

高維度變數下 總體經濟預測模型的建立

指導學生：盧奕盛

指導教授：林常青 博士

研究背景

經濟預測能提供未來景氣變化的參考依據，台灣經濟研究院、中研院經濟所、主計處、中央銀行等機構的總體經濟計量模型，具有預測多個經濟變量及可進行敏感性分析的優點。而在總體經濟季資料長度遠小於資料的維度，為避免過度配適，過去文獻基於經濟理論挑選變數後再排除不顯著的變數。

Fair (1996), “Computational Methods for Macroeconometric Models,”

吳中書等人 (2001)，「臺灣總體經濟計量動態季模型」

林建甫 (2006)，「台灣總體經濟金融模型之建立」

林建甫 (2010)，「總體經濟計量模型的建立與應用」

研究背景

徐士勛等人 (2005) 運用特徵萃取的降維方式，使用國內各市場的變數組成擴散指標，建立經濟成長率的預測模型。

陳宜廷等人 (2011) 在預測更新模型與擴散指標模型比較結果發現，預測經濟成長率的模型以主計處預測值作為解釋變數下，加入擴散指標能提升預測表現。

Li and Chen (2014) 於動態因子模型採用特徵選取方法，透過 Lasso 迴歸交叉驗證的方式，為模型挑選對預測重要的因子。

因子模型相對於總體經濟計量模型最大的差異，在於無法描繪出實際變數之間的影響關係，也就無法進行模擬政策施行等敏感性分析。

研究動機與架構

透過特徵選取以及萃取的降維方式，降低總體經濟計量模型設定上的困難，以及提升模型預測的能力。

1. 採用兩階段變數選擇方法，經 Lasso 迴歸交叉驗證篩選出重要的變數，再利用逐步迴歸法找尋出最高的調整後 R^2 變數組合，以此設定各迴歸解釋變數。
2. 透過主成分分析組成國外市場因子，加入模型中迴歸選擇變數範圍。

總體經濟計量模型

將模型設定的定義式與結構方程式(迴歸估計式轉換成當期函數)，每一期聯立求解。

求解方式：

1. 靜態求解：採用向後一期預測，內生變數落後期皆代入真實值。(樣本內)
2. 動態求解：內生變數落後期皆代入預測值。(樣本外)

舉例說明：令一小型模型內生變數有GDP,C,I，外生變數有G， $GDP = C + I + G$

迴歸估計式：

- $C = \hat{\beta}_0 + \hat{\beta}_1 C(-1) + \hat{\beta}_2 GDP + \hat{\beta}_3 D(I)$
- $I = \hat{\beta}_0 + \hat{\beta}_1 I(-1) + \hat{\beta}_2 G + \hat{\beta}_3 pch(C)$



聯立方程式：

- $C - \hat{\beta}_2 GDP - \hat{\beta}_3 I = \hat{\beta}_0 + \hat{\beta}_1 C(-1) - \hat{\beta}_3 I(-1)$
- $I - \frac{\hat{\beta}_3}{C(-4)} C = \hat{\beta}_0 + \hat{\beta}_1 I(-1) + \hat{\beta}_2 G - \hat{\beta}_3$
- $GDP - C - I = G$
- 每一期求解GDP , C , I

註： $D(I) = I - I(-1)$, $pch(C) = \frac{C - C(-4)}{C(-4)}$

模型定義式的設定

每一期聯立方程式有包含如下6條定義式：

$$1. \text{ GDP} = \text{CF} + \text{CO} + \text{IBF} + \text{IG} + \text{IPC} + \text{J} + \text{CG} + \text{EX} - \text{M}$$

$$2. \text{ real GDP} = \frac{\text{GDP}}{\text{PGDP}} \times 100$$

$$3. \text{ growth rate} = \frac{\text{real GDP} - \text{real GDP}(-4)}{\text{real GDP}(-4)}$$

$$4. \text{ NE} = (1 - 0.01 \text{ NU}) \text{ NF}$$

$$5. \text{ PDT} = \frac{\text{GDP}}{\text{NE}} \text{ (勞動生產力指數)}$$

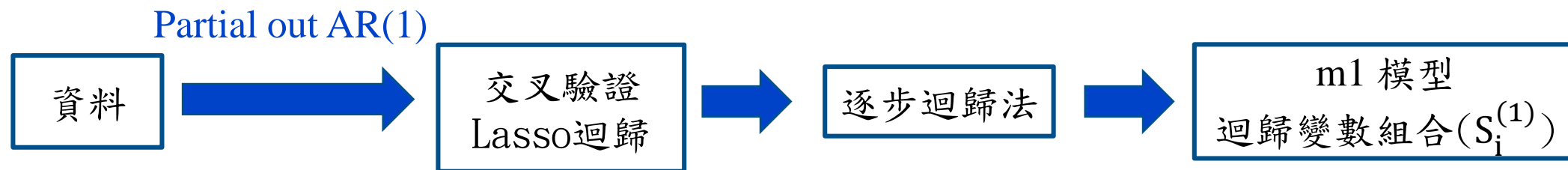
$$6. \text{ ULC} = \frac{\text{PWM}}{\text{PDT}} \text{ (單位產出勞動成本指數)}$$

結構方程式的變數選擇流程

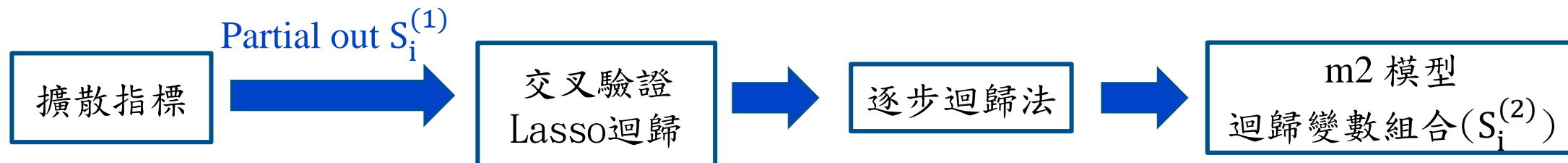
使用重要的經濟及金融數據的當期值、落後一期項、一階差分項、年增率，以及季節虛擬變數共206個變數(X)，其中有28個內生變數要各別進行系統化的挑選解釋變數。

以下為兩種模型設定中，每一條迴歸的變數選擇流程：

1. 國內均衡 (m1) 模型：



2. 經國外調整 (m2) 模型：

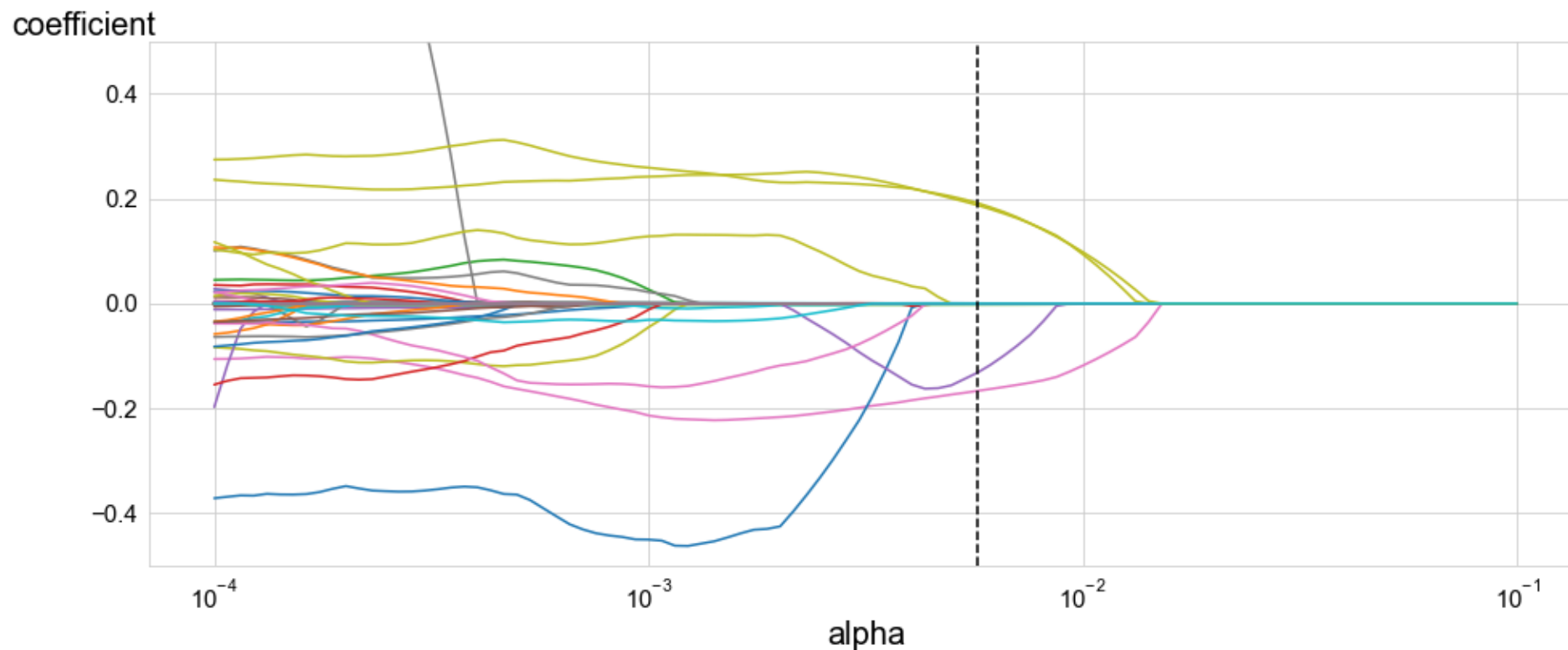


Lasso 迴歸

$$\text{minimize}_{\beta_1, \beta_2, \dots, \beta_k} \underbrace{\frac{1}{T} \sum_t \left(y_t - \beta_0 - \sum_k \beta_k x_{tk} \right)^2}_{\text{最小平方方法的誤差項}} + \underbrace{\alpha \sum_k |\beta_k|}_{\text{懲罰項}}$$

- α : 懲罰項權重
- 目標式: 讓誤差總和與懲罰項加總最小

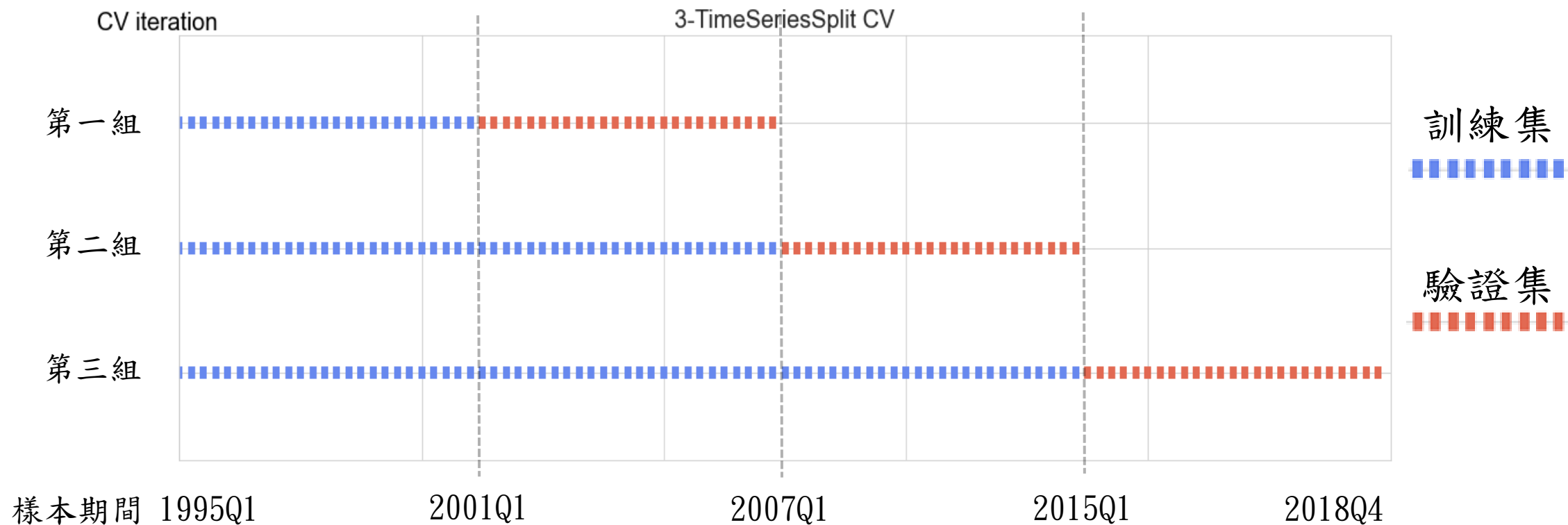
Lasso 迴歸



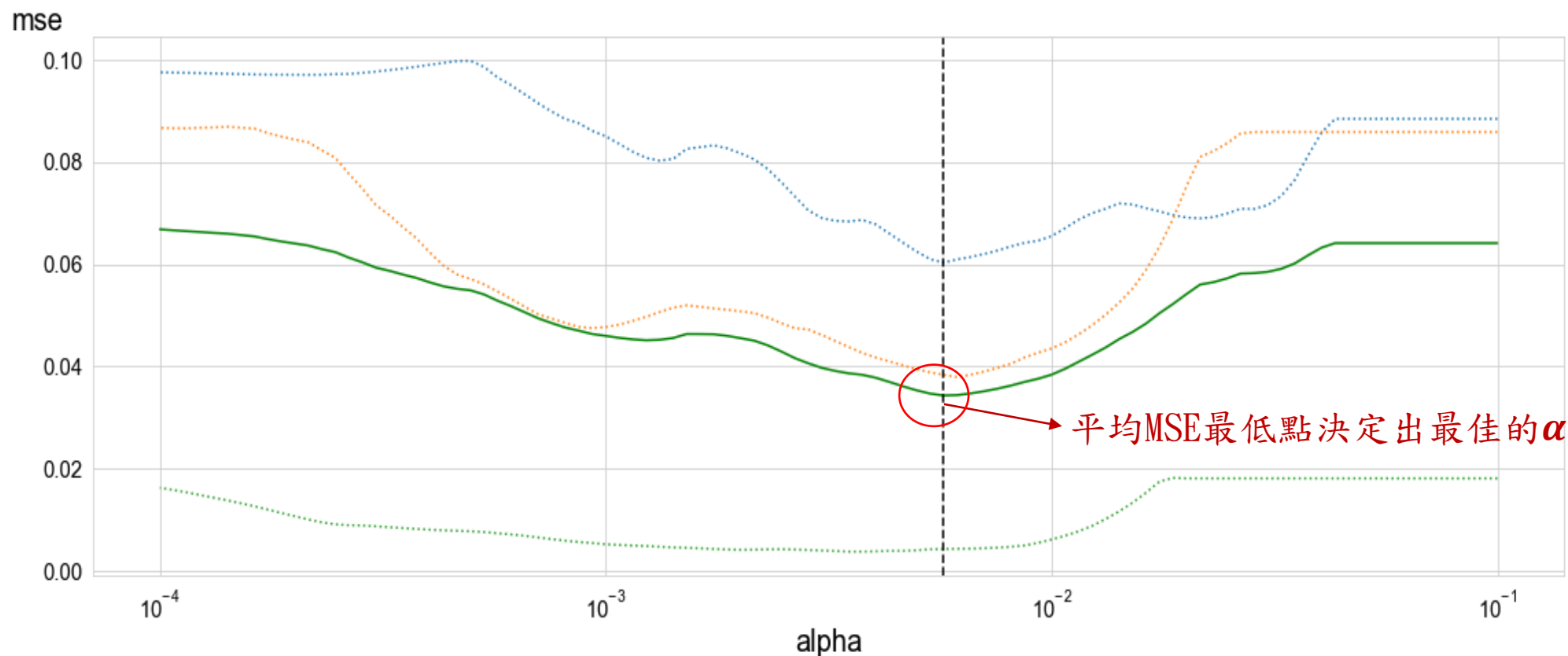
- $\alpha \uparrow \Rightarrow$ 懲罰項 $\uparrow \Rightarrow$ 係數收縮至0
- 如何決定最佳的 α ? \leftarrow 交叉驗證(Cross Validation)

交叉驗證

樣本(內)期間為1995Q1到2018Q4共96期，如圖分成三組樣本資料：



Lasso迴歸交叉驗證



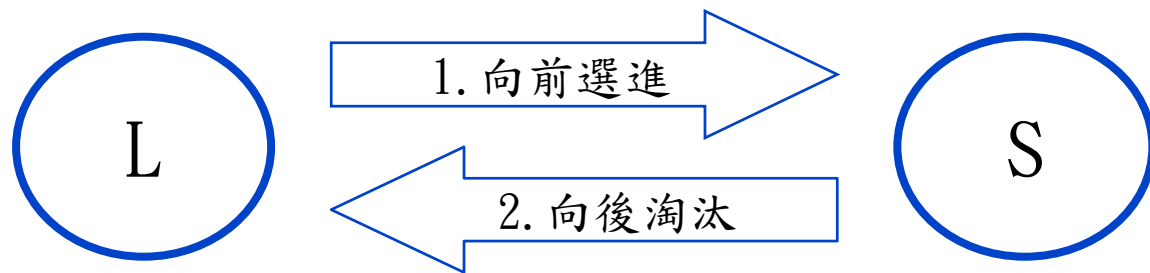
$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2$$

虛線：三組樣本在不同 α 水準的MSE表現

實線：平均MSE路徑

逐步迴歸法

- Lasso迴歸中非零係數的變數集合 $L = \{x_1, x_2, \dots, x_p\}$
 - 逐步迴歸法選進的變數集合為 $S = \{s_1, s_2, \dots, s_k\}$
1. 向前選進: Y 各別對 $S \cup x_i$ 迴歸, $i = 1, 2, \dots, p$, 這 p 個迴歸中, 如果最高的 R_a^2 比 Y 對 S 迴歸的高, 則選進該變數。
 2. 向後淘汰: 任一 s_i 邊際貢獻 = Y 對 S 迴歸的 $R_a^2 - Y$ 對 $S \setminus \{s_i\}$ 迴歸的 R_a^2 , 若最低貢獻的變數不是 s_k 時, 將該變數放回 L 集合。



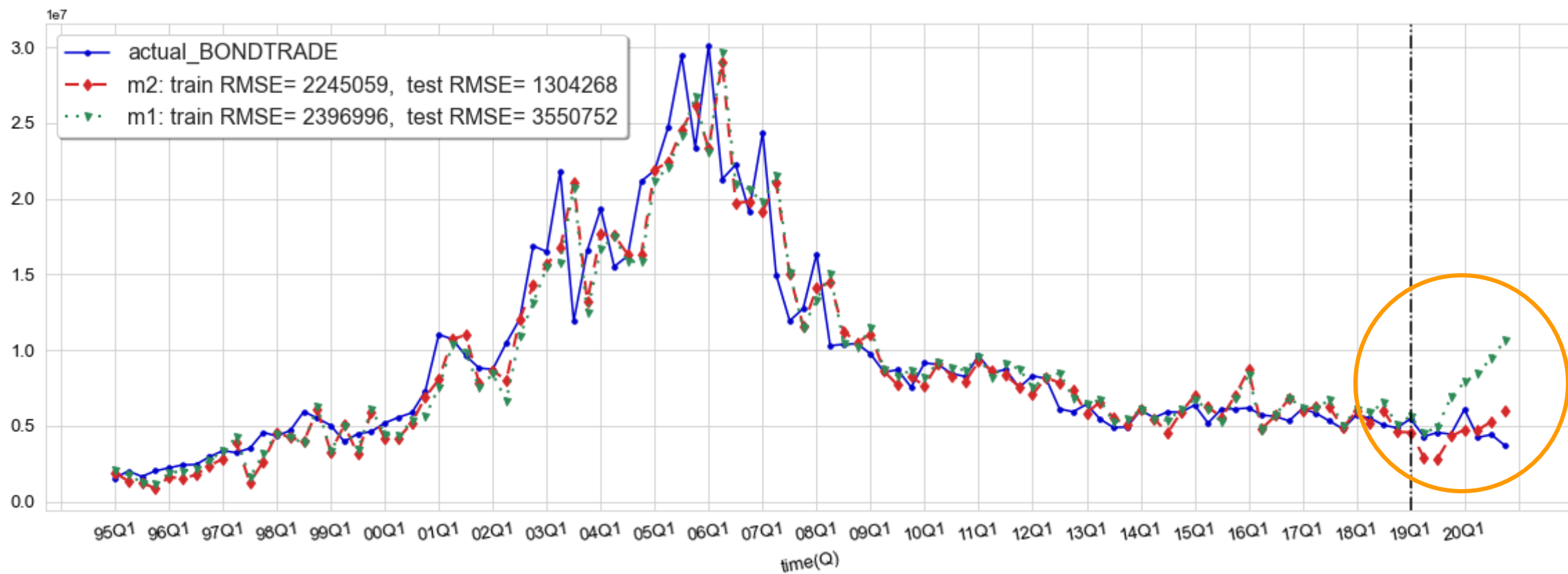
比較模型加入擴散指標前後的表現

- 樣本內期間: 1995Q1~2018Q4
樣本外期間: 2019Q1~2020Q4
- 兩模型設定差別在於 m2 模型有7條迴歸有選入擴散指標。那在同樣的34個內生變數中，有3個變數在樣本外預測的RMSE，在 m2 模型是下降20%以上。

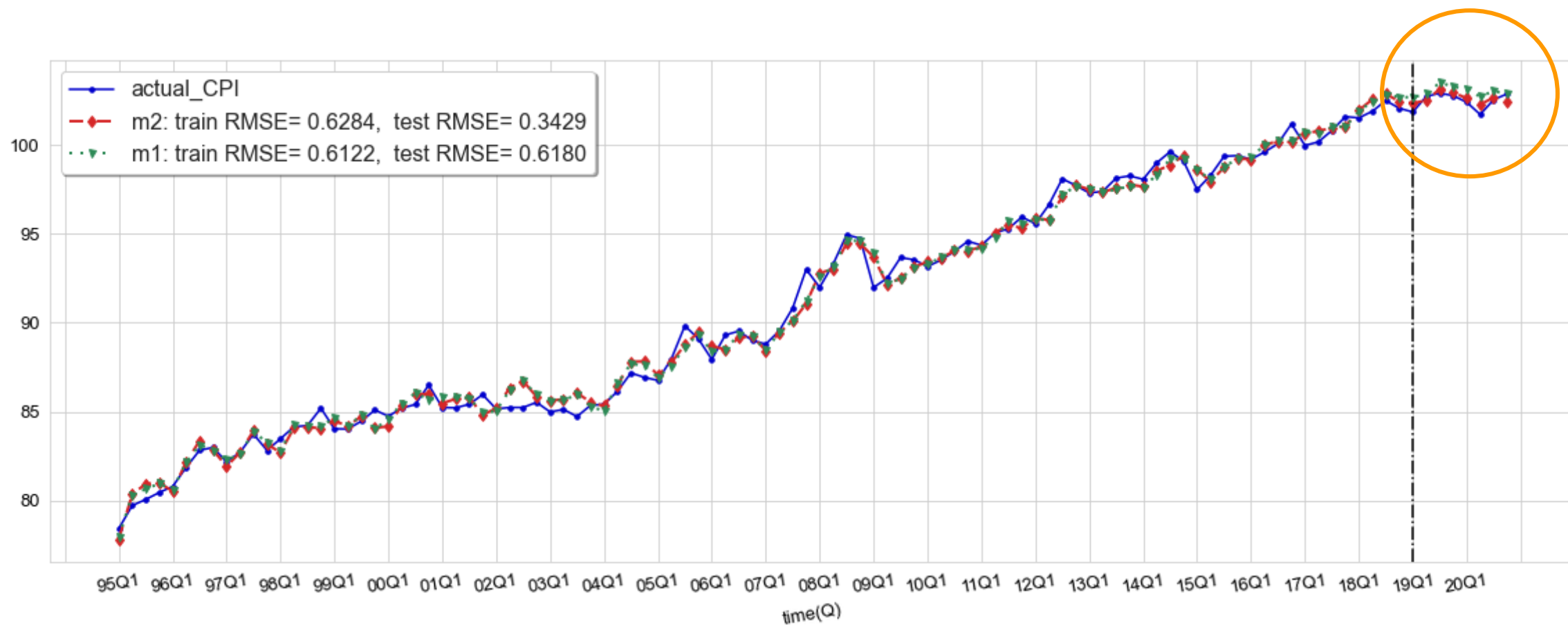
m2 相較於 m1 模型的 RMSE 比例

變數	樣本內預測	樣本外預測
	靜態求解	動態求解
債券市場總成交值 BONDTRADE*	0.937	0.367
消費者物價指數 CPI*	1.026	0.555
實質國內生產毛額 real GDP	0.961	0.769
製造業國內生產毛額 GDPMFG*	0.935	0.927
民間固定資本形成毛額 IBF*	0.861	0.945
國內生產毛額平減指數 PGDP*	0.974	0.990
全體貨幣機構放款 LOAN*	0.926	1.080
金融帳 FA*	1.173	1.293

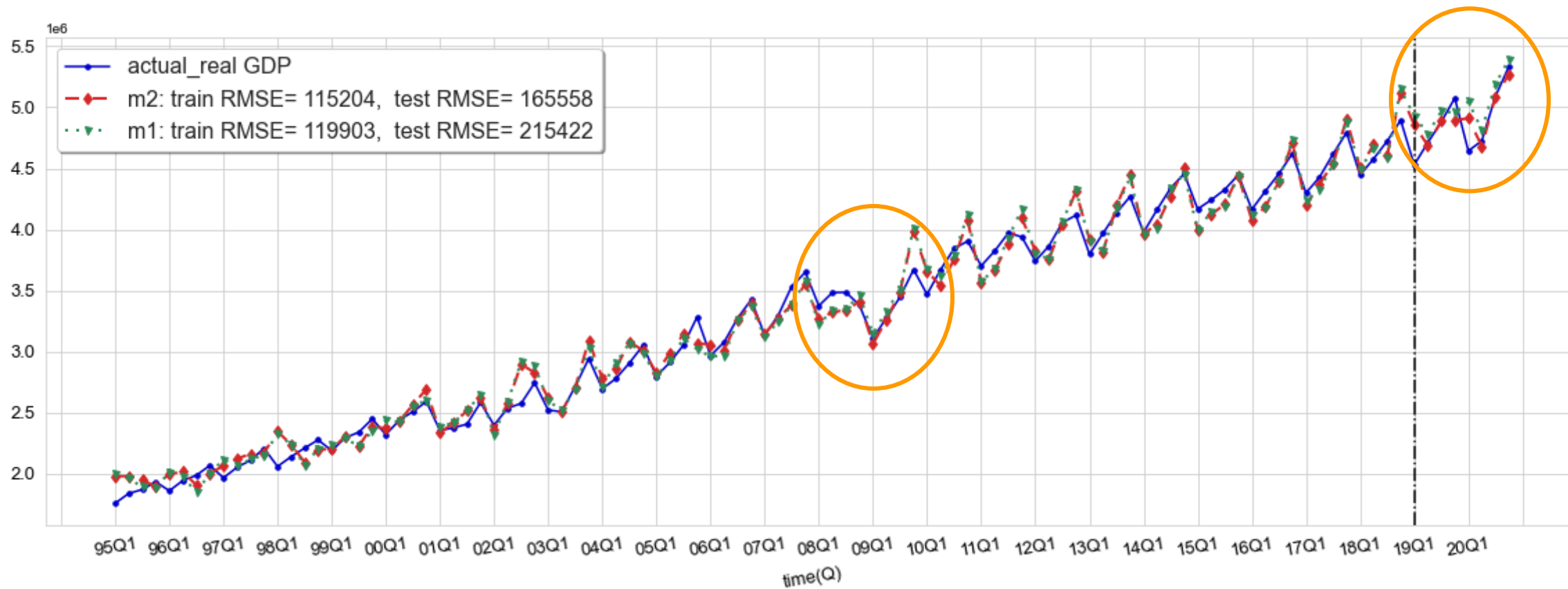
債券市場交易總成交值 $m1,m2$



消費者物價指數 $m1, m2$



實質國內生產毛額 $m1, m2$



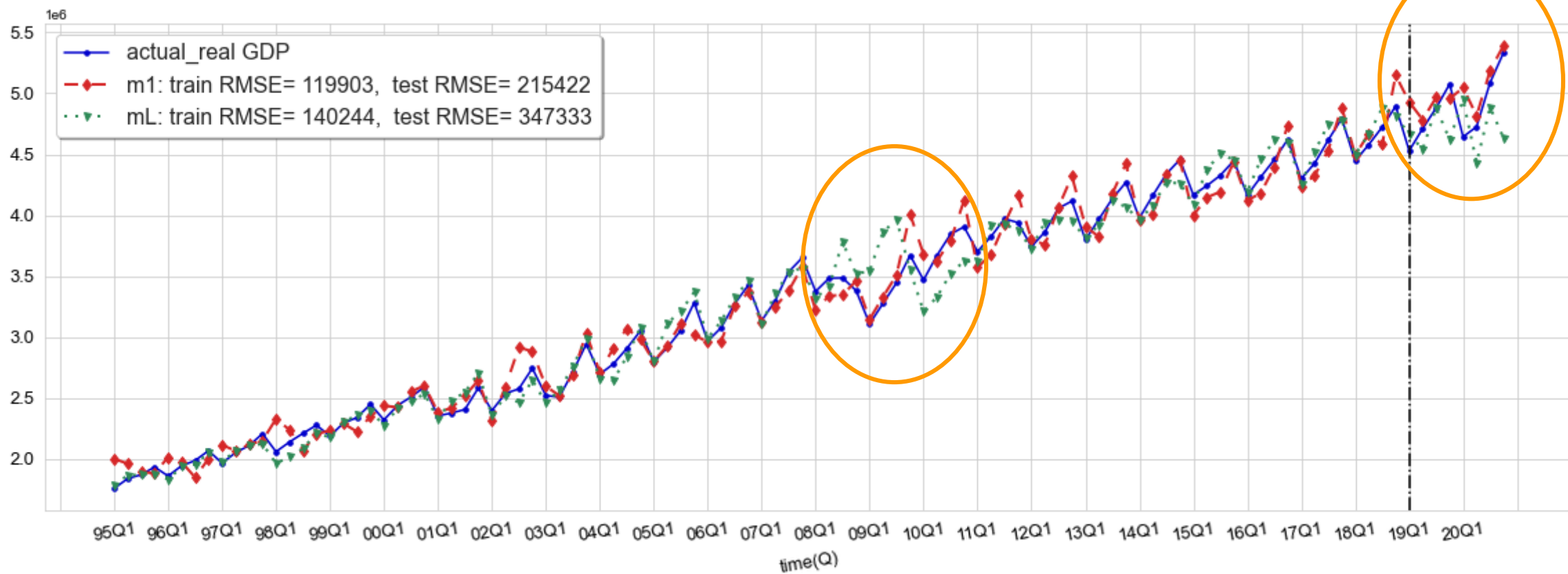
比較採用變數選取方法前後的表現

- mL 模型的迴歸變數組合，是參考林建甫 (2006) 在 2004Q4 以前樣本期間所設定，mL 共有 25 個內生變數，相比 m1 少設定的 9 個內生變數，是受限於無解釋變數的資料。
- 25個中有20個是 m1 模型樣本外預測較佳。

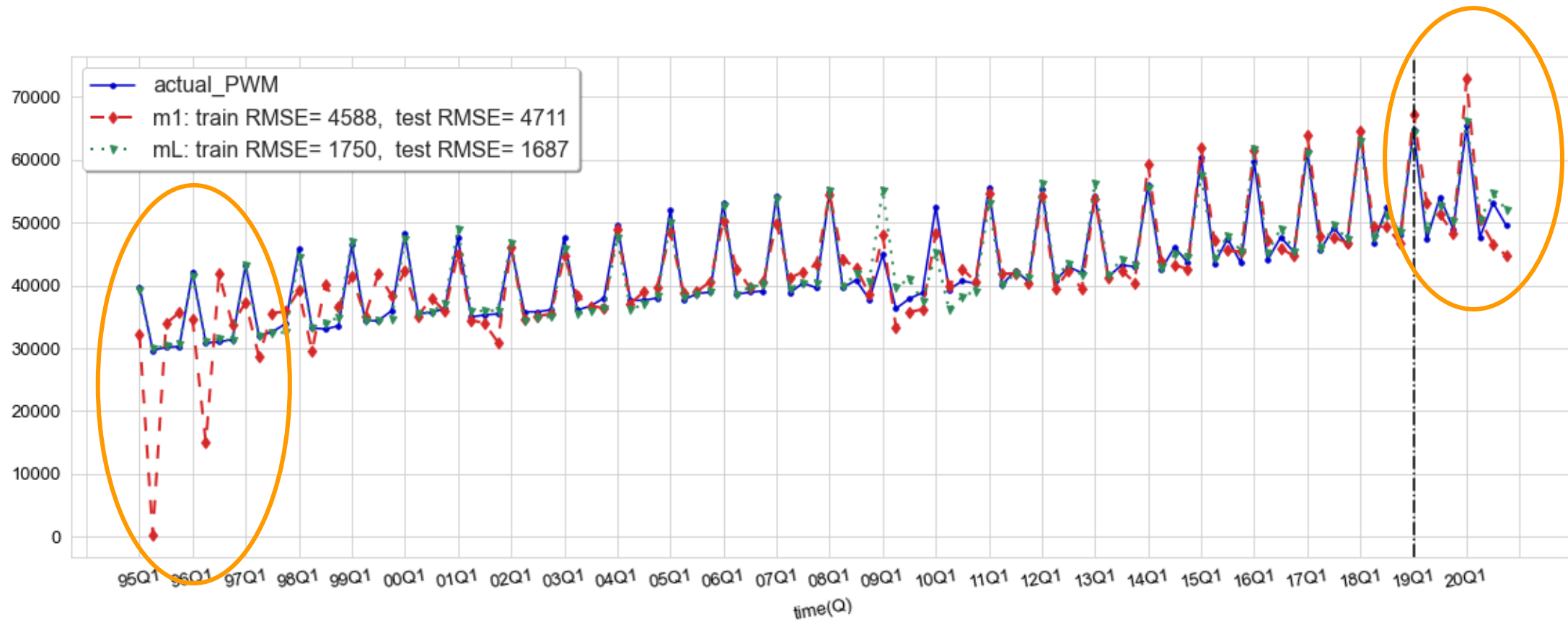
m1 相較於 mL 模型的 RMSE 比例

變數	樣本內預測	樣本外預測
	靜態求解	動態求解
核心消費者物價指數 CPIZF	0.702	0.091
固定資本消耗準備 DEP	0.100	0.126
貨幣總計數 M2	1.150	0.192
實質國內生產毛額 real GDP	0.855	0.620
經濟成長率 growth rate	0.636	0.978
勞動力人口 NF	1.235	1.381
民間固定資本形成毛額 IBF	1.213	1.417
製造業平均薪資 PWM	2.622	2.792

實質國內生產毛額 $m1, mL$



製造業平均薪資 $m1, mL$



結論

使用主成分分析、Lasso 迴歸與逐步迴歸法，簡化模型中迴歸解釋變數設定，當有更新的資料時，方便可再執行變數選擇與模型求解，以資料特徵幫助模型重新挑選變數。經比較模型預測表現顯示，採用變數選擇方法、加入擴散指標皆可以提升預測表現。

在比較模型中樣本外預測 (動態求解) 的表現，為避免外生變數給定預測值造成的誤差，影響到內生變數的預測值，因此在動態求解時，外生變數是代入真實值。

未來在追求預測表現的目標上，可再嘗試不同的交叉驗證方法，或是應用Midas 迴歸，較高頻率的月資料能即時更新經濟成長率的預測值。

吳俊毅與朱浩榜 (2020)，「即時預報台灣經濟成長率:MIDAS 模型之應用」

Thanks for Listening .