市场数据缺失补全设计文档

### 路透市场数据缺失补全

#### 数据范围：

由路透，中债登，风险集市和金交所（手工）等4个不同来源提供的市场数据。

#### 功能要求：

在进入市场风险应用数据库后，根据预定义的规则，自动实现缺失数据补全。不同类型的产品有不同的缺失补全规则。

#### 补全规则：

##### 路透（FX、FXOPTION、IRS）市场数据补全规则

1) RIC码存在，部分数据缺失补全，按以下的规则进行。交易日期，收市日期，取业务数据时间，交易时间取系统时间点，如果最新价类型缺失，统一赋值99。

|  |  |  |  |
| --- | --- | --- | --- |
| ASK | BID | PAC | 取值规则 |
| √ | √ | ╳ | （ASK+BID）/2 |
| √ | ╳ | √ | ASK |
| √ | ╳ | ╳ | ASK |
| ╳ | √ | √ | BID |
| ╳ | √ | ╳ | BID |
| ╳ | ╳ | √ | PAC |
| ╳ | ╳ | ╳ | 进入市场数据缺失补全规则 |

2）按上述表中，当三种价格缺失的时候，进行以下的规则的市场数据缺失补全。

当该RIC码的RIC码类型为OUTRIGHT（不允许为0）时，如果价格为0，当作该价格丢失，按规则补全价格，如果RIC码类型为SPREAD(允许为0)，则允许价格为0。

主要分为以下2种类型：

* 取昨天数据，以配置表（IAL.INT\_RIC\_CFG）中的GROUP\_ID分组判定，缺失大于或等于曲线允许最大缺失值时，所有曲线点沿用昨天的数据补全。注：交易日期、交易时间、收市日期取当天跑批的系统时间。
* 左右找平或中间插值，以GROUP\_ID判定，曲线缺失点小于曲线允许最大缺失值时，使用下面的3种规则进行补全。
  + 1. 最左边数据缺失，取已有最左边一点的数据进行等值补全；
    2. 最右边数据缺失，取已有最右边一点的数据进行等值补全；
    3. 中间数据缺失，取最近有数值的两点进行线性插值补全；线性插值中，每个点的期限的天数由配置表提供。线性插值计算公式：y=y1+(y2-y1)\*(x-x1)/(x2-x1)

##### 中债登市场数据补全规则

1. 前置处理：
   1. 将源表的曲线名称（CRVNM），曲线类型（CRVTP），期限描述（TERMDESC）拼接成唯一的标识，并加载进接口表的RIC字段。
   2. 将源表的收益率（YLD），统一加载接口表的UNI\_ASK,UNI\_BID,PRI\_ACT 中。
   3. 最新价类型，统一标识为99
   4. 接口表的交易日期，交易时间，取业务数据日期和系统时间点。
2. RIC码缺失补全，按以下的规则进行。主要分为以下2种类型：
   * 取昨天数据，以配置表（IAL.INT\_RIC\_CFG）中的GROUP\_ID分组判定，缺失大于或等于曲线允许最大缺失值时，所有曲线点沿用昨天的数据补全。注：交易日期，交易时间，收市日期取当天跑批的系统时间。
   * 左右找平或中间插值，以GROUP\_ID判定，曲线缺失点小于曲线允许最大缺失值时，使用下面的3种规则进行补全。
3. 最左边数据缺失，取已有最左边一点的数据进行等值补全；
4. 最右边数据缺失，取已有最右边一点的数据进行等值补全；
5. 中间数据缺失，取最近有数值的两点进行线性插值补全；线性插值中，每个点的期限的天数由配置表提供。线性插值计算公式：y=y1+(y2-y1)\*(x-x1)/(x2-x1)

##### LIBOR市场数据补全规则

1. 前置处理：
   1. 将源表的数据代码（SYMBOL），币种（CURRENCY）拼接成唯一的标识，并加载进接口表的RIC字段。
   2. 将源表的报价（OFFEREDRATE），统一加载接口表的UNI\_ASK,UNI\_BID,PRI\_ACT 中。
   3. 最新价类型，统一标识为：99
   4. 接口表的交易日期，交易时间，取业务数据日期和系统时间点。
   5. 当该RIC码的RIC码类型为OUTRIGHT（不允许为0）时，如果价格为0，当作该价格丢失，按规则补全价格，如果RIC码类型为SPREAD(允许为0)，则允许价格为0。
2. RIC码缺失补全，按以下的规则进行。

主要分为以下2种类型：

* + - 取昨天数据，以配置表（IAL.INT\_RIC\_CFG）中的GROUP\_ID分组判定，缺失大于或等于曲线允许最大缺失值时，所有曲线点沿用昨天的数据补全。注：交易日期，交易时间，收市日期取当天跑批的系统时间。
    - 左右找平或中间插值，以GROUP\_ID判定，曲线缺失点小于曲线允许最大缺失值时，使用下面的3种规则进行补全。

1. 最左边数据缺失，取已有最左边一点的数据进行等值补全；
2. 最右边数据缺失，取已有最右边一点的数据进行等值补全；
3. 中间数据缺失，取最近有数值的两点进行线性插值补全；线性插值中，每个点的期限的天数由配置表提供。线性插值计算公式：y=y1+(y2-y1)\*(x-x1)/(x2-x1)

##### 手工补录市场数据补全规则

1. 前置处理：
   1. 将源表的数据代码（SGE\_RIC）作为唯一的标识，并加载进接口表的RIC字段。
   2. 将源表的报价（CD\_VAL），统一加载接口表的UNI\_ASK,UNI\_BID,PRI\_ACT 中。
   3. 最新价类型，统一标识为：99
   4. 接口表的交易日期，交易时间，取业务数据日期和系统时间点。
   5. 当该RIC码的RIC码类型为OUTRIGHT（不允许为0）时，如果价格为0，当作该价格丢失，按规则补全价格，如果RIC码类型为SPREAD(允许为0)，则允许价格为0。
2. RIC码缺失补全，按以下的规则进行。

主要分为以下2种类型：

* + - 取昨天数据，以配置表（IAL.INT\_RIC\_CFG）中的GROUP\_ID分组判定，缺失大于或等于曲线允许最大缺失值时，所有曲线点沿用昨天的数据补全。注：交易日期，交易时间，收市日期取当天跑批的系统时间。
    - 左右找平或中间插值，以GROUP\_ID判定，曲线缺失点小于曲线允许最大缺失值时，使用下面的3种规则进行补全。

1. 最左边数据缺失，取已有最左边一点的数据进行等值补全；
2. 最右边数据缺失，取已有最右边一点的数据进行等值补全；
3. 中间数据缺失，取最近有数值的两点进行线性插值补全；线性插值中，每个点的期限的天数由配置表提供。线性插值计算公式：y=y1+(y2-y1)\*(x-x1)/(x2-x1)

##### 大宗商品（LME）市场数据补全规则

1. LME前置处理：
   1. 将源表的RIC加载进接口表的RIC字段。
   2. 将源表的卖价（ASK\_PRICE）加载到接口表的UNI\_ASK字段，买价（BID\_PRICE）加载到接口表的UNI\_BID 字段，结算价（SETTLEMENT\_PRICE） 加载到接口表的PRI\_ACT字段。

|  |  |  |  |
| --- | --- | --- | --- |
| ASK | BID | PAC | 取值规则 |
| √ | √ | ╳ | 优先顺序PRI\_ACT1  ->STL\_PRICE |
| √ | ╳ | √ | ASK |
| √ | ╳ | ╳ | BID: ASK  PAC: PRI\_ACT1  ->STL\_PRICE |
| ╳ | √ | √ | BID |
| ╳ | √ | ╳ | ASK :BID  PAC: PRI\_ACT1  ->STL\_PRICE |
| ╳ | ╳ | √ | STL\_PRICE |
| ╳ | ╳ | ╳ | 进入市场数据缺失补全规则 |

* 1. 最新价类型，统一标识为：99
  2. 接口表的交易日期，交易时间，取业务数据日期和系统时间点。
  3. 当该RIC码的RIC码类型为OUTRIGHT（不允许为0）时，如果价格为0，当作该价格丢失，按规则补全价格，如果RIC码类型为SPREAD(允许为0)，则允许价格为0。

1. RIC码缺失补全，按以下的规则进行。

主要分为以下2种类型：

* + 取昨天数据，以配置表（IAL.INT\_RIC\_CFG）中的GROUP\_ID分组判定，缺失大于或等于曲线允许最大缺失值时，所有曲线点沿用昨天的数据补全。注： 交易日期，交易时间，收市日期取当天跑批的系统时间。
  + 左右找平或中间插值，以GROUP\_ID判定，曲线缺失点小于曲线允许最大缺失值时，使用下面的3种规则进行补全。

1. 最左边数据缺失，取已有最左边一点的数据进行等值补全；
2. 最右边数据缺失，取已有最右边一点的数据进行等值补全；
3. 中间数据缺失，取最近有数值的两点进行线性插值补全；线性插值中，每个点的期限的天数由配置表提供。线性插值计算公式：y=y1+(y2-y1)\*(x-x1)/(x2-x1)

##### 大宗商品（OTHCOM）市场数据补全规则

1. OTHCOM前置处理：
   1. 将源表的RIC加载进接口表的RIC字段。
   2. 将源表的最新价1（PRIMARY\_ACTIVITY）加载到接口表的UNI\_ASK字段、 UNI\_BID 字段，结算价（SETTLEMENT\_PRICE） 加载到接口表的PRI\_ACT字段。(注：此处加载，是为了后续曲线补全统一取值)

|  |  |  |  |
| --- | --- | --- | --- |
| ASK | BID | PAC | 取值规则 |
| √ | √ | ╳ | 优先顺序  PRI\_ACT1  ->STL\_PRICE |
| √ | ╳ | √ | 优先顺序LAST\_PIRCCE  ->STL\_PRICE |
| √ | ╳ | ╳ | BID: LAST\_PIRCCE  ->STL\_PRICE  PAC: PRI\_ACT1  ->STL\_PRICE |
| ╳ | √ | √ | LAST\_PIRCCE  ->STL\_PRICE |
| ╳ | √ | ╳ | ASK : LAST\_PIRCCE  ->STL\_PRICE  PAC: PRI\_ACT  ->STL\_PRICE |
| ╳ | ╳ | √ | LAST\_PIRCCE  ->STL\_PRICE |
| ╳ | ╳ | ╳ | 进入市场数据缺失补全规则 |

* 1. 最新价类型，统一标识为：99
  2. 接口表的交易日期，交易时间，取业务数据日期和系统时间点。
  3. 当该RIC码的RIC码类型为OUTRIGHT（不允许为0）时，如果价格为0，当作该价格丢失，按规则补全价格，如果RIC码类型为SPREAD(允许为0)，则允许价格为0。

1. RIC码缺失补全，按以下的规则进行。

主要分为以下2种类型：

* + 取昨天数据，以配置表（IAL.INT\_RIC\_CFG）中的GROUP\_ID分组判定，缺失大于或等于曲线允许最大缺失值时，所有曲线点沿用昨天的数据补全。注：交易日期，交易时间，收市日期取当天跑批的系统时间。
  + 左右找平或中间插值，以GROUP\_ID判定，曲线缺失点小于曲线允许最大缺失值时，使用下面的3种规则进行补全。

1. 最左边数据缺失，取已有最左边一点的数据进行等值补全；
2. 最右边数据缺失，取已有最右边一点的数据进行等值补全；
3. 中间数据缺失，取最近有数值的两点进行线性插值补全；线性插值中，每个点的期限的天数由配置表提供。线性插值计算公式：y=y1+(y2-y1)\*(x-x1)/(x2-x1)

##### 掉期期权（SWAPTION）市场数据补全规则

1. 前置处理：
   1. 将源表的数据代码（RIC）作为唯一的标识，并加载进接口表的RIC字段。
   2. 将源表的报价（TX\_PRICE），统一加载接口表的UNI\_ASK,UNI\_BID,PRI\_ACT 中。
   3. 最新价类型，统一标识为：99
   4. 接口表的交易日期，交易时间，取业务数据日期和系统时间点。
   5. 当该RIC码的RIC码类型为OUTRIGHT（不允许为0）时，如果价格为0，当作该价格丢失，按规则补全价格，如果RIC码类型为SPREAD(允许为0)，则允许价格为0。
2. RIC码缺失补全，按以下的规则进行。

主要分为以下2种类型：

* + - 取昨天数据，以配置表（IAL.INT\_RIC\_CFG）中的GROUP\_ID分组判定，缺失大于或等于曲线允许最大缺失值时，所有曲线点沿用昨天的数据补全。注：交易日期，交易时间，收市日期取当天跑批的系统时间。
    - 左右找平或中间插值，以GROUP\_ID判定，曲线缺失点小于曲线允许最大缺失值时，使用下面的3种规则进行补全。

1. 最左边数据缺失，取已有最左边一点的数据进行等值补全；
2. 最右边数据缺失，取已有最右边一点的数据进行等值补全；
3. 中间数据缺失，取最近有数值的两点进行线性插值补全；线性插值中，每个点的期限的天数由配置表提供。线性插值计算公式：y=y1+(y2-y1)\*(x-x1)/(x2-x1)

#### 设计总体思路

1. 处理的总体数据流，如下图所示。

将处理缺失补全的整个逻辑，封装在存储过程里面，生成相应的结果表。



1. 市场数据统一接口数据表，表结构如下：

|  |  |  |
| --- | --- | --- |
| **英文字段名** | **中文名称** | **说明** |
| RIC | 路透标示码 |  |
| GROUP\_ID | 组合ID |  |
| CURV\_TERM\_STD | 标准期限 |  |
| UNI\_ASK | 买价 |  |
| UNI\_BID | 卖价 |  |
| PRI\_ACT | 最新价 |  |
| PRI\_ACT\_TYPE | 最新价类型 |  |
| QUOTE\_DATE | 交易日期，针对最新价 |  |
| QUOTE\_TIME | 交易时间，针对最新价 |  |
| UNI\_BID\_ASK\_DT | 买价卖价日期 |  |
| INSERT\_TYPE | 插值类型 |  |
| BATCH\_NO | 批次 |  |
| DATA\_SRC | 数据来源 |  |
| DATE\_DT | 业务数据日期 |  |
| ETL\_DATE | 数据加载日期 |  |

1. 设置缺失补全的配置表，用以保存缺失补全的基本字段属性，并作为缺失补全数据的参考基础。

缺失补全的配置表如下：

|  |  |  |
| --- | --- | --- |
| **字段英文名称** | **字段中文名称** | **字段说明** |
| RIC | RIC码 |  |
| GROUP\_ID | 组合ID |  |
| CURV\_TERM | 期限 |  |
| CURV\_TERM\_STD | 标准期限 | 期限需要转换为标准期限后才能进行插值计算 |
| CURV\_POINT | 曲线允许最大缺失值 | 当曲线缺失值大于等于该值时候用昨天的整条曲线代替今天曲线上的所有点 |
| DATA\_SRC | 数据源标识 | 数据来源表的缩写 |
| RIC\_TYPE | RIC码类型 | OUTRIGHT --不允许为0  SPREAD --允许为0 |

1. RIC缺失补全结果表

|  |  |  |
| --- | --- | --- |
| **字段英文名称** | **字段中文名称** | **字段说明** |
| RIC | RIC码 |  |
| GROUP\_ID | 组合ID |  |
| UNI\_ASK | 买价 |  |
| UNI\_BID | 卖价 |  |
| PRI\_ACT | 最新价 |  |
| PRI\_ACT\_TYPE | 最新价类型 | 1-买入报价  2-卖出报价  3-购买价格 |
| QUOTE\_DATE | 报价日期 |  |
| QUOTE\_TIME | 报价时间 |  |
| UNI\_BID\_ASK\_DT | 收市日期 |  |
| INSERT\_TYPE | 插值类型 | 0-没缺失正常值  1-取当天数据值 2-取昨天数据值  3-左找平值  4-取昨天数据值，但无法取值，全部赋值为0 5-线性插值  6-右找平值。  格式：000或010或333或222 |
| BATCH\_NO | 批次 |  |
| DATA\_SRC | 数据来源 |  |
| DATE\_DT | 业务数据日期 |  |
| ETL\_DATE | 数据加载日期 |  |

1. 处理总体步骤：

* 补齐源数据的RIC。若无则跳过这步。
* 补齐源数据三种价格缺少部分数据情况。若无则跳过这步
* 补全曲线缺失点大于等于允许最大缺失值的数据。若无则跳过这步
* 补全曲线左右两端缺失需要左右找平的数据。若无则跳过这步。
* 补全曲线中间缺失需要线性插值的数据。

#### 假设条件

* 每种曲线都有标准的期限全集定义，否则无法判断是否存在缺失。-因为每条曲线都都需要提前购买，所以理论上我们可以提前将相关曲线的的期限全集提前更新到相关的参数表中。

#### 特殊处理

* 初始化因为没有昨日数据，针对需要取昨日数据补全的要求，必须手工先完善数据再进行处理。