壹、專案起源

貳、資料來源與儲存

參、資料處理

肆、回歸分析

貳、資料來源與儲存

在此一小節，我們共會有三類的資料檔案，依序為台灣股價資料、因子價格資料，以及想要測試的投資組合，以下依序做細部解釋：

參、資料處理

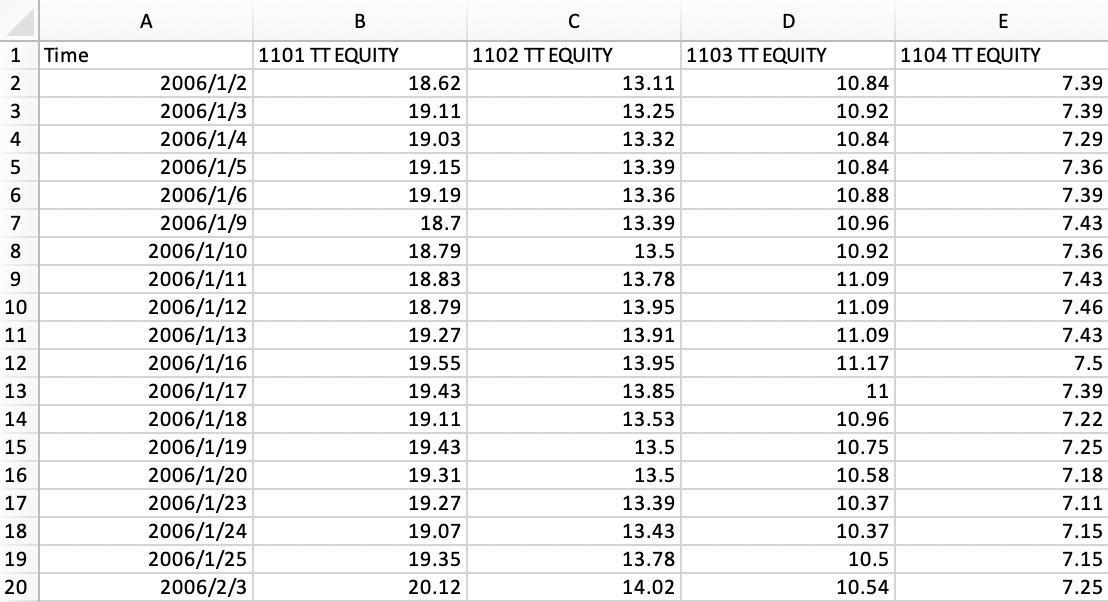
資料處理的部份，可以視為應變數與自變數兩區塊。其中應變數為要預測的投資組合報酬，單純只是計算給定一投資組合，隨者時間改變的報酬率 ; 自變數為用來預測的因子，則因為了滿足回歸分析的模型性質，或以利解釋結果，需對資料做較多的加工，而每一個加工的流程，我們都定義為一個函數。

* 應變數資料：

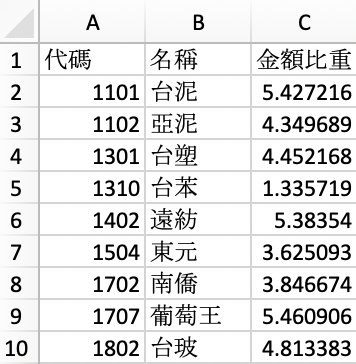
Step 1:

首先，我們在從Bloomberg抓下股價資料時，就先利用了excel的vlookup將股價時間對齊了（若當日無交易，則會以前一天收盤價自動填入，因為將vlookup第四個參數設為1，可以補入最相近者，而剛好是前一天收盤價）。

所以我們在用程式讀取時，預設股價檔案的長相如下：



我們也會有一組想要分析風險與報酬性質的投組，輸入格式如下：



Step 2:

我們計算報酬的思路為：

1. 我們會計算每一天的股票的“ 1 + 報酬率 ”，如：台積電連續三天的價格為100、110、121，那第一天的“ 1 + 報酬率 ”為110 / 100 =1.1 ; 第二天的“ 1 + 報酬率 ”為121 / 110 =1.1。
2. 想要得到投資組合的“ 1 + 報酬率 ”就將每一檔股票的“ 1 + 報酬率 ”乘上持有的權重。
3. 想要得到投資組合的累積報酬率，就將每一天的“ 1 + 報酬率 ”相乘。如果整個投組只有台積電，投組從第一天到第三天的報酬就會是 1.1 \* 1.1 – 1 = 0.21

Note: 我們丟進去的stock\_price為一個DataFrame，最後返回的portfolio\_return也是一個DataFrame。

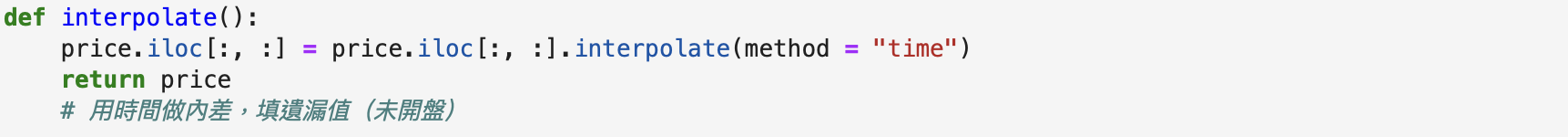
以下為此部分程式碼：



* 應變數資料：

函數1: 補遺漏值

我們拿取的資料可能因所屬國家不同，休市時間不一，因此我們利用時間長短來做內差法，補足空缺的資料。



函數2: 做標準化

因為不同的因子，其資料尺度不同，將資料做標準化，將來回歸係數的大小可以來拿比較對於自變數的影響大小。



函數3: 做ICA

ICA 為將原本有一定程度相依的因子，轉化為獨立的因子的技術。如匯率與利率的關係數不等於0，而ICA套件所做的事就是計算出一組新的因子，其因子間的相關係數為0，而新產生的因子與原因子的相關性要最大。

