

總體經濟多因子股市預測模型

蘇彥庭

一、報告摘要

此篇報告透過總體經濟變數，採用加權最小二乘法建立多因子模型，預測滬深 300 指數月報酬率的漲跌。該模型的建立採用簡化計量方法，將各個總體經濟時間序列變數的趨勢轉化為 1 及 0 兩類的狀態變數。並將此狀態變數序列對滬深 300 指數進行單變數迴歸，檢定總體經濟變數是否與股市之間具有顯著關係。找出對滬深 300 指數有顯著關係的總體經濟變數後，以加權最小二乘法的逐步迴歸方式在樣本期內建立出多因子模型，預測樣本外的滬深 300 指數的月報酬率漲跌。

回測資料期間為 2002 年 1 月至 2015 年 8 月，樣本期採用移動窗格的方法，保證 60 個月的資料長度，每月重新建立多因子模型進行預測，樣本外投資期間為 2007 年至 2 月至 2015 年 8 月。本報告回測的預測準確率為 66.99%，預測漲的條件下實際漲的準確率為 66.15%，預測跌的條件下實際跌的準確率為 68.42%。在未考慮交易成本及停損的情況下，單純做多的累積收益率為 526.67%，年化收益率為 14.47%，最大回撤率為 -32.30%。而多空皆做的累積收益率為 1899.02%，年化收益率為 24.67%，最大回撤率為 -26.72%。

二、總體經濟因子

此篇報告使用 Wind 總體經濟資料庫的資料，挑選出表 1 的總體經濟因子。大致上可將所有總體經濟因子區分為八大面向，分別是貨幣政策、工業生產、採購經理人指數、物價、匯率、對外貿易、利率及總體經濟指數。另外考慮到官方發佈總體資料會有落後的問題，表 1 整理出各指標的落後期數，是以預測月份做

為基準期(T)。舉例來說，生產者物價指數(PPI)的落後期為T-2，則代表若模型要預測11月份的報酬率漲跌時，只能使用9月(含)以前的資料來進行迴歸預測。

表 1 總體經濟因子

類別	總體經濟變數	資料來源	Wind 資料庫 數據變數名稱	落後期數
貨幣政策	流通中現金	中國人民銀行	M0: 同比	T-2
	狹義准貨幣	中國人民銀行	M1: 同比	T-2
	廣義准貨幣	中國人民銀行	M2: 同比	T-2
	狹義准貨幣與 廣義准貨幣的 增速差	中國人民銀行	M1: 同比-M2: 同比	T-2
工業生產	工業增加值	國家統計局	工業增加值: 當月同比	T-2
	全國發電量	國家統計局	產量: 發電量: 當月值: 同比	T-2
採購經理人指數	採購經理指數	國家統計局	PMI	T-1
	採購經理指數 產成品庫存	國家統計局	PMI: 產成品庫存	T-1
	採購經理指數 購進價格	國家統計局	PMI: 主要原材料購進價格	T-1
	採購經理指數 供應商配送時間	國家統計局	PMI: 供應商配送時間	T-1
物價	消費者 物價指數	國家統計局	CPI: 當月同比(月)	T-2

	生產者 物價指數	國家統計局	PPI:全部工業品:當月同 比	T-2
	企業商品 交易價格指數	國家統計局	CGPI:當月同比(月)	T-2
匯 率	外匯匯率	中國人民銀行	平均匯率: 美元兌人民幣:同比	T-1
	外匯儲備	中國人民銀行	官方儲備資產: 外匯儲備:同比	T-2
對 外 貿 易	進出口差額	海關總署	貿易差額:當月同比	T-2
	進口額	海關總署	進口金額:當月同比	T-2
	出口額	海關總署	出口金額:當月同比	T-2
利 率	短期貸款餘額	中國人民銀行	金融機構: 短期貸款餘額:同比	T-2
	上海銀行間 同業拆放利率	全國銀行間 同業拆借利率	SHIBOR:3個月:月平均值	T-1
	10年國債到期 收益率與1年國 債到期收益率 利差	中國債券資訊網	(中債國債到期收益率:10 年-中債國債到期收益 率:1年):月平均值	T-1
總 體 經 濟 指 數	總體經濟景氣 指數:先行指數	國家統計局	總體經濟景氣指數: 先行指數	T-2
	總體經濟景氣 指數:一致指數	國家統計局	總體經濟景氣指數: 一致指數	T-2
	消費者信心 指數	國家統計局	消費者信心指數:月	T-2
	消費者預期 指數	國家統計局	消費者預期指數:月	T-2

	投資者信心指數	中國證券投資者 保護基金公司	投資者信心指數:總指數	T-2
--	---------	-------------------	-------------	-----

三、狀態變數計算方法

狀態變數主要是依據總體經濟指標的趨勢來判斷，若指標呈上升趨勢，則狀態變數為 1，若指標呈下降趨勢，則狀態變數為 0。然而在現實世界中，投資者是無法知道未來的價格走勢，因此無法確定現在的股市是處於上升趨勢還是下降趨勢。因此本報告利用過去的價格序列，透過簡單的計量方法來判別當前的趨勢。此處給定一個趨勢參數 j ，若當前月份的總體經濟指標數值大於(小於)過去 j 個月，則當前月份視為是上升(下降)趨勢，若無法判斷當前月份是否大於或小於前面 j 個月的數值，則將前一個月的狀態變數賦值給當前月份。因此當 j 的取值較小時，可以立即判斷出趨勢，但識別的趨勢是屬於短期趨勢；當 j 的取值較大時，雖可以識別出大波段的趨勢，但此時識別的趨勢會有落後於股市的現象。故本報告在回測因子顯著性時，趨勢參數範圍為 $j=1$ 至 4。圖 1 是在參數 $j=3$ 之下，以 M1 與 M2 增速差因子做趨勢判斷方法的範例圖形，可以很明顯看出此方法在做上升趨勢判斷時會有稍微落後的現象。

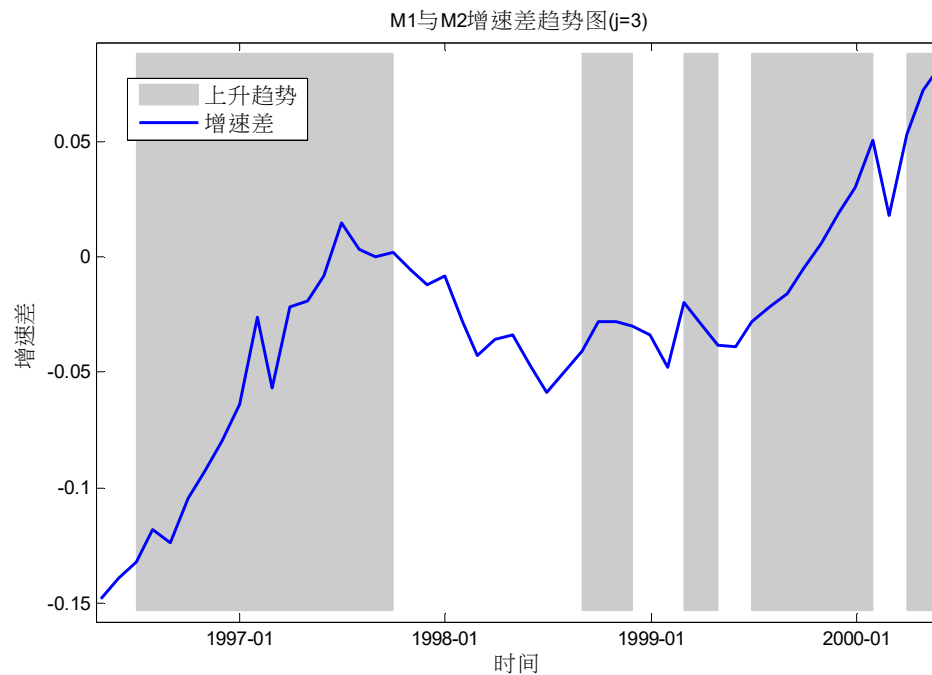


圖 1 M1 與 M2 增速差趨勢圖

四、加權最小二乘法

本研究挑選因子的迴歸模型採用加權最小二乘法。加權最小二乘法迴歸的主要是基於普通最小二乘的基礎上，給不同的樣本點賦予不同的權重之後再以加權的誤差平方和最小做為準則估計參數。此處加權的方式採用線性加權的方式，計算方式如下：

$$\text{weight}(t) = \frac{t}{T} * \lambda + (1 - \lambda), t = 1, \dots, T \quad (1)$$

其中， t 是樣本點發生的時間， T 是總樣本長度， λ 是參數。圖 2 是模擬在不同參數 λ 下的各期權重變化，當 λ 值參數設定愈大，愈早期的權重衰減速度會愈快。本報告將 λ 參數設置為 0.9。

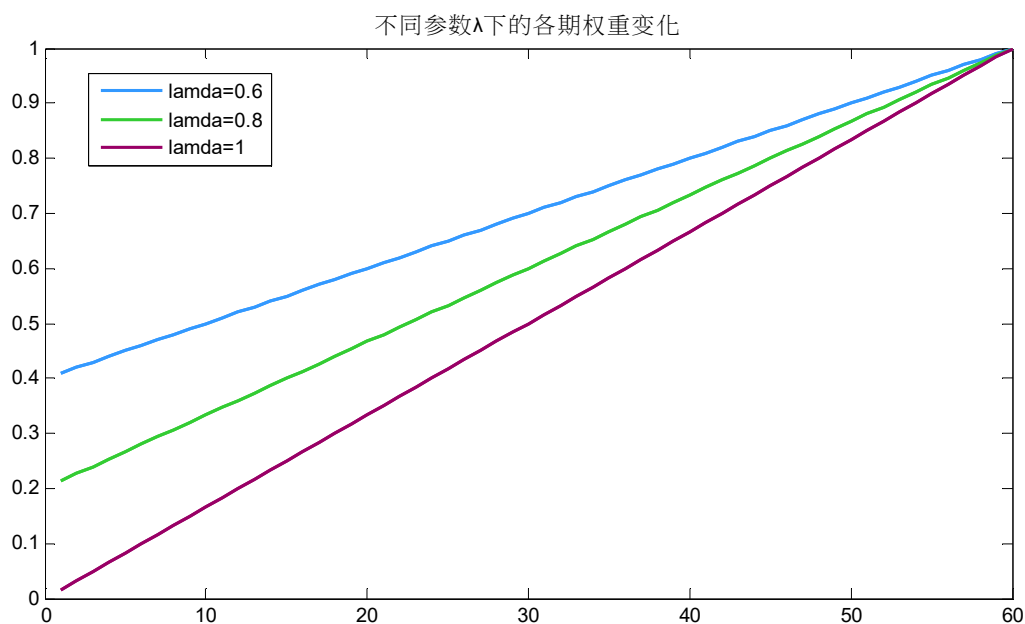


圖 2 不同參數 λ 下的各期權重變化

五、回測期間設定

回測資料期間為 2002 年 1 月至 2015 年 8 月，採用移動窗格的方法，保證 60 個月的資料長度，每月重新建立多因子模型進行預測，樣本外投資期間為 2007 年至 2 月至 2015 年 8 月。回測期間設定如圖 3 所示。



圖 3 回測期間設定示意圖

六、單因子顯著性檢定

在各總體經濟指標序列的趨勢識別出來後，將狀態變數做為解釋變數，滬深 300 指數的月報酬率做為被解釋變數，進行單變數加權最小二乘法迴歸，檢測各個總體經濟指標在樣本期間內是否對股市月報酬率顯著。此處考慮到官方發佈資料的時間，因此會將解釋變數以落後期的方式進行迴歸。例如說在 10 月底欲預測 11 月的股市漲跌，此時只有 9 月中旬公佈的 8 月份 PPI 指標能使用。以預測月份做為基準期，大部分的總體經濟指標公佈時間會落後 2 期(例如 M1、M2、PPI 等)，少部分則是落後 1 期(PMI、Shibor 利率等)，故單因子顯著性檢定的迴歸方程式如式(2)所示。

$$Y_t = \alpha + \beta_1 X_{j,t-k} + \varepsilon \quad (2)$$

其中趨勢參數 $j=1, 2, \dots, 5$ ，落後參數 $k=1, 2$

Y_t ：滬深 300 指數 t 期月報酬率

$X_{j,t-k}$ ：總體經濟指標在趨勢參數 j 下的 $t-k$ 期狀態變數

七、多因子模型

跑完樣本期內的單因子顯著性檢定後，各總體經濟指標會有在不同趨勢參數組合(j, k)下的 t 統計量值。在每個總體經濟因子中，選擇最大 t 統計量絕對值的參數組合(j, k)做為計算總體經濟因子狀態變數的最適參數。另外，若各因子的最大 t 統計量絕對值未大於 1.96，代表該總體經濟因子在樣本期內不顯著，不納入下一階段的多因子模型篩選。

經過單因子顯著性檢定的初篩後，接下來將顯著的總體經濟因子及其最適參數所計算出來的狀態變數，做為逐步迴歸的解釋變數，而滬深 300 指數月報酬率做為被解釋變數。值得注意的是，此處做的逐步迴歸模型是以加權最小二乘法做為基礎進行。經過逐步迴歸建立出多因子模型後，將樣本外欲預測月份的因子狀態變數代入多因子模型，得出預期報酬率。若預期報酬率為正值，則為買入訊號，若預期報酬率為負值，則為賣出訊號。

八、回測結果

樣本外投資期間為 2007 年至 2 月至 2015 年 8 月，共計有 103 筆預測資料。本報告將預測結果進行簡單的回測報酬計算，不考慮交易成本及停損，計算方式如式(3)所示。

$$R_{cum} = (1 + P_1 R_1) \times (1 + P_2 R_2) \times \dots \times (1 + P_T R_T) - 1 \quad (3)$$

R_{cum} ：策略累積報酬率

P_t ：t 月的預測漲跌訊號，若預測為漲則為 1，若預測為跌則為-1

R_1 ：股市指數在 t 月的報酬率

回測結果如表 2 所示，累計收益圖如圖 4 及圖 5 所示。

表 2 回測結果表

年份	準確率	預測漲之下 實際漲準確率	預測跌之下 實際跌準確率
2007	90.91%	90.00%	100.00%
2008	58.33%	40.00%	71.43%
2009	91.67%	91.67%	—
2010	66.67%	75.00%	62.50%
2011	66.67%	42.86%	100.00%
2012	41.67%	45.45%	0.00%
2013	58.33%	50.00%	75.00%
2014	58.33%	75.00%	50.00%
2015	75.00%	75.00%	75.00%
全部	66.99%	66.15%	68.42%
	累計收益率	年化收益率	最大回撤率
只做多投資結果	526.67%	14.47%	-32.30%
多空皆做投資結果	1899.02%	24.67%	-26.72%

樣本外多因素模型擇時效果(只做多)

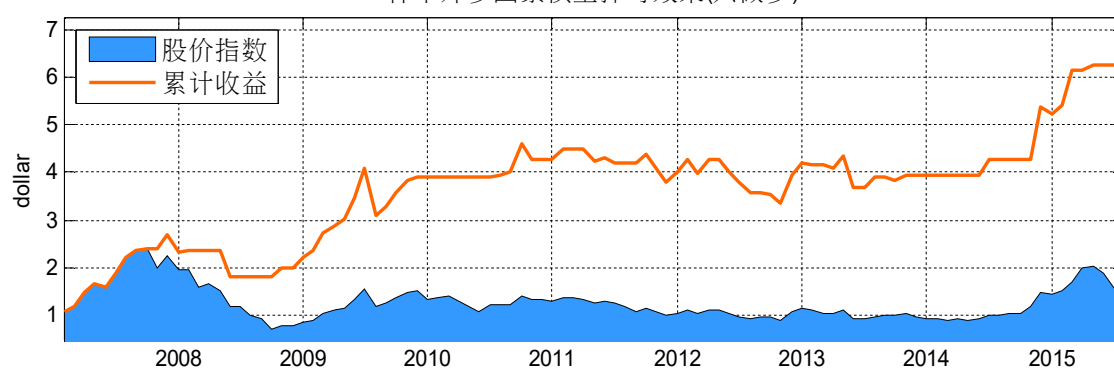


圖 4 策略只做多累積收益圖

樣本外多因素模型擇時效果(多空皆做)

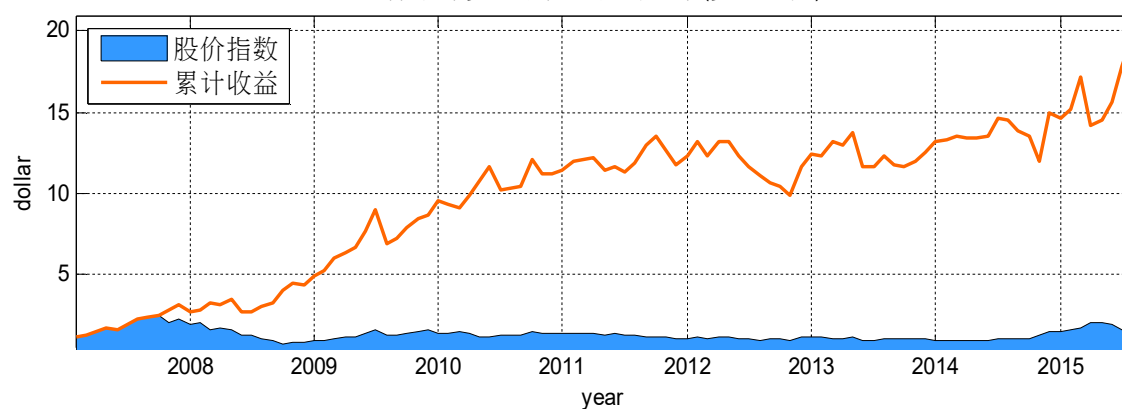


圖 5 策略多空皆做累積收益圖

八、參數敏感度分析

本模型主要參數有下列二個：權重參數 λ 及趨勢參數 J 。表 3 及表 4 列出在固定其他參數下，單獨調整權重參數 λ 及趨勢參數 J ，對各年準確率及年化報酬率的影響。其中表格內標示粗體的部分表示本模型回測所使用的參數結果。

表 3 權重參數 λ 下的各年準確率及年化報酬率

lamda 參數	0.5	0.7	0.8	0.9	0.1
2007	81.82%	90.91%	81.82%	90.91%	72.73%
2008	41.67%	41.67%	58.33%	58.33%	58.33%
2009	83.33%	83.33%	91.67%	91.67%	91.67%
2010	66.67%	66.67%	66.67%	66.67%	66.67%
2011	66.67%	66.67%	66.67%	66.67%	66.67%
2012	50.00%	50.00%	41.67%	41.67%	66.67%
2013	50.00%	45.45%	45.45%	58.33%	58.33%
2014	63.64%	58.33%	58.33%	58.33%	50.00%
2015	75.00%	75.00%	75.00%	75.00%	75.00%
年化報酬率	14.86%	16.41%	21.54%	24.67%	21.78%

表 4 趨勢參數 J 下的各年準確率及年化報酬率

jmax 參數	3	4	5	6	7	8	9
2007	90.91%	90.91%	63.64%	72.73%	81.82%	81.82%	81.82%
2008	50.00%	58.33%	41.67%	58.33%	50.00%	50.00%	58.33%
2009	83.33%	91.67%	91.67%	75.00%	91.67%	75.00%	83.33%
2010	66.67%	66.67%	66.67%	66.67%	66.67%	58.33%	50.00%
2011	75.00%	66.67%	66.67%	66.67%	66.67%	66.67%	58.33%
2012	50.00%	41.67%	41.67%	50.00%	50.00%	50.00%	50.00%
2013	58.33%	58.33%	58.33%	58.33%	58.33%	58.33%	50.00%
2014	58.33%	58.33%	58.33%	58.33%	58.33%	58.33%	66.67%
2015	62.50%	75.00%	62.50%	62.50%	62.50%	62.50%	62.50%
年化報酬率	21.84%	24.67%	13.09%	16.10%	18.98%	14.55%	17.92%