列模型

ARMA 模型（Auto-Regressive and Moving Average Model）是研究时间序列的重要方法由自回归模型简称AR模型与滑动平均模型简称MA模型为基础“混合”构成，是时间序列模型的一种，是由美国统计学波克斯(Box)和金肯(Jenkins)在20世纪70年代提出，在预测领域有非常广泛的应用。

1. ARIMA油价预测模型

我们以2004年3月-2012年6月WTI月度原油价格序列数据建立ARIMA（8,1,0）模型来预测2012年7月至11月期间内原油价格。模型方程如下：



模型估计结果如表所示。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **模型統計資料** | | | | | | | | | | | | | |
| 模型 | 預測變數數目 | 模型適合度統計資料 | | | | | | | | Ljung-Box Q(18) | | | 離群值數目 |
| 平穩 R 平方 | R 平方 | RMSE | MAPE | MAE | MaxAPE | MaxAE | 標準化 BIC | 統計資料 | DF | 顯著性 |
| 油价-模型\_1 | 1 | .078 | .982 | 2.967 | 3.091 | 2.200 | 17.916 | 13.938 | 2.315 | 7.916 | 10 | .637 | 0 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ARIMA 模型參數** | | | | | | | | |
|  | | | | | 估計 | SE | T | 顯著性 |
| 油价-模型\_1 | 油价 | 無轉換 | 常數 | | 4.711 | 11.648 | .404 | .686 |
| AR | 落後 1 | .187 | .048 | 3.884 | .000 |
| 落後 2 | .014 | .049 | .296 | .767 |
| 落後 3 | .011 | .049 | .222 | .824 |
| 落後 4 | .036 | .049 | .730 | .466 |
| 落後 5 | .016 | .049 | .318 | .750 |
| 落後 6 | .027 | .049 | .559 | .577 |
| 落後 7 | .004 | .049 | .078 | .938 |
| 落後 8 | .170 | .048 | 3.515 | .000 |
| 差異 | | 1 |  |  |  |
| 时间序列 | 無轉換 | 分子 | 落後 0 | .000 | .000 | -.395 | .693 |

对样本内原油价格月度数据进行预测，得到样本区间内即2004年3月-2012年6月的月度结果，如图4.9所示，从样本内预测精度的指标上看，样本容量共436，该模型样本内的预测误差是2.97%左右，可以看出时间序列模型样本期内拟合效果较好。

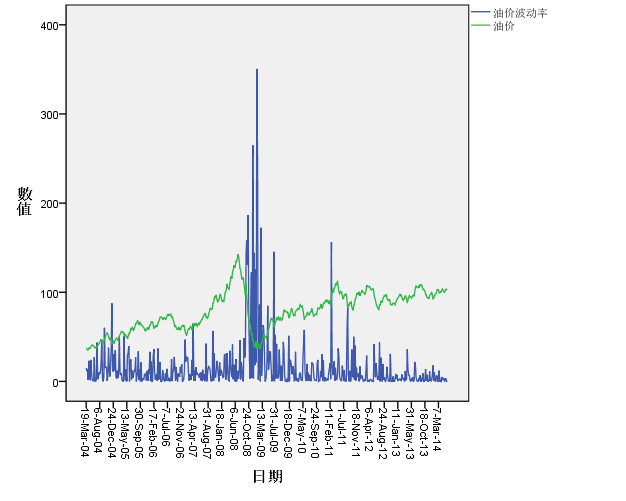
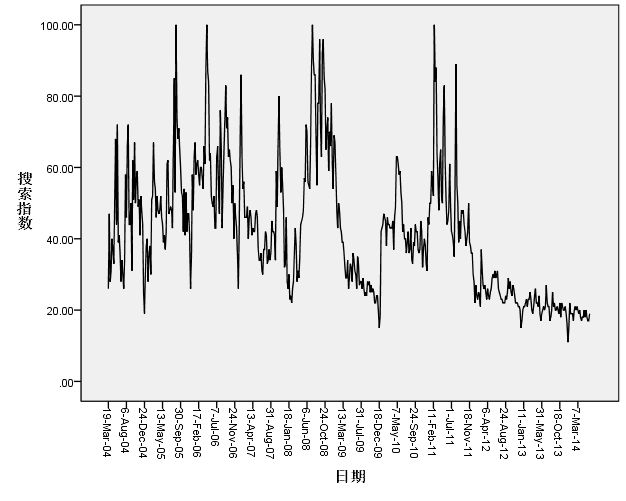
为检验该模型的预测效果，对样本外的原油价格进行预测，预测结果如下表所示。从预测精度的指标上看，样本外的预测误差为6.70%。预测精度显示无论样本内拟合情况还是样本外的预测情况，该回归模型的预测精度均在93%以上，预测效果较好。

1. 加入搜索指数的ARIMA油价预测模型

搜索数据准备：

为了准确反映互联网中对于原油价格的关注度，采用人工范围划定的方式选取了若干关键词，包括‘crude oil’, ‘WTI’, ‘crude oil price’, ‘NYMEX oil price’,以及‘NYMEX crude oil’。时间与油价的选取时间相同，为2004年3月-2012年11月。经过统计分析，发现各个关键词得到的搜索指数序列较为相似，序列间的线性相关系数较高。其中，‘crude oil’对应的搜索量最大，且包含更多的具体信息，因此选取‘crude oil’作为最能代表互联网搜索关注度的关键词。

下图为搜索指数以及原油价格波动的变化趋势，可以看到，搜索指数序列与原油价格波动序列具有一定的相似性，在2008年9月全球金融危机期间，油价变动值达到了一个高点，而期间的搜索指数也出现了明显的升高，表现了类似的趋势。



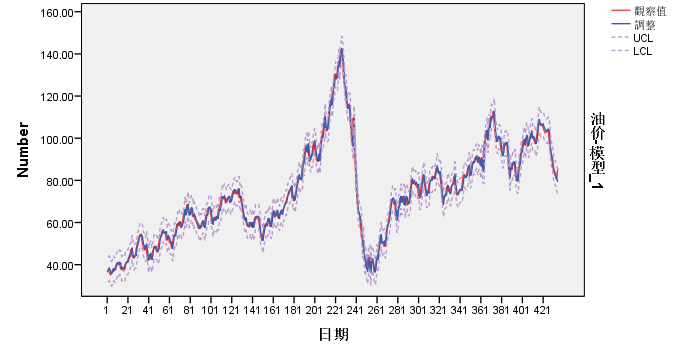
采用加入了搜索指数的ARIMA(8,1,0)模型来预测2012年7月至11月期间内原油价格。其模型方程为：



对样本内原油价格月度数据进行预测，得到样本区间内即2004年3月-2012年6月的月度结果，如图4.9所示，从样本内预测精度的指标上看，样本容量共436，该模型样本内的预测误差是2.95%左右，可以看出时间序列模型样本期内拟合效果较好。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **模型統計資料** | | | | | | | | | | | | | |
| 模型 | 預測變數數目 | 模型適合度統計資料 | | | | | | | | Ljung-Box Q(18) | | | 離群值數目 |
| 平穩 R 平方 | R 平方 | RMSE | MAPE | MAE | MaxAPE | MaxAE | 標準化 BIC | 統計資料 | DF | 顯著性 |
| 油价-模型\_1 | 1 | .090 | .983 | 2.947 | 3.115 | 2.210 | 18.007 | 13.775 | 2.301 | 6.671 | 10 | .756 | 0 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ARIMA 模型參數** | | | | | | | | |
|  | | | | | 估計 | SE | T | 顯著性 |
| 油价-模型\_1 | 油价 | 無轉換 | 常數 | | 1.369 | .550 | 2.490 | .013 |
| AR | 落後 1 | .169 | .048 | 3.507 | .001 |
| 落後 2 | .012 | .049 | .253 | .801 |
| 落後 3 | .006 | .049 | .131 | .895 |
| 落後 4 | .026 | .049 | .543 | .588 |
| 落後 5 | .004 | .049 | .075 | .940 |
| 落後 6 | .018 | .049 | .359 | .720 |
| 落後 7 | -.004 | .049 | -.091 | .928 |
| 落後 8 | .161 | .048 | 3.328 | .001 |
| 差異 | | 1 |  |  |  |
| 搜索指数 | 無轉換 | 分子 | 落後 0 | -.027 | .011 | -2.526 | .012 |



为检验该模型的预测效果，对样本外的原油价格进行预测，并与不加入搜索指数的模型对比，预测结果如下表所示。

表4.21: 原油价格序列的ARIMA模型估计结果-样本外

|  |  |  |
| --- | --- | --- |
|  | 模型1 | 模型2 |
| *MAE* | 5.80 | 4.95 |
| *MAPE* | 6.16 | 5.36 |
| *RMSE* | 6.70 | 5.73 |

从预测精度的指标上看，样本外的预测误差为5.73%。预测精度显示无论样本内拟合情况还是样本外的预测情况，该回归模型的预测精度均在94%以。与不加入搜索指数的模型相比样本外的预测精度提升了近1%，并且模型平稳的R方从0.73提升到0.90，说明了加入搜索的模型能更好刻画原油价格序列，搜索数据对于预测原油价格具有一定作用。