

doi:10.16104/j.issn.1673-1891.2021.03.006

基于多尺度主成分分析的 ARIMA 原油价格预测方法

袁 力

(芜湖职业技术学院基础部,安徽 芜湖 241003)

摘 要:石油作为“工业芯片”,原油价格波动会对全球的经济与政治安全造成影响,准确地预测原油价格未来信息一直备受各方关注。提出基于多尺度主成分分析(MSPCA)的 ARIMA 原油价格预测方法,考虑原油期货价格与现货价格之间的相关性,采用原油期货价格和现货价格序列组成的二维数据作为原始数据,数据经过 MSPCA 后利用 ARIMA 进行预测。该方法利用了小波变换的多尺度分析能力,PCA 的降维统计能力和 ARIMA 模型对非平稳时间序列的预测能力,实验证实该预测方法优于经典 ARIMA 方法和 Holt's 指数平滑法,有效地提高原油价格预测精度。

关键词:原油价格预测;多尺度主成分分析;ARIMA 模型;期货与现货价格;互信息

中图分类号:F416.22;F764.1 **文献标志码:**A **文章编号:**1673-1891(2021)03-0028-05

A Multi-scale Principal Component Analysis-based ARIMA Method for Forecasting Crude Oil Prices

YUAN Li

(Department of Basic Education, Wuhu Institute of Technology, Wuhu, Anhui 241003, China)

Abstract: The fluctuation of crude oil prices has an impact on the global economic and political security. Accurate prediction of the future crude oil prices has been all parties' concern. This paper proposes a multi-scale principal component analysis (MSPCA)-based ARIMA method for crude oil price forecasting, which uses two-dimensional crude oil price series as data source, uses the ARIMA for forecasting after MSPCA and fully considers the correlation between crude oil futures price and spot price. This method uses the multi-scale analysis ability of wavelet transform, the dimension reduction statistical ability of PCA, and the prediction ability of ARIMA model for non-stationary time series. The experiment proves that the ARIMA method based on MSPCA is better than ARIMA method and Holt's exponential smoothing method, and thus can effectively improve the accuracy of forecasting crude oil prices.

Keywords: crude oil price forecasting; multi-scale principal component analysis; ARIMA model; futures price and spot price; mutual information

0 引言

石油作为“工业芯片”,对政治和经济影响深远,其价格波动会影响政治经济稳定。原油价格的变化与走势,往往受到除供需关系之外的各方面因素的综合影响,有较大的不确定性。因此,准确地预测原油价格未来信息一直备受各方关注,也是国内外学者的研究重点之一。原油价格的传统预测方法主要基于经济学与博弈论角度的基本油价理论。1977 年 Levhari^[1]提出 Hotelling 关于枯竭资源

经济学理论的扩展,计量经济学中 ARIMA 模型、指数平滑等方法被相继提出并应用于油价预测。2010 年吴虹^[2]提出了基于 ARIMA 和支持向量机的预测模型,该方法对预测复杂时间序列有良好效果。2016 年熊熊^[3]在 ARIMA 模型中加入季节因子建立国际原油价格预测模型,实验预测结果精确度较高,模型较为精确。人工智能领域的发展为原油价格预测带来了新思路新想法,2017 年范丽伟^[4]提出了一种基于主成分分析的小波神经网络模型(PCA-WNN),用于 WTI 原油现货价格预测,实验验证所

收稿日期:2021-04-09

基金项目:安徽省自然科学基金资助项目(KJ2020A0916);安徽省质量工程资助项目(2019mooc396);芜湖职业技术学院教学研究资助项目(WZ[2020]jy20)。

作者简介:袁力(1992—),女,安徽舒城人,讲师,硕士,研究方向:数据分析等。

提 PCA-WNN 方法对原油价格预测的有效性。

对时间序列采用多尺度分解方法,其多分辨率分析和增强平稳能力能更好地对时间序列进行预测,2005 年梁强等^[5]利用小波分析对布伦特(Brent)原油价格进行长期预测,对原油价格时间序列进行分解和重构,实验证实小波分析在油价长期预测中的有效性。2015 年王书平等^[6]利用经验模态分解将原油价格序列分解并重构成高频、中频、低频和趋势 4 个部分,高中低频与趋势分别用 BP 神经网络和 ARIMA 模型进行预测后再集成,实验显示该组合模型预测优于 ARIMA 模型等单一模型。2020 年刘金培等^[7]提出一种基于区间时间序列小波多尺度分解,结合 Holt's 指数平滑方法、ARIMA 模型和支持向量机回归的组合预测方法,该模型有效提高预测精度。

多尺度主成分分析(MSPCA)^[8]结合了小波变换的多分辨率分析能力和 PCA 的降维能力,充分利用了原油价格序列中蕴含的空间相关性和时间相关性,2017 年 Rua 等^[9]将它用于预测 GDP 数据。而 ARIMA 模型中的差分步骤可以平稳非平稳数据,更贴合数据特点,使模型获得更好的预测效果。因此,本文提出基于 MSPCA 的 ARIMA 原油价格预测方法,考虑原油期货价格与现货价格之间的相关性,采用原油期货价格和现货价格序列组成的二维数据作为数据源。实验证实基于 MSPCA 的 ARIMA 方法的预测优于单一的 ARIMA 和 Holt's 指数平滑预测方法,有效地提高原油价格预测精度。

1 多尺度主成分分析

1.1 小波分析

小波分析作为一种时频信号分析的工具,是傅立叶分析的继承和发展,自其出现以来,在信号和数据处理等领域广泛应用。对原始数据 $x(t)$,小波变换 $WT_x(a, \tau)$ 有如下形式:

$$WT_x(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \varphi^* \left(\frac{t - \tau}{a} \right) dt \quad (1)$$

式中: a 为频率缩放因子; $\varphi(t)$ 为母小波函数,对母小波的伸缩和平移得到小波基函数,小波基函数的适当选择可最优表示原始信号的特征^[10]; $\varphi^*(t)$ 是对 $\varphi(t)$ 取共轭; τ 为移因子。

当容许条件 $c_\varphi < \infty$ 满足时,对小波变换 $WT_x(a, \tau)$ 进行逆变换得到 $x(t)$,

$$x(t) = \frac{1}{c_\varphi} \int_0^\infty \frac{da}{a^2} \int_{-\infty}^{\infty} WT_x(a, \tau) \varphi_{a, \tau}(t) d\tau \quad (2)$$

1.2 主成分分析(PCA)

PCA 在 1901 年由 Pearson 引入,作为一种降维方法,其主要目的是希望用较少的变量去解释原始数据中的大部分变异,将诸多相关性很高的变量转换成彼此相互独立或不相关的新变量。通常选出比原始变量个数少且能解释大部分数据的变换的几个新变量,即主成分,并用以解释原始数据。

首先对原始数据 $X_{n \times p}$ 进行标准化处理得到 $\bar{X}_{n \times p}$,计算处理后数据的相关系数矩阵 $R_{p \times p}$ 及该矩阵的特征值 $\lambda_1, \lambda_2, \dots, \lambda_p$ 和特征向量 $a_i = (a_{i1}, a_{i2}, \dots, a_{ip})$ ($i = 1, 2, \dots, p$),即得 p 个主成分。实际中可根据累积贡献率或特征值大小等准则选择保留的主成分个数。

1.3 多尺度主成分分析(MSPCA)

MSPCA 将小波变换抽取数据确定性特征的能力与 PCA 抽取多元变量共有模式的能力相结合,首先对二维数据的每一列进行一维小波分解,生成高频和低频小波系数,在每一层,对小波系数进行 PCA,选择合适数量的主成分,在重构系数后进行小波逆变换数据,最后再进行一次 PCA 及重构,流程如图 1 所示。

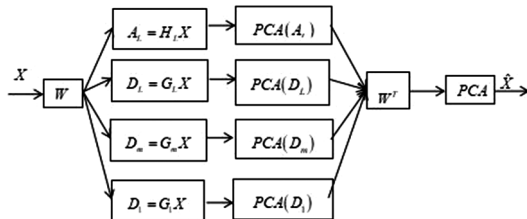


图 1 多尺度主成分分析流程

在 MSPCA 中,每个时间点仅使用到该时间点的信息,而不是对原始多维信号整个样本进行小波分解,以便处理后的数据能更好地反映出原始样本的特征。

1.4 WTI 原油现货价格与期货价格相关性分析

期货市场是在现货市场的基础上发展,进一步发展必然不能脱离现货市场,期货与现货价格之间有着动态关系^[11]。为了衡量期货与现货价格之间的相关性,本文采用互信息^[12](Liu 等,2001)作为度量标准。2 个变量 s_1 和 s_2 之间的互信息 $I(s_1; s_2)$ 可表示为:

$$I(s_1; s_2) = \iint p(s_1, s_2) \log \frac{p(s_1, s_2)}{p(s_1) p(s_2)} ds_1 ds_2 \quad (3)$$

式中: $p(s_1)$ 和 $p(s_2)$ 分别为 s_1 和 s_2 的概率密度函数; $p(s_1, s_2)$ 为 s_1 和 s_2 的联合概率密度函数,并利用 Peng 等^[13]所述方法计算互信息值 $I(s_1; s_2)$,互

信息值越大,表示变量之间的相关性越强。

为了说明 MSPCA 对原始数据的影响,笔者分别计算原始的期货与现货价格序列之间的互信息,和经过 MSPCA 后的期货与现货价格序列之间的互信息。选取 3 组 WTI 原油价格,分别为同一时间段的期货与现货价格序列数据,计算互信息,得到结果如表 1 所示。

表 1 原始和 MSPCA 处理后的期货与现货价格序列之间的互信息值

	第 1 组数据	第 2 组数据	第 3 组数据
原始数据	5.539 4	5.448 0	5.488 7
MSPCA 处理后数据	6.248 5	6.035 1	5.787 3

观察表 1 数据可以发现,经过 MSPCA 处理后的数据中的期货与现货价格序列之间的互信息值均有提高,说明期货与现货价格序列之间的相关性变强。MSPCA 利用小波变换抽取了数据的确定性特征,利用 PCA 抽取多元变量共有模式,处理后的期货与现货价格序列之间变强的相关性,进一步说明了经过 MSPCA 处理后的数据能更好地提取原始样本的共有确定性特征。

2 单项预测方法

2.1 自回归积分移动平均过程

自回归积分移动平均过程(ARIMA)是时间序列分析中的经典模型,主要用于非平稳时间序列的分析^[3]。设 $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ 是非平稳时间序列,若存在正整数 d ,使得 $\nabla^d X_t = W_t$,而 $\{W_t, t = 0, \pm 1, \pm 2, \dots\}$ 是自回归移动平均过程(ARMA(p, q))序列,则称 X_t 是 ARIMA(p, d, q) 序列。这时, X_t 满足

$$\varphi(B) \nabla^d X_t = \theta(B) \varepsilon_t \quad (4)$$

若 $\nabla^d X_t$ 为平稳序列,但均值 $\mu \neq 0$,则 $\nabla^d X_t - \mu$ 为平稳的零均值序列,满足

$$\varphi(B) (\nabla^d X_t - \mu) = \theta(B) \varepsilon_t, t > d \quad (5)$$

此时,称 X_t 为一般 ARIMA(p, d, q) 序列,若 μ 未知,可用 $\nabla^d X_t$ 的平均值估计。其中 ε_t 为白噪声时间序列, d 为差分阶数, p 为自回归系数, q 为移动平均系数。ARIMA(p, d, q) 模型对原始数据进行的 d 阶差分操作,平稳化数据,使模型更加精准。

建立 ARIMA 模型后,在原来时间序列观测值 $x_t, x_{t-1}, x_{t-2}, \dots$ 的基础上,对未来序列进行向前 $l(l > 0)$ 步预测,得到 x_{t+l} ,显然预测步长越大,预测精度越差。但由于会不断获得新的观测值 x_{t+1}, x_{t+2}, \dots ,将新的观测值 x_{t+1}, x_{t+2}, \dots 加入历史数据

中,就能提高预测精度。

2.2 Holt's 指数平滑法

Holt's 指数平滑法^[6]是对一次指数平滑值再进行指数平滑,模型如下:

$$\begin{cases} \hat{y}_{t+l} = \hat{y}_t + l\hat{x}_t \\ \hat{y}_t = \alpha y_t + (1 - \alpha) (\hat{y}_{t-1} + \hat{x}_{t-1}) \\ \hat{x}_t = \beta (\hat{y}_t - \hat{y}_{t-1}) + (1 - \beta) \hat{x}_{t-1} \end{cases} \quad (6)$$

式中: \hat{y}_t 和 \hat{y}_{t-1} 分别为 t 和 $t - 1$ 时刻预测值; \hat{x}_t 和 \hat{x}_{t-1} 分别为 t 和 $t - 1$ 时刻增量预测值; α 和 β 为平滑系数; y_t 为 t 时刻真实值; \hat{y}_{t+l} 为 $t + l$ 时刻预测值; l 为预测步长。

Holt's 指数平滑法利用历史数据的特点,能较好地预测数据的短期趋势,而 ARIMA 模型克服了指数平滑法难以确定指数平滑系数,主观性影响较大等缺点,能对数据进行较长时间的预测。

3 基于 MSPCA 的原油价格预测

原油价格序列有复杂性、非线性、有噪声等特征,且原油价格序列统计特性变化大,有着显著的非平稳性特征,这些特征会对序列预测效果产生影响^[11]。鉴于一个综合预测模型通常由 2 个或 2 个以上的模型组合而成,利用不同方法各自的优势,提高预测性能,将小波变换的多尺度分析能力,PCA 的降维统计能力和 ARIMA 模型对非平稳时间序列的预测能力相结合,本文提出基于 MSPCA 的 ARIMA 的原油价格预测方法。

本文方法包括以下基本步骤:

1) 原油价格序列的小波分解。首先采用标准正交小波变换 W 对原油价格序列 X 进行小波多尺度分解,获得各个尺度的低频及高频小波系数矩阵 $A_L, D_i (i = 1, \dots, L)$ 。

2) 小波系数矩阵的主成分分析和重构。对各个尺度小波系数矩阵 $A_L, D_i (i = 1, \dots, L)$ 进行主成分分析,然后根据 Kaiser 准则选择主成分的数目,最后重构出小波系数矩阵 $\bar{A}_L, \bar{D}_i (i = 1, \dots, L)$ 。

3) 原油价格序列的小波逆变换。对各个尺度小波系数矩阵 $\bar{A}_L, \bar{D}_i (i = 1, \dots, L)$ 作小波逆变换,重构出原油价格序列。

4) 原油价格序列的主成分分析及重构。对小波重构出的原油价格序列再次进行主成分分析和重构,具体步骤与步骤 2 类似,获得最终的原油价格序列 \hat{X} 。

5) 原油价格序列的预测。对上一步中最终重

构出的原油价格序列 X 每一列进行 ARIMA 预测,最后进行模型评价。

4 实证分析

4.1 数据及其预处理

原油价格选取 2 种常用的原油价格序列,即 WTI 原油期货和 WTI 现货价格序列。第 1 组数据是从 2004 年 1 月 18 日—2 月 9 日(共 526 周)的每周数据,由 2×526 个观测值组成(图 2)。第 2 组数据是从 2009 年 1 月 4 日—2018 年 12 月 30 日(共 522 周)的每周数据,由 2×522 个观测值组成(图 3)。第 3 组数据是从 2004 年 1 月 1 日—2019 年 1 月 1 日(共 181 个月)的每月数据,由 2×181 个观测值组成(图 4)。

4.2 模型评价

为了评价本文提出方法的预测效果,度量采用均方根误差(RMSE)和平均绝对百分误差(MAPE)。RMSE 和 MAPE 的计算公式如下:

$$RMSE = \frac{1}{p} \sqrt{\sum_{k=1}^p (x_k - \hat{x}_k)^2} \quad (7)$$

$$MAPE = \frac{100}{p} \sum_{k=1}^p \left| \frac{x_k - \hat{x}_k}{x_k} \right| \quad (8)$$

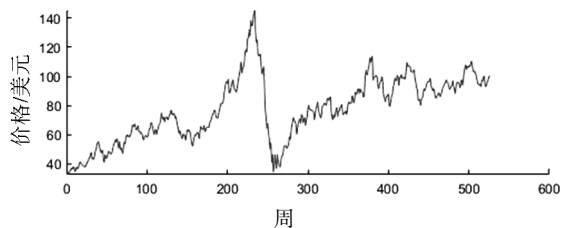
式中: x_k , \hat{x}_k 分别为原始价格和预测价格; p 为被预测的原始价格序列长度。RMSE 与 MAPE 数值越小,就表明预测所得数据与原始数据具有更小的差异。

将本文所提出的基于多尺度主成分分析的 ARIMA 的原油价格预测方法(MSPCA_ARIMA)与经典 ARIMA 预测方法^[3]和 Holt's 指数平滑法^[6]作对比,对训练数据做对比实验,计算预测结果与原来价格序列的 RMSE 和 MAPE,发现本文方法预测精度更高,有一定的预测效果。具体结果如表 2 所示。

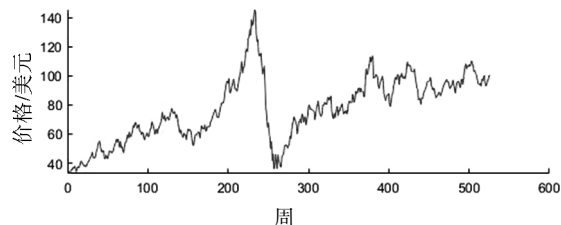
其中第 1 组、2 组和 3 组数据分别选取前 350、300 和 100 个数据点作为训练集,后面的数据作为预测集合。ARIMA 模型和 Holt's 指数平滑法中的预测步长均取向前 1 步预测。

从表 2 可以看出,本文方法取得的 RMSE 和 MAPE 值小于 ARIMA 方法和 Holt's 指数平滑法对应的值,因此本文方法的降噪效果更好,预测精度有提升,且方法对较高频数据处理效果更为明显。为了更好地体现实验效果,在实验中取一组数据其中的某一道序列进行观察。选择表 1 中第 2 组

数据的期货价格序列的预测效果局部放大对比效果如图 5 所示。

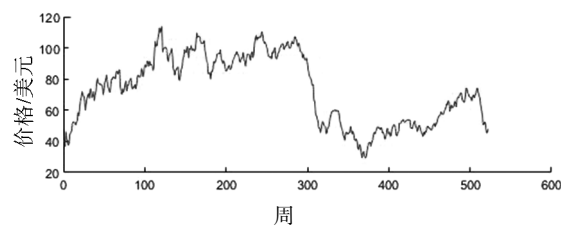


a. 期货价格序列

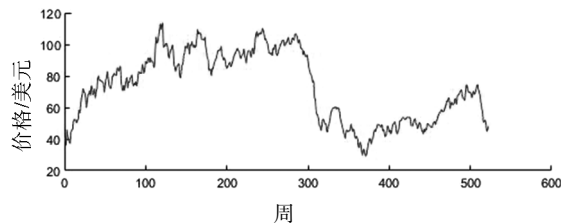


b. 现货价格序列

图 2 第 1 组原油价格序列

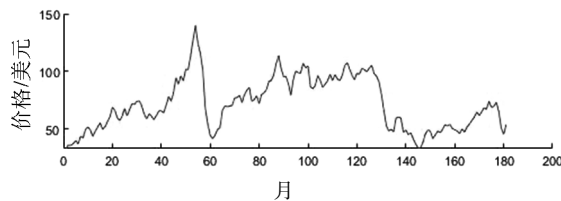


a. 期货价格序列

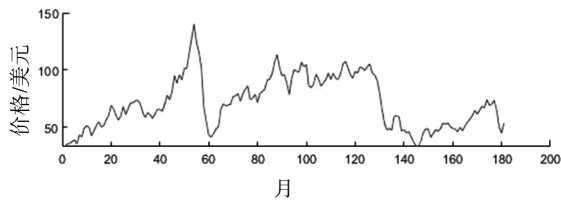


b. 现货价格序列

图 3 第 2 组原油价格序列



a. 期货价格序列



b. 现货价格序列

图 4 第 3 组原油价格序列

表 2 本文方法、ARIMA 方法和 Holt's 指数平滑法预测效果

		第 1 组数据		第 2 组数据		第 3 组数据	
		期货	现货	期货	现货	期货	现货
RMSE	Holt's 指数平滑法	0.081 4	0.081 7	0.070 6	0.071 2	0.232 0	0.233 3
	ARIMA	0.062 8	0.063 8	0.031 5	0.031 4	0.222 4	0.225 8
	MSPCA_ARIMA	0.055 7	0.054 5	0.019 9	0.021 4	0.221 8	0.217 7
MAPE	Holt's 指数平滑法	0.92	0.93	1.58	1.58	2.65	2.66
	ARIMA	0.74	0.75	0.75	0.69	2.60	2.75
	MSPCA_ARIMA	0.62	0.62	0.44	0.45	2.63	2.62

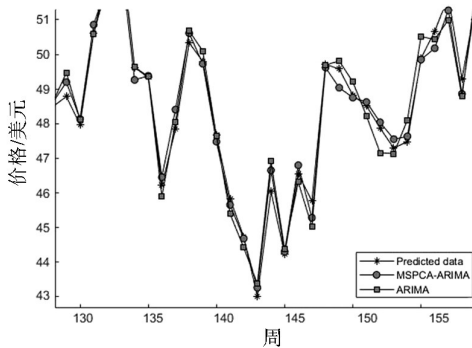


图 5 第 2 组数据期货价格序列的预测效果局部放大对比

从图 5 可以看出,在大多数时间点处,本文方法较 ARIMA 方法更接近真实数据,准确度更高,趋势和实际走势保持一致,能够较好地预测价格序列。综合表 1 和图 5 可以看出,本文方法在误差上较经典的预测方法有一定的减小,预测后的价格序列与

原价格序列具有更小的差异。

5 结语

本文提出基于 MSPCA 的 ARIMA 原油价格预测方法,将小波变换的多尺度分析能力,PCA 的降维统计能力和 ARIMA 模型对非平稳时间序列的预测能力相结合,以原油期货价格和现货价格序列组成的二维数据作为原始数据,充分利用了数据中蕴含的时空相关性,实验证实基于 MSPCA 的 ARIMA 方法的预测优于 ARIMA 方法和 Holt's 指数平滑法,有效地提高原油价格预测精度。这也在一定程度上佐证了组合模型较单一模型合理性和准确性更高。同时,因原油价格受诸多复杂因素影响,如政经形势、突发事件等,在后续研究中,可在本文模型中可加入这些影响因素,提高预测精度。

参考文献:

- [1] LEVHARI D, LIVIATAN N. Notes on Hotelling's economics of exhaustible resources[J]. Canadian Journal of Economics, 1977, 10(2): 177-192.
- [2] 吴虹,尹华.ARIMA 与 SVM 组合模型的石油价格预测[J]. 计算机仿真, 2010, 27(5): 264-266+326.
- [3] 熊熊,李璇.国际原油价格预测研究[J]. 重庆理工大学学报(自然科学版), 2016, 30(1): 48-53.
- [4] 范丽伟,代杰,尹俊超.基于改进型小波神经网络的油价预测[J]. 统计与决策, 2017(12): 85-88.
- [5] 梁强,范英,魏一鸣.基于小波分析的石油价格长期趋势预测方法及其实证研究[J]. 中国管理科学, 2005(1): 31-37.
- [6] 王书平,朱艳云.基于多尺度分析的国际原油价格预测方法研究[J]. 价格月刊, 2015(10): 1-5.
- [7] 刘金培,汪漂,黄燕燕,等.基于区间时间序列小波多尺度分解的组合预测方法[J]. 统计与决策, 2020, 36(19): 5-9.
- [8] BHAVIK R B. Multiscale PCA with application to multivariate statistical process monitoring[J]. AIChE Journal, 1998, 44(7): 1596-1610.
- [9] RUA A. A wavelet-based multivariate multiscale approach for forecasting[J]. International Journal of Forecasting, 2017, 33(3): 581-590.
- [10] 汪金菊,袁力,刘婉如,等.地震信号随机噪声压制的双树复小波域双变量方法[J]. 地球物理学报, 2016, 59(8): 3046-3055.
- [11] 张冠华,丁日佳.布伦特原油期货价格与现货价格的关系研究[J]. 金融理论与实践, 2011(8): 30-33.
- [12] LIU J, MOULIN P. Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients[J]. IEEE Transactions on Image Processing, 2001, 10(11): 1647-1658.
- [13] PENG H, LONG F, DING C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1226-1238.