

美元人民币汇率预测模型



使用不同预测模型，依据相关联数据，对未来美元对人民币汇率走势进行合理预测。



模型设计流程

确定核心的预测模型和其
决定性参数，进行宏观汇
率预测（从IMF获取）

01

02

通过不同的模型和
算法确定附加参数

结合短期和长期的经济
指标和市场数据对宏观
汇率进行预测

03

04

测试最终预测模型

预测模型简介

01

机器学习模型：

- XGB
- 随机森林 (Random Forest)
- 决策树 (Decision Tree)

统计模型：

- 向量自回归 (Vector Autoregressive model)

02

- 考虑不同数据和参数间的**相关性**
- 考虑不同数据间所产生的**信息增益** (information gain) 和**数据的随机性**
- 可同时对海量参数进行模拟，探索**不同因子对预测的影响**
- 容易过度挖掘数据，产生过度拟合以至于降低预测精准性

- 可有效模拟**多因子时间序列模型**中不同时序间关联性，即不同时期不同因子的时间序列关系
- 模拟不同因子在不同时期对预测目标产生的影响，较少考虑因子间相关性 (linear independence)
- 模型关系多为线性关系，现实中数据多为非线性关系

模型数据及衡量指标

宏观
数据

微观
数据

1

宏观数据： GDP，人均GDP，外汇总资产，进出口额度，外汇储备，就业率，各国赤字等相关数据

衡量
标准

2

衡量标准： 经济学意义，模型准确性，数据相关性

3

微观（市场 – market based）数据： 一年期中债国债到期收益率（YTM），1年期美国债到期收益率（YTM），SP500，NASDAQ index，外汇远期，上证指数，VIX等相关数据

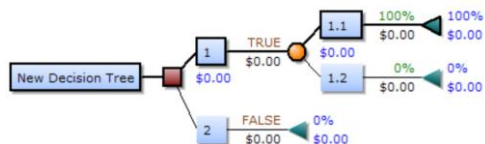
模型介绍

■ XGB (极端梯度提升)

- 根据每次预测的正确与否进行**自我修正并学习所犯错误**，从而正确模型因子间关系（非线性关系，易模拟黑天鹅事件）
- 可同时模拟大量数据，并自动删除无用数据以避免过度模拟
- 准确度高，但参数过多不宜建模

■ 决策树 (Decision Tree)

- 依据数据和因子间所产生的**信息增益 (information gain)** 和**数据熵 (entropy)** 进行汇率走势预测
- 过于**依赖数据**间关系导致**过度拟合**



■ 随机森林 (Random Forest)

- 随机使用决策树（结果优于决策树）模拟不同数据集，并进行加权合并。**可模式市场中汇率的随机走势**
- 需大量数据模型市场走势，但受因子间关系大

■ 向量自回归 (VAR) [3][4]

- 模拟**多因子时间序列**模型中不同因子间时序的**线性关系**
- 过多因子易导致线性模型长生过多误差

数据集



01

- 1yr中债国债到期收益率(中债)(日)
- NASDAQ(日)
- QDII投资额度
- sp500主要指数(日)
- USDCNY人民币汇率(日)
- USD伦敦同业拆借利率(LIBOR)(日)
- 上海银行间同业拆放利率(SHIBOR)(日)
- 上证国债指数(日)
- 中债国开债到期收益率(中债)(日)
- 外汇远期 (1D) 成交统计
- 市场波动率指数(VIX)



02

- US企业债收益率(月)
- US储备资产(月)
- US基准利率
- US外汇储备(月)
- US市场波动率指数(VIX)
- US消费信贷(月)
- US消费者物价指数(CPI)同比(月)
- US联邦政府财政赤字(月)
- US货币供应量(月)
- US进口价格指数(月)
- US采购经理指数(PMI)(月)
- 非农就业人数(月)



03

- Per GDP
- real_fx
- real_monthly_fx
- term_of_trade
- USDCNY平均汇率(月)
- 国内生产总值(季)
- 国际收支总差额累计值(亿美元)
- 经常账户差额当季值(亿美元)



04

- GDP人均GDP
- M0供应量(月)
- 中国CPI当月同比(月)
- 中国出口贸易(月)
- 中国制造业采购经理指数(PMI)(月)
- 中国投资海外证券情况(月)
- 中国进口贸易(月)
- 人口概况(年)
- 人民币存款基准利率
- 出口价格指数HS2分类(月)
- 国家财政支出(月)
- 宏观经济景气指数(月)
- 官方外汇储备(月)
- 工业增加值 宏观预测(月)
- 消费者信心指数
- 规模以上工业增加值当月(月)
- 进口数量指数HS2分类(月)
- 金融机构存款准备金率

向量自回归(VAR)

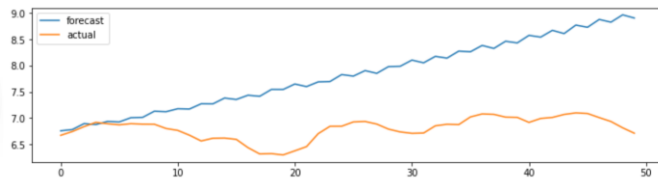


- 度量标准:对称平均绝对误差百分比 (symmetric mean absolute percentage error)

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

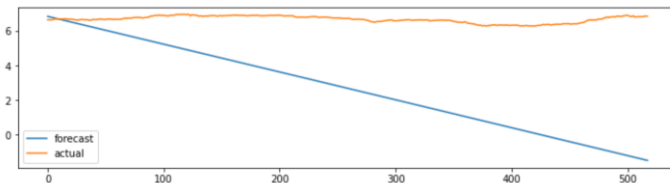


■ 长期 (long run) 预测



误差 : 13.4%

■ 短期 (short run) 预测



误差 : 98.0%

■ 结论:

- 较多的因子导致线性模型误差较大
- 无法较好模型该数据集多因子间时间序列关系

决策树 (Decision Tree)

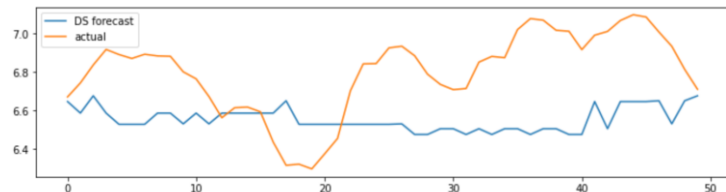


- 度量标准: 对称平均绝对误差百分比 (symmetric mean absolute percentage error)

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$



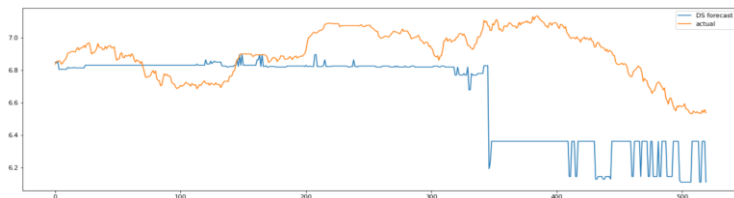
■ 长期 (long run) 预测



误差 : 3.86%

核心预测参数: 中国外汇储备, M0

■ 短期 (short run) 预测



误差 : 4.25%

核心预测参数: SP500, 国开行1年债到期回报率, USD Libor

■ 结论:

- 短期汇率预测精确尚可
- 能模拟汇率趋势, 但波动较大

随机森林 (Random Forest)

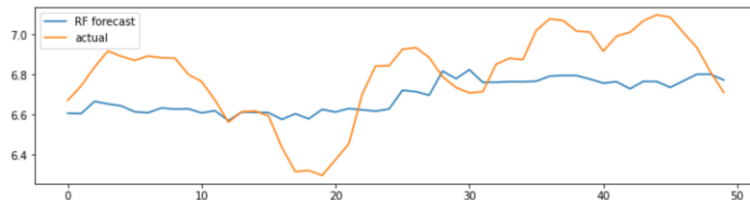


- 度量标准: 对称平均绝对误差百分比 (symmetric mean absolute percentage error)

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$



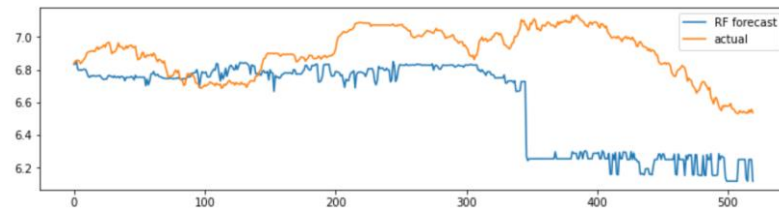
长期 (long run) 预测



误差 : 2.95%

核心预测参数: USD Libor

短期 (short run) 预测



误差 : 4.94%

核心预测参数: USD libor, SP500, NASDAQ, 美债收益率

结论:

- 长期汇率预测精确较高
- 可较好模拟汇率走势, 但突发事件对模型影响大

极端梯度提升 (XGB)

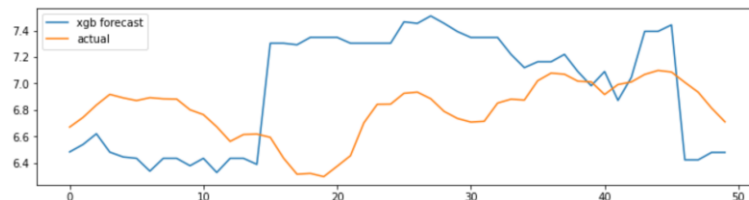


- 度量标准: 对称平均绝对误差百分比 (symmetric mean absolute percentage error)

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$



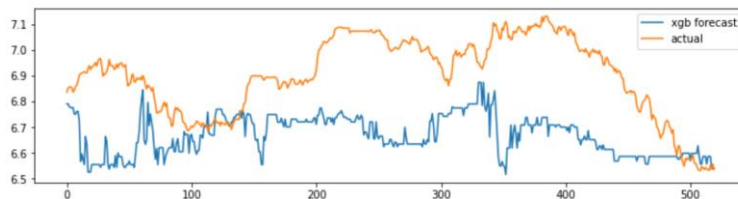
■ 长期 (long run) 预测



误差: 6.33%

核心预测参数: 中国央行基准利率, 美国货币供应量

■ 短期 (short run) 预测



误差: 4.89%

核心预测参数: USD libor, SP500, 1年中国国债到期收益率, SP 500 VIX

■ 结论:

- 短期汇率预测精确最高, 但长期有误差
- 可较好模拟汇率走势和突发事件, 短期误差在可控范围

最终模型



1



XGB进行长期 (long run) 汇率预测

可较好模拟汇率增长趋势，且误差在可接受范围和添加数据随机性

2



基础参数 (IMF 推荐) :

中国人均GDP, 进出口交换比例, 外国总资产 [1]

3



附加参数：从之前四个模型中提取

中国官方外汇储备, 金融机构存款准备金率, 美国货币供应量

4



预测结果:

预测目标: 真实汇率

预测准确度: SMAPE - 50.1%, RMSE - 12.196

最终模型：远期汇率预测

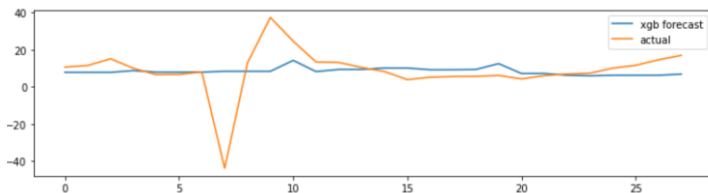


- 度量标准:对称平均绝对误差百分比 (symmetric mean absolute percentage error)

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$



■ 长期 (long run) 预测



误差 : SMAPE - 50.10%, RMSE - 12.20 (IMF RMSE: ~1)

核心预测参数: 中国央行基准利率, 美国货币供应量

■ 结论:

- 实际汇率预测结果误差较大, 需对预测目标进行优化
- 可较好模拟汇率走, 但远期无法精确模拟突发事件
- 数据模型预测远期汇率误差较大, 建议使用经济学模型[1]

改进方法

