数据可视化期末考查作业报告

——美元汇率波动因素和规律初探

何阳 张雁玲 王泷德

1. **问题阐述**

当前，外汇投资是社会中较为常见的一种投资项目，得到人民群众的广泛参与。本可视化工程描述了自2005年至今的美元兑人民币比价走势，同时引入经济统计数据，通过研究汇率波动以及经济统计数据发布时的汇率波动来总结规律，分析原因，以期发现理想的投资时点。

1. **工程设计**
2. **基本理论**

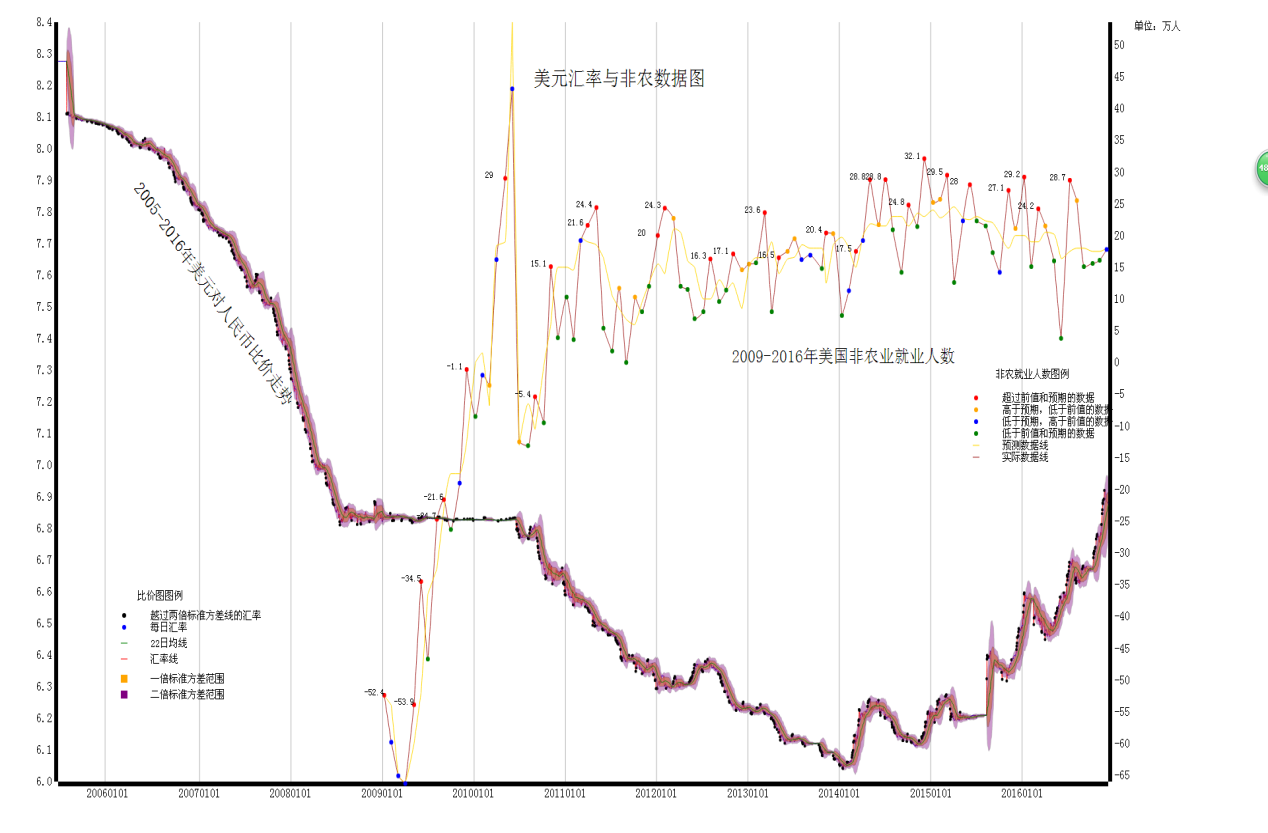
借用切比雪夫不等式的思想，将每日汇率看作随机变量X，其一定时间内的汇率均值视为数学期望E（X），方差为D（X），则  
P（|X-E（X）|>=ε）=<D(X)/ ε2。即如果每日汇率X偏离均值E（X）超过2倍标准方差的概率要小于25%，如果再考虑偏差方向，向上和向下偏离各只有不到12.5%的概率。我们基于以上思想，统计数年来美元对人民币的走势，找出汇率偏离其均值超过2倍标准方差的时点，并分析每一个时点若干天后的汇率波动规律。

本工程又引入了非农就业人数这一经济统计数据，其是美国非农业人口的就业状况的数据指标，可以极大地影响货币市场的美元价值，客观地反映了美国经济的兴衰。本工程区分了该数据的四种情况，即

|  |  |
| --- | --- |
| 高于前值，高于预期 | 低于前值，高于预期 |
| 高于前值，低于预期 | 低于前值，低于预期 |

前值即上个月的数据值；预期即当月数据正式发布前，美国有关部门官方发布的对当月数据的预测值。一般认为数据高于前值、高于预期利好美元，低于前值、低于预期利空美元。该数据发布时间为下月初发布上月的数据。本工程由于篇幅有限，同时工程参与人员经济知识有限，故只对数据四种情况中的两种（高于前值、高于预期；低于前值、低于预期）结合偏离标准方差偏离情况加以了分析。

1. **工程组成**
2. 绘制2005年-2016年美元对人民币的汇率曲线、均线曲线、一倍和二倍标准方差范围。
3. 绘制2009年-2016年美国非农业就业数据曲线，并使用不同颜色标示各月数据可能对经济发生的正面或负面的影响。即前文所述的四种情况中的两种。
4. 汇率数据统计分析——找出偏离均值超过两倍标准方差的时点，并分析每个时点若干天后的涨跌规律，进行统计。
5. 综合分析——分析当非农业就业数据出现支撑美元走高或走弱的讯息时，那些偏离均值超过两倍方差的时点若干天后的走势将受到何样的影响。

****

1. **统计分析项目和方法**
   1. 统计每日汇率偏离汇率均值超过两倍标准方差的时点。
   2. 统计这些汇率偏离汇率均值超过两倍标准方差的时点三天后、五天后、十天后是否继续向增大偏离的方向波动。
   3. 引入非农业就业数据，考察在该数据发布时偏离均值超过两倍标准方差的日子十天后的汇率走势。
2. **工程实现**
3. 数据收集

本工程从网络上收集了2005年-2016年美元兑人民币的每日汇率（以下简称：数据1）和2009年-2016年美国劳工部月度发布的非农业就业数据（以下简称：数据2）作为工程的数据输入。

1. 编码实现

本工程使用d3.js语言进行开发设计，使用Firefox浏览器进行调试和展示，并使用Apache Http Server 来搭建服务器，存储要读取的.csv文件。本工程程序行数：472。

1. 具体实现细节
   * 1. 从两个.csv文件读取数据1和数据2。
     2. 利用数据1的画出2005年-2016年美元兑人民币的汇率曲线等，包括汇率曲线、均线、一倍和两倍标准方差范围。
     3. 利用数据2的画出非农业就业数据曲线以及以各点所属的四种情况之一为各数值点着色。红色代表高于前值、高于预期；绿色代表低于前值、低于预期；橙色代表低于前值、高于预期；蓝色代表高于前值、低于预期。
     4. 统计分析汇率偏离汇率均值超过两倍标准方差的时点三天后、五天后、十天后继续向增大偏离的方向波动的个数和偏离比率。
     5. 统计分析当非农业就业数据出现支撑美元走高或走弱的信号时，那些偏离均值超过两倍方差时点十天后出现继续增大偏离的天数和偏离比率。
     6. 添加交互功能——缩放以及显示当前鼠标指示位置的数据值。
     7. 其他——添加图例、标题和坐标轴。
2. 源代码

另附。

1. **工程结论和讨论**
2. 从2009年-2016年，总共收集了3000天的数据，偏离超过二倍标准方差的天数为552天，所占比例为18.4%。如此高的比例表明，股价不是随机波动的，其中应有系统性的决定因素。

|  |  |  |
| --- | --- | --- |
| 总天数 | 偏离超过二倍方差的天数 | 所占比例 |
| 3000 | 552 | 18.4% |

1. 偏离超过两倍标准方差若干天后汇率变化统计。以十日内继续增大偏离为例，在552个偏离超过二倍方差的日子里，有355天在十天后增大了偏离，这355天增大偏离的平均比率是0.33%，355天占全部偏离超过二倍方差的天数的64.3%。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 天数 | 继续加大偏离比率的平均值 | 天数占所有偏离超过两倍方差的比例 |
| 三日内继续增大偏离 | 292 | 0.18% | 52.9% |
| 五日内继续增大偏离 | 306 | 0.26% | 55.4% |
| 十日内继续增大偏离 | 355 | 0.33% | 64.3% |

1. 偏离均值超过两倍标准方差时点若干天后的波动比率期望。以十日后波动比率为例，即在552个偏离超过二倍方差的日子里，所有这些日子与十天后的汇率相比的平均波动比率。正数为增大偏离，负数为减小偏离。

|  |  |
| --- | --- |
| 三日后波动比率 | 0.035% |
| 五日后波动比率 | 0.062% |
| 十日后波动比率 | 0.126% |

1. 从结论2和3可以得出，当汇率越过两倍标准方差时，其大概率事件是继续增大偏离，而不是减小，并随着考察之后天数的增加，比例逐渐变大。其原因可能是汇率价格此时已经超过了某个重要的心理关口或界限，导致投资者追涨杀跌，这从汇率价格的一些决定因素中可以解释。即一国货币的价值更多地决定于其国家经济的好坏、其国际竞争力、进出口收支和处于国际产业链中的位置。即使是短期有人人为操纵，也需要大量资金。这些因素都较为宏大，一旦因以上因素而导致汇率波动，很难短时间内对冲或扭转该趋势，而必然是在一定时期内保持这一趋势。
2. 在非农业就业数据发布三天之内偏离均值超过两倍标准方差的时点十天后的波动情况。本工程考察了2009年1月至2016年12月之间，共84个月发布的在非农业就业数据。其中有64个月是我们的考察对象，即高于前值、高于预期或低于前值、低于预期。我们查找发布数据后三日内是否有偏离均值超过两倍标准方差的日子，共计查找64\*3=192天，其中21天是偏离均值超过两倍标准方差的日子，占比11%。这21天里有10天是十日后偏离继续增大，这10天平均的偏离比率是0.45%。然而，这21个偏离均值超过两倍标准方差的日子，十日后的平均偏离比率是-0.12%，表明平均来看是缩小了偏差。以上数据表明，在非农业就业数据发布后，市场波动较平常增加，增大偏差和减小偏差的比率都加大了，但波动方向并不与非农业就业数据所代表的含义正相关。可以这样解释，在该数据发布后，市场可能会借机对该数据做一轮短期行情。而由于该行情没有其他“宏大”因素的支持，市场各方都知道这是短期行情，故会高抛低吸，及时对冲，产生涨跌交互的状态。而这192个被考察的日子里，只有11%的日子偏离均值超过两倍标准方差，小于18.4%的平均比例，说明该数据发布不导致汇率的大波动，反而是有助于其稳定。这与之前所述波动增加，事实上是辩证统一的。即要么是不偏离，要么是发生大偏离，但会及时回调。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 符合标准的非农数据发布数 | 考察的天数 | 其中有偏离均值超过两倍标准方差的时点个数 | 十日后偏离增大个数 | 偏离增大比率 | 波动期望比率 |
| 64 | 192 | 21 | 10 | 0.45% | -0.12% |

1. 对可视化技术方面的讨论

本文使用的可视化图形是折线图，x轴是时间，y轴是数值，即汇率和非农就业人数。之所以选择折线图，是因为对于汇率、股价波动，传统上是选择这样的折线图，早已为大多数人接受和认可。该图可以看出数值随时间的变化。

本工程将汇率曲线图和非农就业图放置在一个图内，原意是期望直观地看出两曲线之间的关系，比如相似、相逆或其他。然而，经过直观感受和数据分析，发现这两种数据间并无有规律可偱的某种关系。

1. **工程可精进的地方**
2. 本工程所展示的图形为美元对人民币汇率，但其实展示美元指数可能更好，因为该指数可能更受美国国内经济数据的影响。
3. 在交互方面还可加强，观看者目前不能自由选择均线时间、标准方差倍数等参数。
4. 还可增加更多的影响因素来考察汇率波动，这样可能会发现某些数据对于美元价格的影响。