

中国第三产业贡献率时间序列分析

赵镇岳

信息管理学院 161070094

摘要 本研究通过使用季节性 ARIMA 模型对国民经济核算中 1992 年至 2018 年（第一季度）每一季度第三产业贡献率的变动情况进行分析，对中国第三产业的相对发展规律做出了大致的展示与评价，并讨论了季节性 ARIMA 模型在相关内容研究中应用的合理性。

0 引言

产业结构是一个经济体在一定技术条件下形成的专业化生产部门分工，并能在一定意义上决定经济增长的方式^[1]。早在 1949 年著名经济学家 Kuznets^[2]就指出，国民收入应从产业结构的角度去度量。一国国民经济的产业结构由以劳动密集型产业为主的低级结构，向以知识、技术密集型产业为主的高级结构调整和转变的过程及趋势被称作产业结构高级化。此过程的一个重要标志是国民经济发展使得第三产业比重日趋增大^[3]。对我国第三产业 GDP 贡献率的变动规律进行分析能够让我们对中国的产业结构高级化进程有一个直观的认识，以此支持有关部门后续的决策。

1 数据与方法

1.1 数据获取

笔者从中华人民共和国国家统计局国家数据数据库中下载得到了国民经济核算中 1992 年至 2018 年（第一季度）每一季度第三产业增加值增量与 GDP 增量之比，即第三产业贡献率（单位：%）。数据共有 105 期，无缺损值。方便起见，文中描述时均省略该单位。

1.2 方法性工具

1.2.1 ARMA 模型

ARMA^{[4][5]}（自回归滑动平均）模型是美国学者博克斯（George Box）和英国统计学家詹金斯（Gwilym Jenkins）共同建立的一种随机时间序列预测模型，也被称为博克斯-詹金斯法，简称 B-J 法。该模型的建立分为两个过程，即 AR（自回归）过程与 MA（滑动平均）过程。

AR 模型的基本思想为，如果时间序列 $\{Y_t\}$ 满足

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

（ ε_t 为误差项），且满足 $\sigma_\varepsilon^2 > 0, E(\varepsilon_t) = 0$ ，则称时间序列 $\{Y_t\}$ 服从 p 阶的自回归 AR(p) 模型， β_1, \dots, β_p 为自回归系数。

MA 模型的基本思想为，如果时间序列 $\{Y_t\}$ 满足

$$Y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}$$

(ε_t 为t期数据预测误差)，则称时间序列{Y_t}服从q阶移动平均MA(q)模型

因此，ARMA(p,q)模型表达式为：

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}$$

引入自回归算子 $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ 和移动平均算子 $\theta_q(B) =$

$1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ ，其中B为滞后算子，可将模型表示为：

$$\phi_p(B)Y_t = \theta_q(B)\varepsilon_t$$

1.2.2 SARIMA模型

如果序列为不平稳序列并存在趋势，且经过d阶逐期差分可以平稳，可以建立改进后的ARMA模型，即ARIMA(p,d,q)模型，d表示差分阶数，模型的一般形式如：

$$\phi_p(B)\nabla^d Y_t = \theta_q(B)\varepsilon_t$$

如果序列为不平稳序列并存在趋势和季节性变动，且可以通过逐期差分和季节差分平稳化，则可以用更加复杂的ARIMA(p,d,q)(P,D,Q)^s模型描述，模型的一般形式为：

$$\phi_p(B)\Phi_P(B^s)\nabla^d \nabla_s^d Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

其中，P为季节性自回归阶数，Q为季节性移动平均阶数， $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$ ， $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$ 。如果能通过样本数据确定 ϕ 、 Φ 、 θ 、 Θ 的取值，模型就能够被用于预测。

2 模型建立与评价

2.1 序列平稳性与差分

为了判断时间序列的平稳性，笔者借助使用数据绘制的折线图对数据序列进行了初步观察，折线图如图1所示。从时间序列发展的规律来看，其具有明显的季节性，笔者使用SPSS软件对其进行了季节性分解，季节周期为4。若将季节性时间序列的发展看作长期趋势(T)、季节周期(S)、不规则变动(I)三个主要因素共同作用的结果，则有两种模型可被用来对序列的发展规律进行分解，分别是加法模型：

$$y = T + S + I$$

与乘法模型：

$$y = T \times S \times I$$

其中，在乘法模型中，只有长期趋势是与y同计量单位的绝对量，其余因素均为以长期趋势为基础的比率。

如图 2 与图 3，在使用乘法模型分离季节性因子之后，我们可以发现时间序列有明显的上升趋势，这说明原时间序列同样有随着时间推移而上升的趋势。图 4 与图 5 分别展示了对原序列进行自相关、偏自相关检验的结果，序列的自相关、偏自相关程度印证了之前的猜测。以上情况均说明序列是不平稳的，且存在季节性变动的规律。因此，季节性差分的步骤是必不可少的。图 6 展示了对原序列进行一阶季节性差分后的结果，差分后序列的自相关偏自相关情况如图 7 与图 8 所示。

可以看到，序列的自相关函数值随滞后阶数增加呈现出正弦衰减趋势，且大部分落入置信区间内，这些现象支持了序列平稳的假设；但偏自相关函数值仍有部分落入置信区间外，这种现象可能是由数据总量过少引起的。为了进一步检验原序列经一阶季节差分后的平稳性，笔者使用 R 语言对原序列进行了 ADF 检验，输出结果如下：

```
> adf.test(diff)
    Augmented Dickey-Fuller Test
data: diff
Dickey-Fuller = -3.9588, Lag order = 4, p-value = 0.01393
alternative hypothesis: stationary
```

检测结果p值很低，这支持了备选假设，即序列平稳。序列或差分后的序列平稳是使用ARIMA(p,d,q)(P,D,Q)^s模型进行预测的前提条件。以上过程也说明模型中季节性差分的阶数D等于1。

2.2 模型识别及定阶

平稳时间序列的 ARMA 模型定阶可以通过判断自相关、偏自相关函数分布来粗略完成，主要对应关系如下：

模型	AR (p)	MA (q)	ARMA (p, q)
模型表示	$\phi_p(B)Y_t = \varepsilon_t$	$Y_t = \theta_q(B)\varepsilon_t$	$\phi_p(B)Y_t = \theta_q(B)\varepsilon_t$
自相关函数	拖尾	截尾 (q 步)	拖尾、指数衰减和 (或) 正弦衰减
偏自相关函数	指数衰减和 (或) 正弦衰减、截尾	拖尾、指数衰减和 (或) 正弦衰减	拖尾、指数衰减和 (或) 正弦衰减

此外，还有一些准则可作为定阶依据，如 AIC 准则（赤池信息准则）、BIC（贝叶斯信息准则）等。

从自相关、偏自相关函数值分布情况来看，考虑到数据总量少可能带来的误差，自相关函数 1 步之后开始呈现正弦衰减规律，故序列符合 AR 模型要求， p 定为 1；与季节有关的阶数确定则是以相同法则，通过判断与季节周期数有关的滞后期数对应的函数值来确定的， P 被定为 2。类似地，此处 q 被定为 0， Q 被定为 1。结合之前确定的差分期数，模型暂时被确定为

$$ARIMA(1,0,0)(2,1,1)^4$$

而后，笔者使用 R 语言 forecast 包中提供的 auto.ARIMA() 函数对序列依据 AIC、BIC 准则进行了定阶尝试。输出结果如下：

```
> auto.arima(ts,trace=T)
```

ARIMA(2,0,2)(1,1,1)[4] with drift	: Inf
ARIMA(0,0,0)(0,1,0)[4] with drift	: 618.7499
ARIMA(1,0,0)(1,1,0)[4] with drift	: 580.2695
ARIMA(0,0,1)(0,1,1)[4] with drift	: Inf
ARIMA(0,0,0)(0,1,0)[4]	: 622.0981
ARIMA(1,0,0)(0,1,0)[4] with drift	: 596.5554
ARIMA(1,0,0)(2,1,0)[4] with drift	: 576.0045
ARIMA(1,0,0)(2,1,1)[4] with drift	: Inf
ARIMA(0,0,0)(2,1,0)[4] with drift	: 611.3209
ARIMA(2,0,0)(2,1,0)[4] with drift	: 577.4112
ARIMA(1,0,1)(2,1,0)[4] with drift	: 577.5252
ARIMA(2,0,1)(2,1,0)[4] with drift	: 579.6805
ARIMA(1,0,0)(2,1,0)[4]	: 578.1516

Best model: ARIMA(1,0,0)(2,1,0)[4] with drift

Series: ts

ARIMA(1,0,0)(2,1,0)[4] with drift

Coefficients:

	ar1	sar1	sar2	drift
	0.5808	-0.5353	-0.2543	0.3049
s.e.	0.0842	0.0999	0.0975	0.1320
sigma^2 estimated as 16.19: log likelihood=-282.69				
AIC=575.37 AICc=576 BIC=588.45				

该函数从 AIC、BIC 准则的角度出发建议将模型确定为 $ARIMA(1,0,0)(2,1,0)^4$ ，可以看出，由于滑动平均阶数均为 0，这其实是一个 SARIMA 模型退化成的季节性 AR 模型。

至此两个备选模型已经产生，笔者使用两个模型分别对原序列进行了拟合，曲线拟合效果分别见图 9 与图 10，输出结果分别如下所示：

$$ARIMA(1,0,0)(2,1,0)^4$$

ARIMA 模型參數

		估計	SE	T	顯著性
第三产业当季	第三产业當無轉常數	1.220	.539	2.264	.026
贡献率-模型_1	季貢獻率換 AR 落後 1	.581	.084	6.883	.000
	AR, 落後 週期性 1	-.535	.101	5.310	.000
	落後 2	-.254	.101	2.512	.014
	週期性差異 1				

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	統計資料	DF	顯著性	
第三产业当季貢獻率-模型_1	0	.392		37.029	15	.001	0

可以看出，各个参数都显著不为零，且模型整体显著性水平低。不足之处在于平稳的 R 方不高，但在变动规律性不强的时间序列模型中一味追求 R 方很可能会导致过拟和现象的出现。从图像来看，曲线的拟合效果较好，模型整体而言是有价值的。

$$ARIMA(1,0,0)(2,1,1)^4$$

ARIMA 模型參數

		估計	SE	T	顯著性	
第三产业貢獻率-模型_1	第三产业貢獻率	無轉換常數	1.064	.137	7.754	.000
		AR 落後 1	.685	.081	8.422	.000
		AR, 週期性 落後 1	.071	.123	.576	.566
		落後 2	-.040	.120	-.333	.740
		週期性差異 1				
		MA, 週期性 落後 1	.995	1.299	.765	.446

模型統計資料

模型	預測變數數目	模型適合度統計		Ljung-Box Q(18)			離群值數目
		資料	平穩 R 平方	統計資料	DF	顯著性	
第三产业贡献率-模型_1	0	.511	.511	26.334	14	.023	0

可以看出，部分参数显著性较高，模型整体显著性较低，平稳的 R 方较之前更高。经检验，该模型带有常数时效果更好，故此处不讨论不带常数的情况。为了考察两个模型的泛化能力，我删除了原序列最后三期数据并对模型重新进行拟合，并使用其分别对最后三期数据进行预测。如图 11 与图 12 所示。两个模型都成功预测了曲线上升的趋势，但 $ARIMA(1,0,0)(2,1,0)^s$ 的预测效果更好。

综上所述，笔者最终选择的模型为 $ARIMA(1,0,0)(2,1,0)^s$ 模型，拟合的曲线表达式为：

$$(1 - \phi_1 B)(1 - \Phi_1 B^4 - \Phi_2 B^8)(1 - B^4)Y_t = c$$

代入参数展开后得：

$$\begin{aligned} Y_t = & 0.581Y_{t-1} + 0.465Y_{t-4} - 0.270Y_{t-5} + 0.281Y_{t-8} - 0.163Y_{t-9} + 0.254Y_{t-12} \\ & - 0.147Y_{t-13} + 1.220 \end{aligned}$$

2.3 模型检验

为评价模型的优劣，笔者对模型的残差序列进行了检验。残差序列若为白噪声，模型可被认为合理并适于预测，否则说明残差序列中还有某种有规律的信息，模型质量不高。从图 13 可见，残差的自相关、偏自相关函数值大多都落在置信区间内，序列自性相关与 0 无显著性不同，此种现象支持序列为白噪音的假设。图 14 为残差序列的正态分布 P-P 图，由图可见残差基本服从正态分布。以上现象均说明残差为白噪音序列，模型合理且可以投入使用。

3 结果与讨论

笔者使用模型对 2018 年其余季度至 2023 年第四季度的第三产业贡献率进行了预测，预测结果如图 15 所示。

模型预测结果显示，若不受外界原因显著干扰，我国之后五年内第三产业贡献率仍将继续上升，但上升趋势将减缓。刘伟^[1]在对中国经济的增长模式进行实证分析后指出，能最有效拉动中国经济增长的产业几乎全部是第三产业的部门，其中以地质勘察、水利管理、体育卫生等为代表。以上迹象表明，中国产业的升级进程已经取得了长足的发展，且进入了平稳增长的时期。

但同时仍有学者指出，在与发达国家进行横向对比的过程中，中国经济的比较优势仍是劳动力密集型产业，且近几十年的快速发展并没有让要素禀赋产

生质变^[6]。因此，国内三次产业应共同抓住发展机遇，在追求自身质量提升的同时，协同带动产业结构的转变。

4 研究中的不足与思考

ARMA 本质是自回归（AR）模型与滑动平均（MA）模型的组合，它被用来对随机变量序列的变动规律进行拟合以对未来数据进行预测。笔者引入该模型对中国第三产业贡献率季度数据进行了分析，在操作过程中发现该类模型并不适合对变动频繁且规律性不强的数据进行预测，且不适合进行长期预测。

其根本原因显而易见：AR 模型与 MA 模型的基本行为相同，都是基于序列自身数据的挖掘。模型而并没有试图建立序列变动与外部任何变量之间的联系，因此它们无法揭示影响序列发展的根本因素（或者说，序列变动的内在原因）。

具体而言，AR 模型内部存在一种类似条件概率的思想，即通过之前数据的分布情况推断出本期数据的值；MA 模型的核心则是“反馈”，本期数据的预测误差被用来修正下一期数据的预测值（也正因此，每当序列发生方向性变动，预测值总会需要短暂的“反应时间”来进行调整）。

当序列变动的规律性不强时，条件概率机制便会失效，AR 模型的性能将大大降低；当序列变动频繁时，由于反馈机制自身具有滞后性，MA 模型无法将滞后及时消化，误差将“如影随形”，无法被摆脱。

从本文中最终确定的模型对近五年数据的预测结果来看，模型在失去真实数据的控制之后，所预测出的序列很快会进入一种“自我规划”状态。受“惯性”作用，序列开始以平稳的轨迹推进，并不再会产生较大的上下的波动或趋势性的变动。该平稳轨迹反映出的其实是因变量在外界条件不再变化时的一种理想化发展状况。这在某些业务背景下，是有一定作用的，但在绝大多数情况下并没有太大现实意义。因此，在一般情况下，该模型只适合在序列的某一阶段发展趋势确定的情况下，进行短期预测。

参考文献

- [1] 刘伟, 李绍荣. 产业结构与经济增长[J]. 中国工业经济, 2002(5):14-21.
- [2] Kuznets S. National Income and Industrial Structure[J]. Econometrica, 1949, 17:205-241.
- [3] 李江帆, 曾国军. 中国第三产业内部结构升级趋势分析[J]. 中国工业经济, 2003(3):34-39.
- [4] 王燕编著. 应用时间序列分析[M]. 2015
- [5] 史代敏, 谢小燕主编. 应用时间序列分析[M]. 2011
- [6] 李钢, 廖建辉, 向奕霓. 中国产业升级的方向与路径—中国第二产业占 GDP 的比例过高了吗 [J]. 中国工业经济, 2011(10):16-26.

附：全部图片

图 1 第三产业贡献率序列折线图

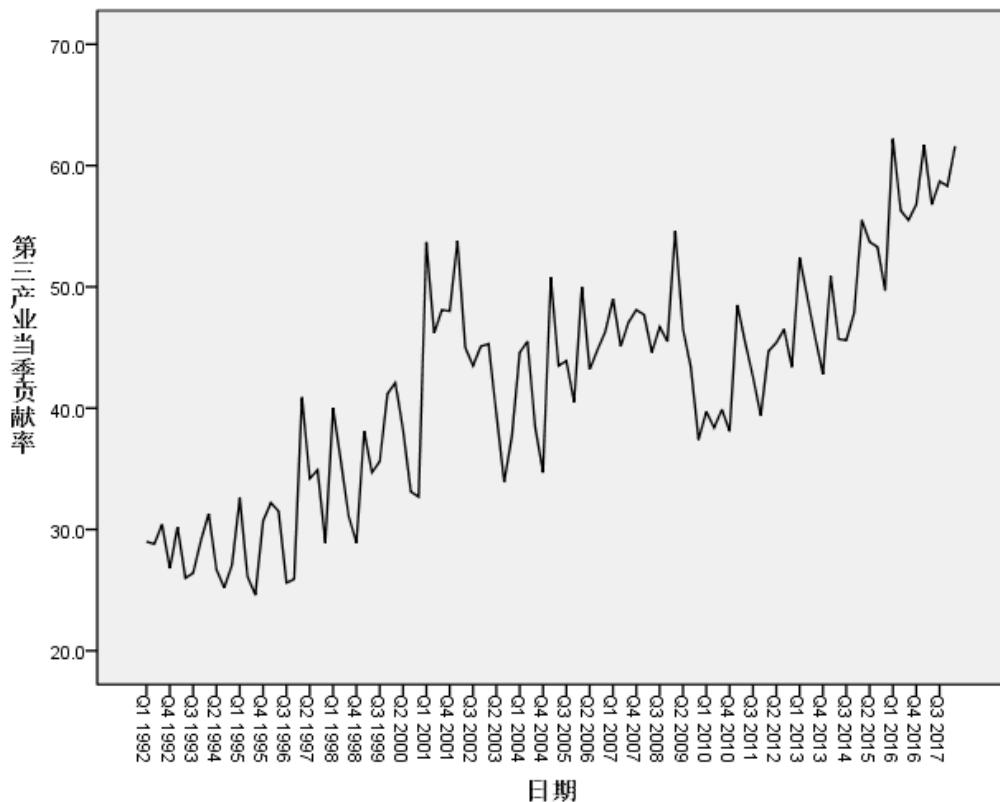


图 2 季节性因子

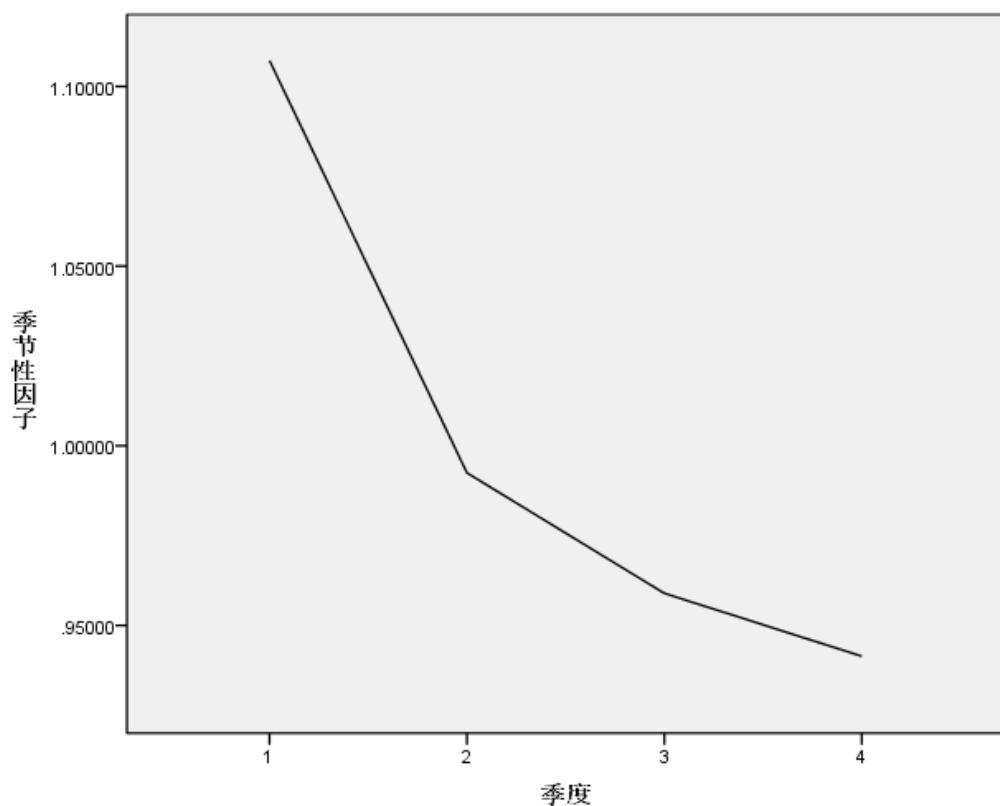


图 3 分离季节因素后的第三产业贡献率序列

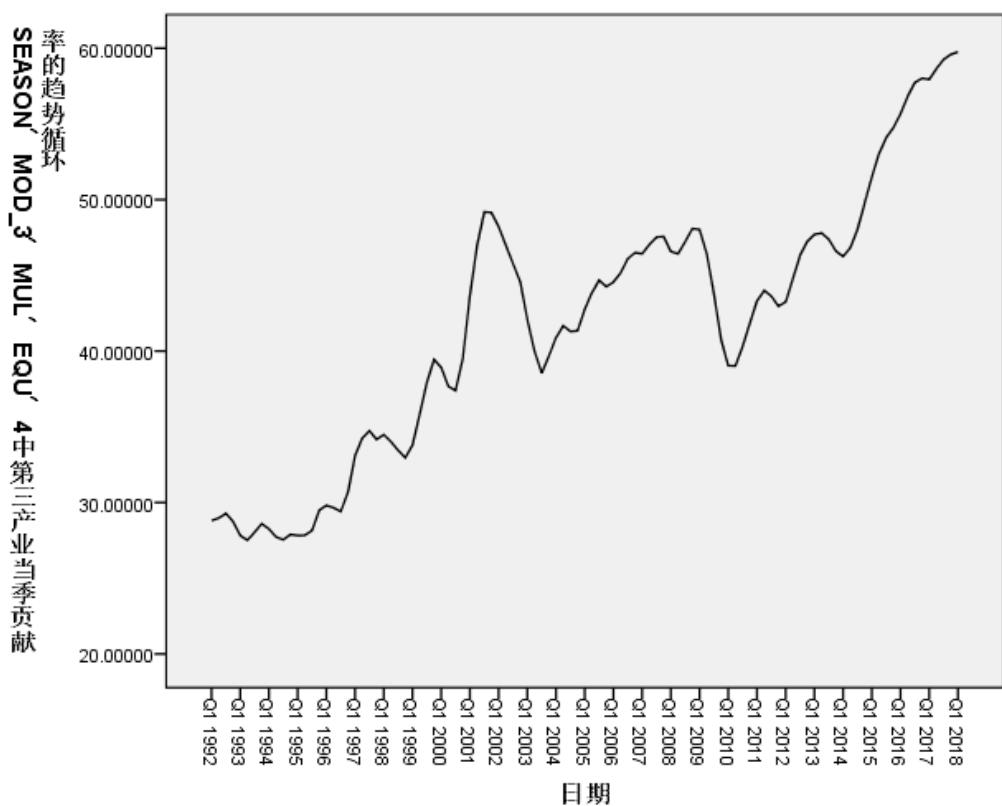


图 4 第三产业贡献率序列自相关检验

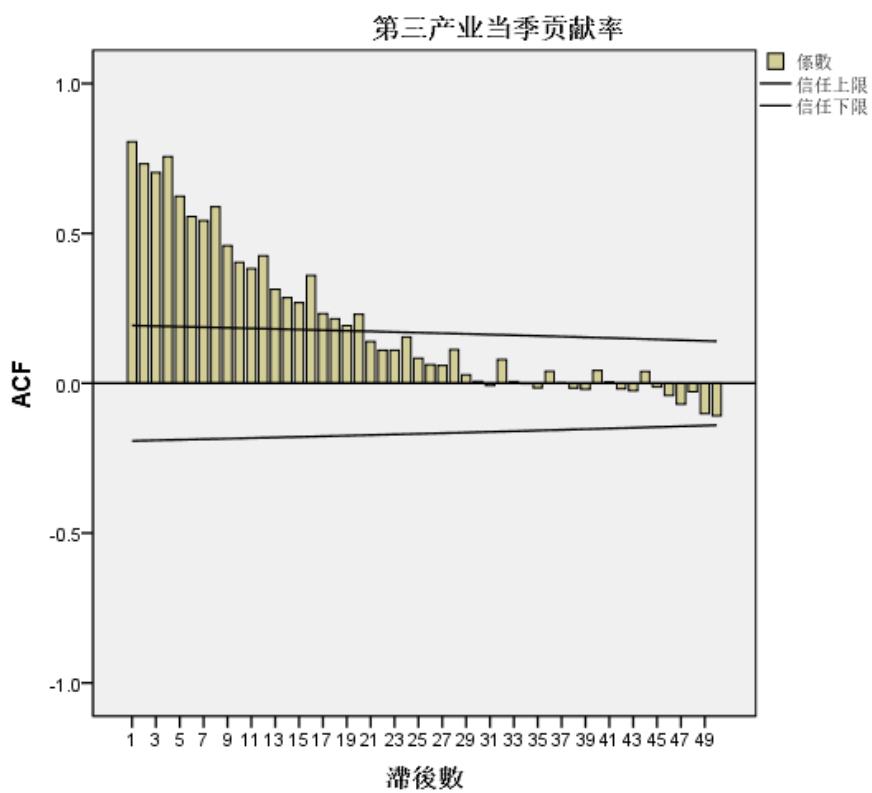


图 5 第三产业贡献率序列偏自相关检验

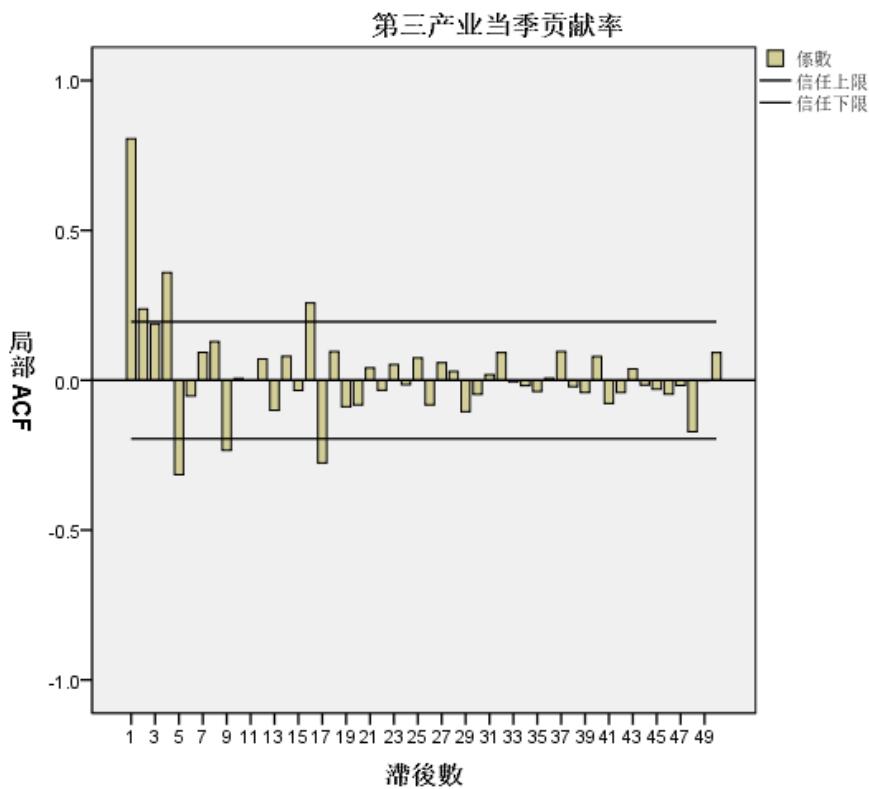


图 6 一阶季节性差分后的第三产业贡献率序列

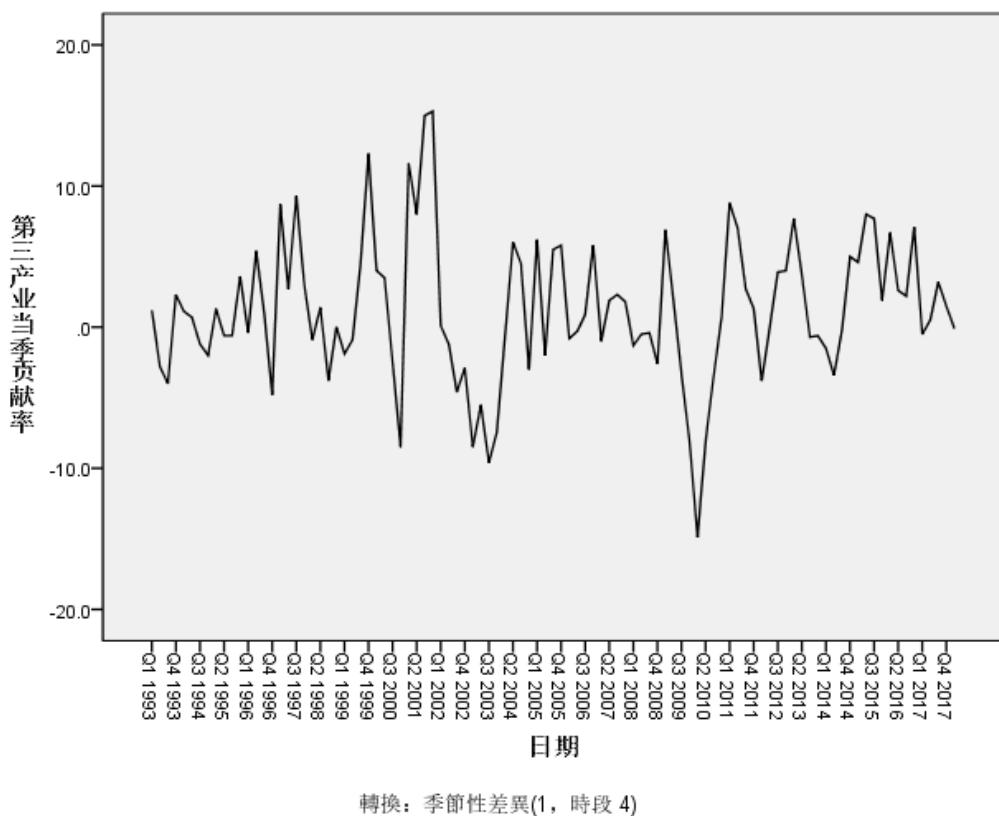


图 7 差分后的第三产业贡献率序列自相关检验

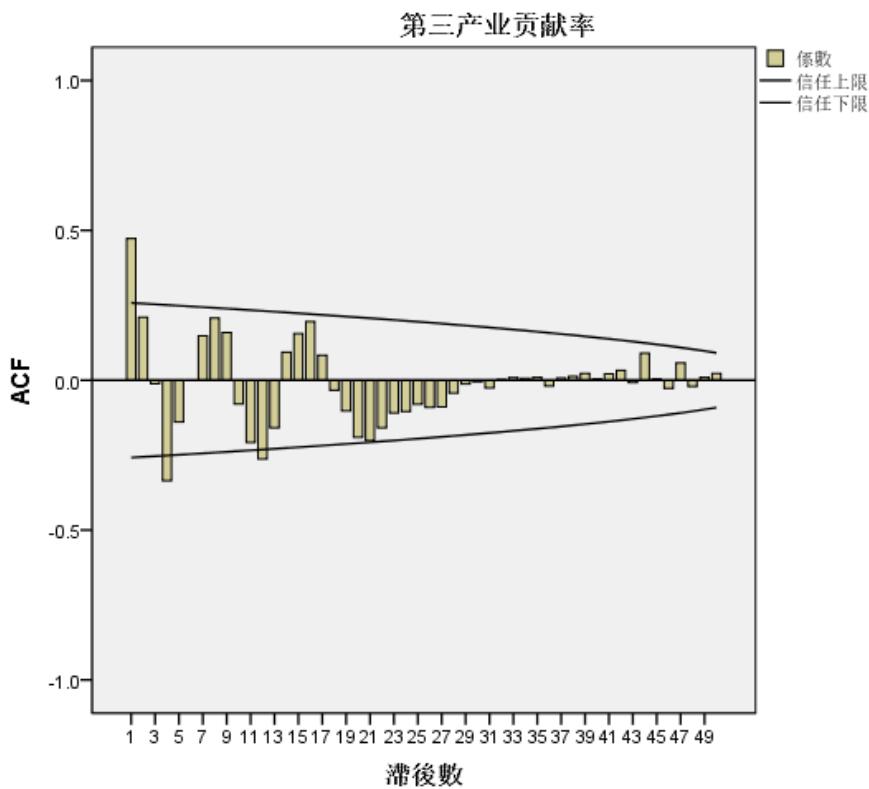


图 8 差分后的第三产业贡献率序列自相关检验

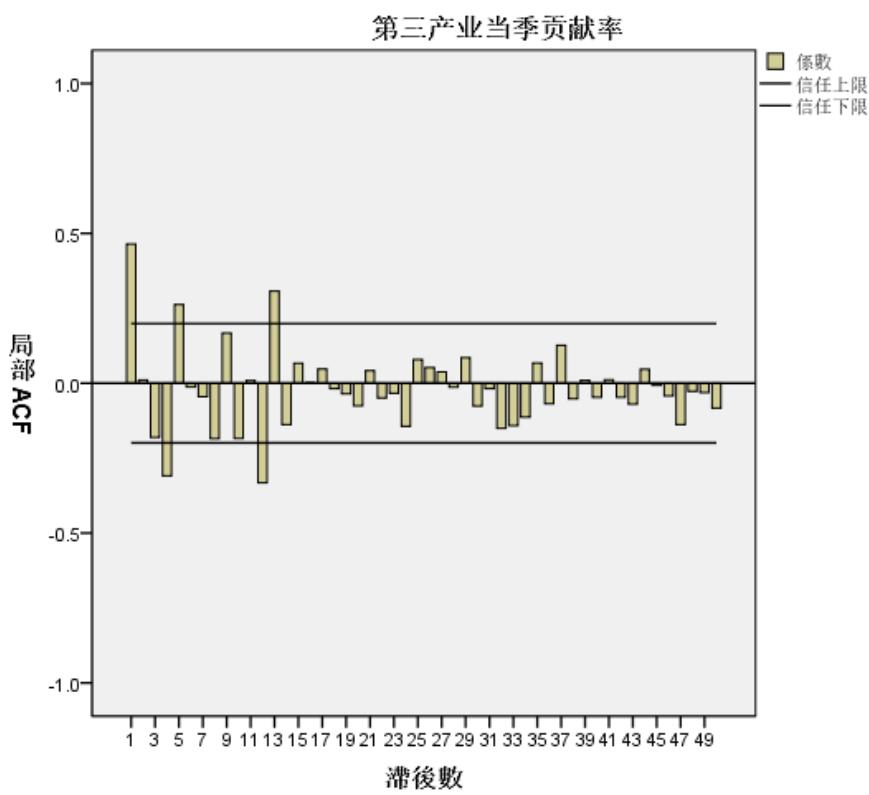


图 9 $ARIMA(1,0,0)(2,1,0)^4$ 模型拟合效果

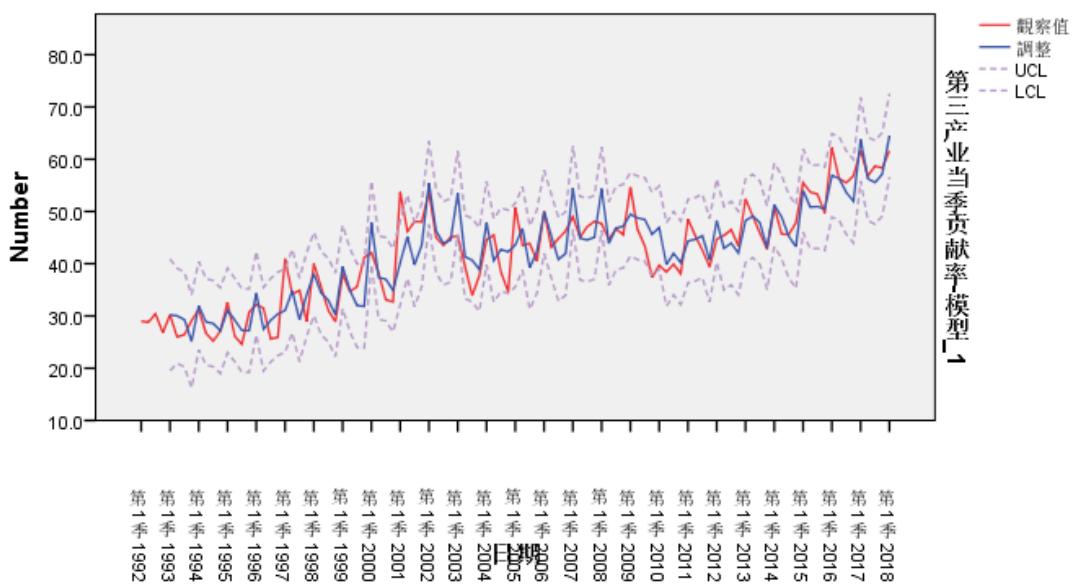


图 10 $ARIMA(1,0,0)(2,1,1)^4$ 模型拟合效果

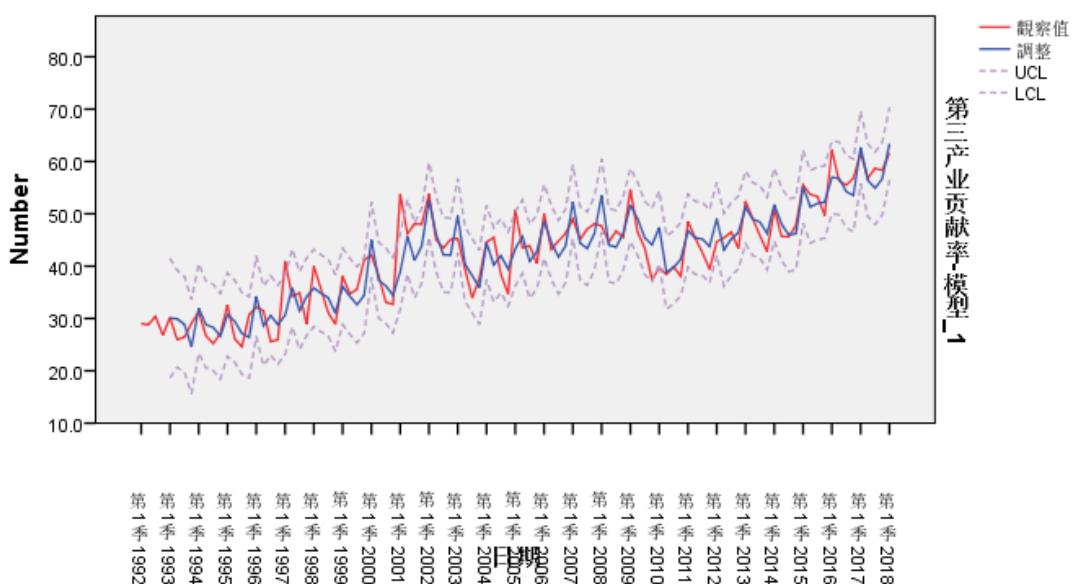


图 11 $ARIMA(1,0,0)(2,1,0)^4$ 模型模拟预测效果

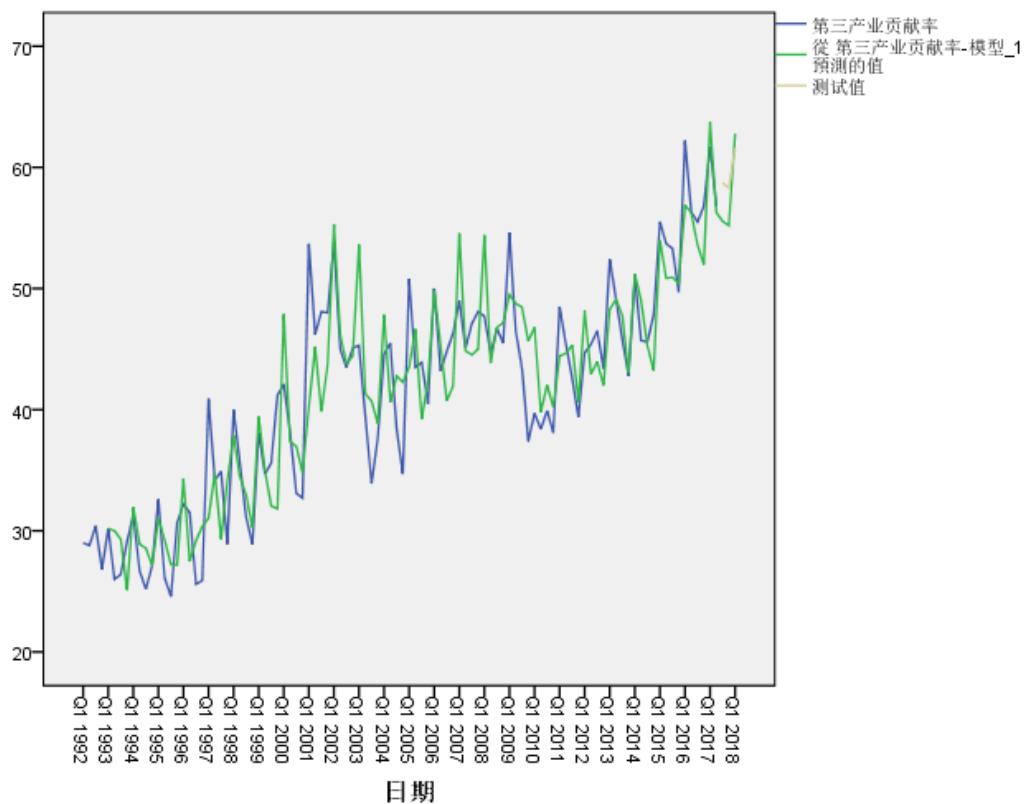


图 12 $ARIMA(1,0,0)(2,1,1)^4$ 模型模拟预测效果

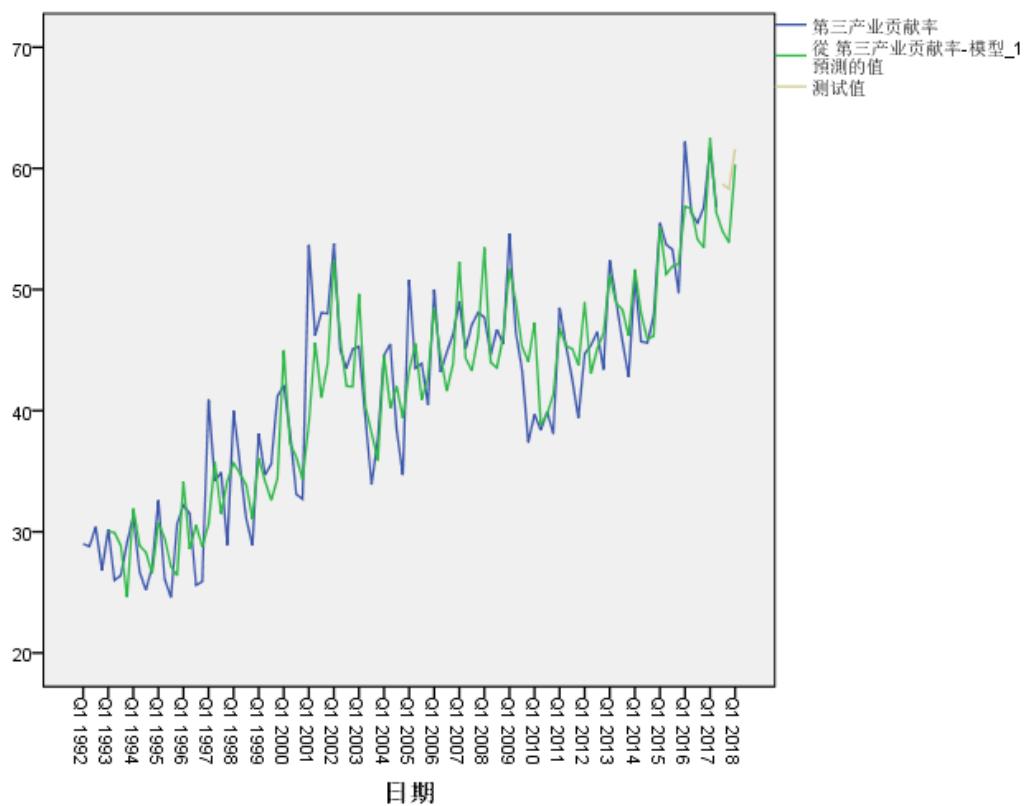


图 13 ARIMA(1,0,0)(2,1,0)⁴模型残差检验

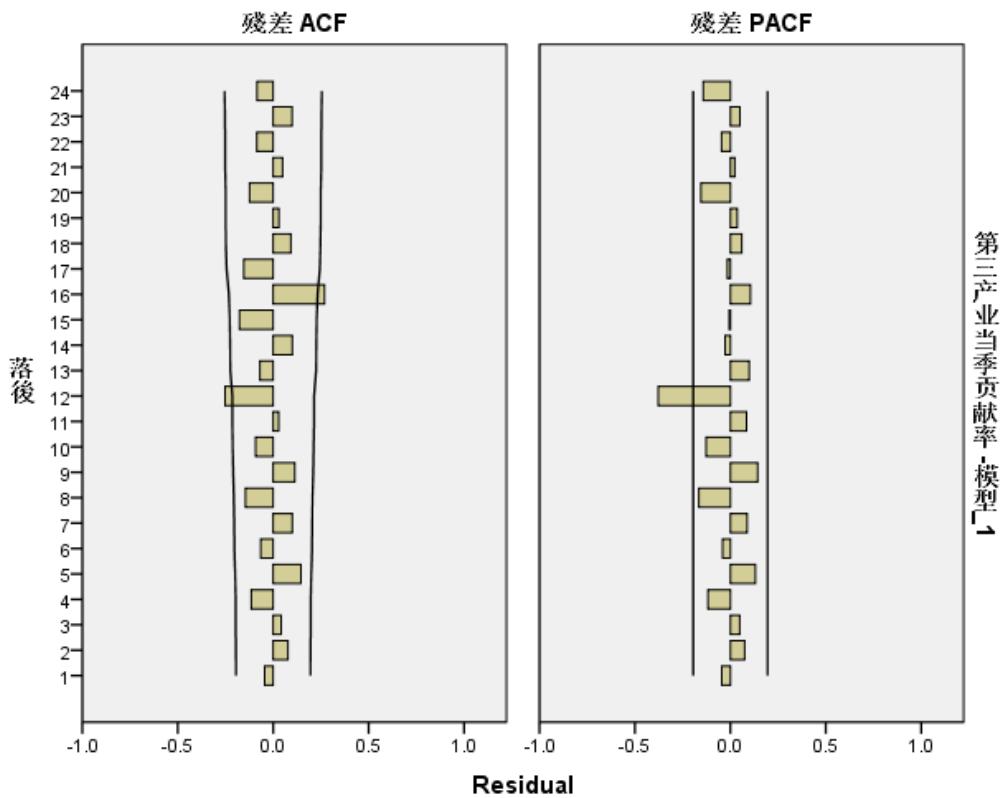


图 14 ARIMA(1,0,0)(2,1,0)⁴模型残差 P-P 图

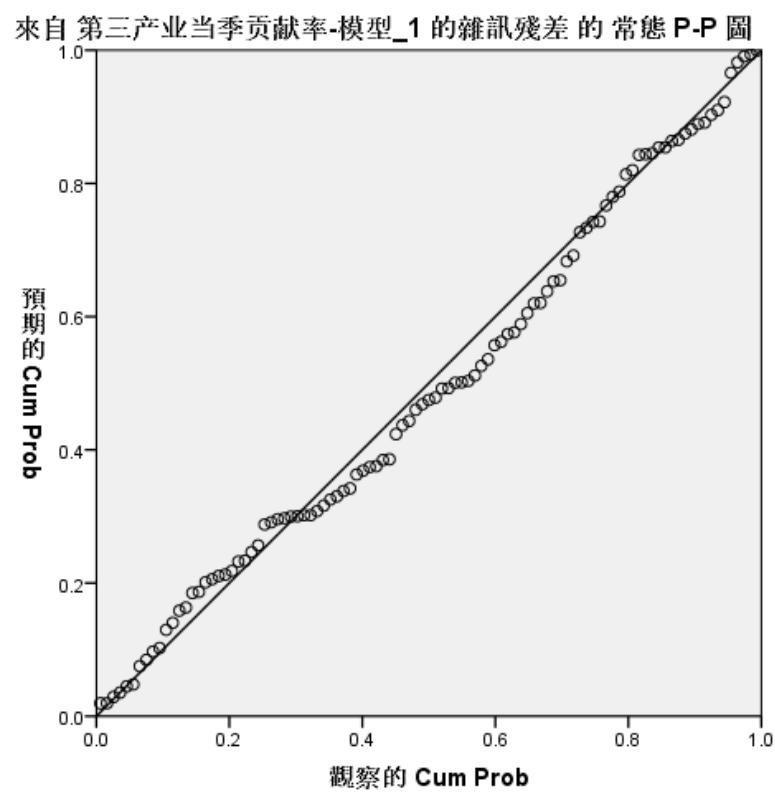


图 15 $ARIMA(1,0,0)(2,1,0)^4$ 模型预测结果

