

基于传递函数模型的油田产油量预测方法

王滨¹, 赵伟¹, 计小宇², 张以根¹, 侯春华¹, 颜子¹, 邢绍猷¹, 彭国红¹

(1.中国石化胜利油田分公司地质科学研究所, 山东东营 257015; 2.中国人民大学商学院, 北京 100872)

摘要:针对油田产油量预测方法无法考虑因素动态关系的局限性,首次应用时间序列传递函数模型建立了考虑因素动态关系的多因素油田产油量预测模型。应用传递函数模型的原理并按照建模步骤,以胜利油区某油田新井产油量预测为例,说明了传递函数模型数据准备、模型识别、参数估计、诊断校验及预测过程。研究结果表明,该预测模型既能在繁杂因素中剔除无关因素的干扰,识别影响油田产油量预测的主控因素,又能考察变化非同步且具有滞后性的因素动态关系;拟合程度好,预测精度高达98.4%。

关键词:产油量预测 时间序列 传递函数模型 互相关分析 主控因素

中国分类号:TE319

文献标识码:A

文章编号:1009-9603(2014)05-0077-04

考虑统计对象和信息来源2个方面,预测油田产油量的方法可分为基于单因素的预测和基于多因素的预测。其中,基于单因素的预测主要采用曲线拟合法^[1-5],但因其往往只考虑时间因素,无法研究各种影响因素的作用,因此多用于常规油田产油量预测;基于多因素的预测考虑了开发指标与其他相关影响因素的关系,一般有多个输入变量和一个输出变量,可采用多元线性模型^[6]、神经网络模拟^[7]和支持向量机^[8]等方法,但其不足之处是建立的模型均为静态模型,影响因素的关系是在同一时间截面上的。然而,实际上一个因素的变化对另一个因素产生的影响并不一定会立刻显现,原因导致结果的发生需要一定的过程,因素之间的影响可能非同步且具有滞后性。对因素之间的这种动态关系,前人所研究的模型都无法描述。笔者采用时间序列分析中传递函数模型进行油田产油量预测。与其他产油量预测方法对比,该模型既能在繁杂因素中剔除无关因素的干扰,识别影响油田产油量的主控因素,又能考察变化非同步且具有滞后性的因素的动态关系;该模型产油量预测精度高,适应性强,可作为一种新的方法应用于油田产油量预测中。

1 传递函数模型的建立

1.1 建模原理

传递函数模型(ARIMAX)是一种多变量时间序

列分析方法,从形式上可以说是时间序列(ARIMA)模型与回归分析模型的结合。它综合了时间序列与回归分析的优点,不但可以考虑变量与自身历史的关系,还可以考虑其他变量与所研究变量的关系。这种关系,并不是简单的、同一时间截面上的一对一的关系,而是考虑了变量之间在时间上具有领先或滞后特点的动态关系。因此传递函数模型也称为动态回归模型。

在多变量时间序列分析中将因变量序列视为一个动态系统的输出,自变量序列视为该系统的输入。如果系统只有一个输入变量,则单变量的传递函数模型表达式^[9]为

$$y_t = \mu + \frac{\omega(B)B^b}{\delta(B)}x_t + n_t \quad (1)$$

其中

$$\omega(B) = (\omega_0 + \omega_1 B + \dots + \omega_s B^s) \quad (2)$$

$$\delta(B) = (1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r) \quad (3)$$

式中: t 为时间,月; y_t 为因变量,即需要预测的时间序列; μ 为方程常数项的均值,可以根据模型的实际情况确定是否必要; $\omega(B)$ 和 $\delta(B)$ 为待估计模型系数简化表达式; B 为后移算子; x_t 为自变量序列; n_t 为模型残差或噪声; b, s, r 为模型对应的参数,可通过互相关函数识别。

多变量传递函数的表达式为

$$y_t = \mu + \sum_i \frac{\omega_i(B)}{\delta_i(B)} B^{b_i} x_{i,t} + n_t \quad (4)$$

收稿日期:2014-07-15。

作者简介:王滨,女,高级工程师,硕士,从事油气田开发及中长远开发规划工作。联系电话:(0546)8717970, E-mail:wangbin265.slyt@sinopec.com。

基金项目:国家科技重大专项“胜利油田特高含水期提高采收率技术”(2011ZX05011-002)。

式中: $\omega_i(B)$ 和 $\delta_i(B)$ 为变量 x_i 的待估计模型系数简化表达式; B^k 为变量 x_i 的滞后算子; $x_{i,t}$ 为时间为 t 时的变量 x_i 。

对于初步建立的传递函数模型, 残差序列并不一定是白噪声, 还可以进一步提取的信息是识别出残差的时间序列模型。修正后最终的传递函数模型为

$$y_t = \mu + \sum_i \frac{\omega_i(B)}{\delta_i(B)} B^k x_{i,t} + \frac{\theta(B)}{\phi(B)} a_t \quad (5)$$

式中: $\theta(B)$ 为移动平均算子; $\phi(B)$ 为自回归算子; a_t 为白噪声序列。

建立传递函数模型的关键是识别对因变量有影响的自变量, 以及每一个自变量对因变量的影响方式, 即模型中参数 r, s, b 的确定。其中参数 b 是因变量与自变量之间在时间上的滞后数值, 反映了二者之间的动态关系。对于每一个自变量, r, s, b 的识别是通过观察因变量与自变量的互相关函数图来判断的; r, s, b 确定以后再估计模型的参数, 要求所有参数均显著; 当输出变量与输入变量之间的关系确定后, 还需要检验对应模型的残差, 直到整个模型的残差可以被判定为白噪声为止; 当模型残差与自变量之间的互相关检验判定为无相关性时, 说明变量中全部信息已经被提取完毕, 模型拟合充分。

1.2 建模步骤

建立传递函数模型主要包括4个步骤: ①数据的准备以及序列的预白噪声化。初步分析与因变量有关的自变量, 准备好相应数据, 并对自变量进行预白噪声化。②因变量与自变量的互相关分析, 初步识别传递函数模型。对自变量和因变量进行互相关分析, 剔除不相关自变量, 对保留进入模型的自变量识别对应参数。③传递函数模型的估计。估计模型参数, 检验模型参数的显著性, 并以此为依据不断修正模型直到模型系数均显著。④传递函数模型的校验。通过模型残差的自相关检验, 建立残差的时间序列模型来修正传递函数模型, 直到模型残差为白噪声, 并检验残差序列与自变量的相关性, 确定最终合理的模型。

2 应用实例与分析

2.1 数据来源

选取胜利油区某油田1990—2012年开发指标相关数据。其中1990—2011年的数据作为历史数

据用来验证模型的拟合程度, 而2012年的数据作为对比数据用来验证模型的预测精度。以月度数据为研究对象, 用来增加模型样本点, 提高预测精度。预测的目标因变量为新井单井月平均产油量, 在得到新井平均单井月产油量之后, 给定新井井数, 并将每月产油量累加就可以得到年产油量。

根据实际开发经验, 初步确定对新井单井月平均产油量产生影响的因素。其中模型的自变量主要包括反应开发阶段的综合含水率和新井含水率; 反应物质基础的单井控制地质储量和地质储量采出程度; 反应能量补充状况的月注水量; 反应开发条件的投产新井井数。应用月度数据就会出现不同月份结果的差异性, 因此必须引入12个代表月份的虚拟变量。该模型的计算全部通过统计软件SAS实现。

2.2 数据的标准化及预白噪声化

模型中不同因素之间的量纲和数量级相差较大, 相互之间无可比性, 对后续模型参数的估计影响较大, 因此需对数据进行无量纲化处理, 处理方法为标准化方法, 具体表达式为

$$x_i = \frac{x_i - \bar{x}_i}{\sigma(x_i)} \quad (6)$$

式中: x_i 为原自变量序列; \bar{x}_i 为原自变量序列的平均值; $\sigma(x_i)$ 为原自变量序列的标准差。

建立传递函数模型的关键是识别对因变量有影响的自变量, 可以通过变量间的互相关分析实现。任何一个时间序列都存在较强的自相关性, 即便是2个从理论上没有任何内在联系的时间序列也存在较强的自相关性。将2个近似线性关系的时间序列画在一起, 其线性相关关系明显, 但以此为依据分析的结果是不可靠的, 必须消除变量自相关对互相关分析的影响, 其方法就是对自变量进行预白噪声化处理。对具有反馈现象的系统, 即自变量受到因变量影响的系统, 建立传递函数模型, 并对因变量进行预白噪声化^[10-11]。该模型的优点是各种因素通过互相关分析, 剔除变量自相关的干扰, 找出真正影响预测对象的主控因素。

模型数据选取月度数据, 其具有明显的趋势性和季节性, 是非平稳时间序列。经过12个月季节差分后, 新井投产井数为ARIMA(1,0,0)(1,1,0)¹², 单井控制地质储量为ARIMA((0,12),1,(0,3,12)), 地质储量采出程度为ARIMA((0,12),1,(0,12)), 月注水量为ARIMA((0,11),1,1)(0,1,1)¹², 综合含水率为ARIMA(1,0,0)(1,1,0)¹², 新井含水率为ARI-

MA(1,0,0)(1,1,0)¹²,12个代表月份的虚拟变量是不需要预白噪声化。

2.3 因变量与自变量的互相关分析

自变量预白噪声化后,即可以进行自变量与因变量的互相关分析。分析结果表明:①投产井数与新井月平均产油量的互相关系数呈指数形式衰减,滞后1期和滞后2期的互相关系数分别为-0.2816和-0.16,均超过2倍标准差,显著不为0,说明投产井数领先新井月平均产油量变化,领先时间为1个月。投产井数对应的模型参数 r,s 和 b 分别为1,0和1。②新井含水率与新井月平均产油量同期的互相关系数为-0.2316,显著不为0,其他时期的互相关系数在2倍标准差之内。新井含水率对应的模型参数 r,s 和 b 皆为0。③综合含水率与新井月平均产油量在滞后12期的互相关系数为-0.1683,超过2倍标准差,显著不为0,其他时期的互相关系数在2倍标准差之内。综合含水率对应的模型参数 r,s 和 b 分别为0,0和12。

互相关分析结果表明,单井控制地质储量、地质储量采出程度、月注水量以及12个虚拟变量与月平均产油量的互相关系数均在2倍标准差之内,相关性不显著。因此,将这些变量剔除,最后仅选择投产井数、新井含水率和综合含水率3个自变量进入模型,其中2个变量的参数 b 都大于0,说明自变量的影响有滞后期,如油田综合含水率的滞后期为12个月,综合含水率是油田层面的宏观变量,需要12个月的时间才要达到影响新井产油量的程度,与油田投产新井的含水率大约12个月能上升到开发区块同等水平的规律是一致的。

2.4 传递函数模型估计

根据互相关分析结果初步建立传递函数模型表达式为

$$(1-B^{12})y_t = \omega_{1,0}(1-B^{12})zshs_{t-12} + \omega_{2,0}(1-B^{12}) \times xjhs + \frac{\omega_{3,0}}{(1-\delta_{1,1}B)}(1-B^{12})tcjs_{t-1} + n_t \quad (7)$$

式中: $\omega_{1,0}$ 为综合含水率对应的模型待估计系数; $zshs$ 为标准化的综合含水率; $\omega_{2,0}$ 为新井含水率对应的模型待估计系数; $xjhs$ 为标准化的新井含水率; $tcjs$ 为标准化投产井数; $\omega_{3,0}$ 和 $\delta_{1,1}$ 为投产井数对应的模型待估计系数。

建模时对因变量进行了差分,使其平稳化,模型中不含常数项,可采用SAS软件对模型进行估计(表1)。从表1中可以看出,模型系数 T 检验值对应 P 值都小于0.05,即在0.05的显著水平下系数都显

著,说明考察的自变量及初步识别的模型是合理的。同时,初步建立的模型标准差为0.538,赤池信息准则(AIC)值为416.3。

表1 传递函数模型参数估计值及校验

模型待估计系数参数	估计值	标准差	T值	P值	滞后期	变量	移位
$\omega_{1,0}$	-1.29	0.49	-2.61	0.0095	0	zhhs	12
$\omega_{2,0}$	-0.28	0.04	-6.49	<0.0001	0	xjhs	0
$\omega_{3,0}$	-1.00	0.25	-4.08	<0.0001	0	tcjs	1
$\delta_{1,1}$	0.48	0.20	2.45	0.0149	1	tcjs	1

2.5 传递函数模型的校验

传递函数模型估计后结果显示,模型的残差项不是白噪声且存在规律,需进一步对模型残差进行分析和识别。残差的自相关和偏自相关研究结果表明,12期内的残差序列是一阶自回归和一阶移动平均模型。滞后12期的自相关系数较大,而滞后24期的自相关系数很小。因此,残差可能具有一阶的季节移动。将残差的时间序列模型考虑进模型(表2),模型系数检验值对应 P 值都小于0.05,即在0.05的显著水平下系数都显著。同时残差的白噪声检验结果表明,模型的残差已经是白噪声序列,模型满足各项检验要求。

表2 考虑残差ARIMA模型的传递函数模型参数估计值及校验

模型参数	估计值	标准差	T值	P值	滞后期	变量	移位
MA1,1	0.78	0.05	14.86	<0.0001	1	新井单井平均月产油	0
MA2,1	0.70	0.05	14.26	<0.0001	12	新井单井平均月产油	0
AR1,1	0.97	0.02	42.88	<0.0001	1	新井单井平均月产油	0
$\omega_{1,0}$	-1.35	0.46	-2.92	0.0038	0	综合含水率	12
$\omega_{2,0}$	-0.25	0.05	-4.97	<0.0001	0	新井含水率	0
$\omega_{3,0}$	-0.73	0.18	-4.05	<0.0001	0	投产井数	1
$\delta_{1,1}$	-0.49	0.21	-2.29	0.0228	1	投产井数	1

在考虑残差后,标准差估计值由原来的0.538降低到0.455;AIC值由416.3降到318.9,降低了23%,模型的拟合效果显著提高。

传递函数模型要求模型的残差序列与自变量序列不存在相关性。从表3可以看出,残差序列与输入序列完全无关的概率几乎为1,该模型是可行的。对于未通过校验的模型,需要调整模型参数,直到校验结果满足要求。

表3 残差与输入变量的互相关系数

滞后期	P值		
	新井含水率	投产井数	综合含水率
5	0.798	0.785	0.469
11	0.042	0.761	0.238
17	0.134	0.940	0.165
23	0.278	0.941	0.322
29	0.319	0.982	0.340
35	0.383	0.991	0.369
41	0.497	0.998	0.515
47	0.549	0.994	0.598

经过参数估计和模型校验,得到最终的传递函数模型为

$$(1-B^{12})y_t = -1.35(1-B^{12})zshs_{t-12} - 0.25(1-B^{12})xjhs - \frac{0.73}{(1+0.49B)}(1-B^{12}) \times tcjs_{t-1} + \frac{(1-0.78B)(1-0.7B^{12})}{1-0.97B}a_t \quad (8)$$

从式(8)中可以看出,新井平均单井月产油的季节差分序列与经季节差分的前12期综合含水率、同期新井含水率以及上1期投产井数有关。由于差分的实质是自回归,因此新井平均单井月产油的季节差分序列还与其自身历史有关。

2.6 传递函数模型的预测

传递函数模型中因变量与自变量的互相关分析,对传递函数模型进行预测时,若自变量领先因变量较多,则不用对自变量进行预测,只需用到自变量的实际观测值即可。当预测需要首先提供自变量未来值时,则需要对自变量进行预测。此时,可通过自变量的ARIMA模型,预测自变量数值,将其代入预测模型计算因变量的预测值,可以通过SAS软件自动实现。在实际应用中并不希望对自变量进行预测,而是人为控制并检验预测结果的敏感程度,对其相应结果进行风险分析。胜利油区某油田检验投产井数波动对新井产油量的影响时,利用ARIMA模型预测除控制自变量之外的其他自变量,然后将预测的自变量值以及可以人为控制的自变量值带入传递函数模型中,计算因变量的预测值。

根据标准化的过程,将预测的新井单井月平均产油量逆向还原,并根据历史上的实际投产井数,换算得到新井年产油量,与实际值对比检验方法的精度(图1)。从模型的拟合结果可以看出,实际值与拟合值的吻合程度较高。计算拟合的平均相对误差仅为2.5%。应用该方法,已知2012年投产新井数时,所计算的2012年预测误差仅为0.3%,说明应用传递函数模型进行产油量预测的精度较高。

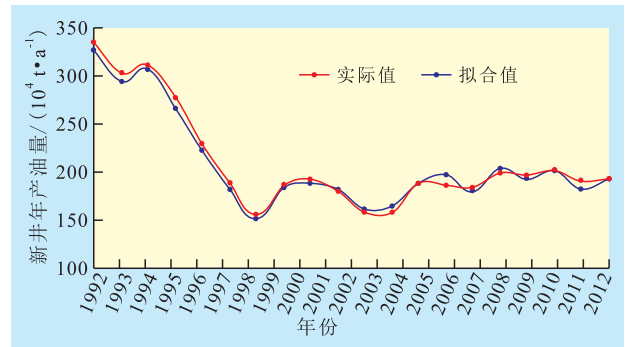


图1 新井年产油量实际值与拟合值对比

3 结束语

基于传递函数模型的油田产油量预测方法,是应用数学方法描述了油田开发指标的规律性和不确定性。以胜利油区某油田的新井产量为研究对象,针对油田实际情况,不仅研究了开发指标与其主要影响因素的关系,还研究了多种影响因素随时间变化对开发效果的影响程度,并且明确了主控因素对效果影响的后效性(非同步性),解决了开发中带有随机性变量的预测问题。

参考文献:

- [1] 刘朝霞,张禹坤,蒋平.Weibull与HCZ预测模型在聚合物驱产油量预测中的应用[J].油气地质与采收率,2007,14(6):76-79.
- [2] 张家良,余元洲,邢卫东,等.Logistic模型的改进及在油藏可采储量预测中的应用[J].油气地质与采收率,2006,13(3):56-58.
- [3] 姚志刚.改进的Logistic模型在原油产量预测中的应用[J].油气地质与采收率,2010,17(3):93-98.
- [4] 殷世荣,宫健,刘凤喜.基础油的预测方法及应用[J].油气地质与采收率,2001,8(3):40-42.
- [5] 赵连素,金春洪,徐茂斌.水驱特征曲线与产量衰减曲线在滨南油田利7块的综合应用[J].油气地质与采收率,2003,10(1):45-47.
- [6] 刘秀婷,杨军,程仲平,等.油田产量预测的新方法及其应用[J].石油勘探与开发,2002,29(4):74-76.
- [7] 吴辛根,葛家理.应用人工神经网络预测油田产量[J].石油勘探与开发,1994,21(3):75-79.
- [8] 李卓,刘斌,刘铁男.支持向量机在油田产量预测中的应用[J].大庆石油学院学报,2005,29(5):96-100.
- [9] 丁岚,杨军.基于传递函数的产品价格预测模型[J].统计与信息论坛,2010,25(6):28-31.
- [10] 肖枝洪,郭明月.时间序列分析与SAS应用[M].武汉:武汉大学出版社,2009.
- [11] 王耀东,张德远,张海雄.经济时间序列分析[M].上海:上海财经大学出版社,1996.