

以回歸模型印證高等教育的外部效益

鄭泊聲* 陳致宇† 倪楷恩‡ 白崇佑§

May 31, 2022

摘要

本研究的主旨是希望了解高等教育是否具有外部效益，以印證高等教育的補助政策是否使社會福利最大化。我們採取的方式是以一地區的平均薪資作為被解釋變數，並以該地區的就業人口中大專畢業生的比例或其他類似資料，作為主要的解釋變數。以台灣各縣市 2020 年的資料來進行回歸估計，在控制住其他各項縣市的特徵後，我們發現提高一地區的大專畢業生佔就業人口比例，確實會提升該地區的平均薪資。我們假設平均薪資的提升，代表不光是受高等教育者的薪資提升，甚至是其他未受高等教育的就業者，其薪資也都會因為高該地區的大專畢業生變多而提升，因此高等教育具有外部效益。除此之外，我們另取 2018-2020 的 Panel data 資料進行回歸估計，也發現類似的結果。最後，我們取整個台灣 (不分縣市) 的 1982-2021 時間序列資料，探討政府教育支出對家戶平均所得的影響，也得到正面的結果。

*b07611002 國立臺灣大學經濟學系

†b09505045 國立臺灣大學工程科學與海洋工程學系

‡b08208033 國立臺灣大學政治學系

§b07501026 國立臺灣大學土木工程學系

Abstract

The main topic of this work is to proof whether higher education has a spillover effect or not. We firstly used cross sectional data in 20 Taiwanese cities to estimate the effect of college worker share on average wage. We found it does has a positive significant effect. We then turn to panel data in 2018-2020 with approximately the same model where an even higher positive significant effect was found. Finally we used a time series dataset in Taiwan from 1982 to 2020 and also discovered similar results.

目錄

1	前言與研究目的	1
1.1	資料來源與參考文獻	1
2	2020 年的多變數回歸模型	2
2.1	以大專畢業生佔就業人口比例做為主要解釋變數	4
2.2	以各縣市人口中受高等教育者比例做為主要解釋變數	6
3	Instrumental Variable	9
4	2018-2020 Panel Data	9
4.1	一階差分法估計 (First Differencing Estimation)	13
4.2	隨機效果法估計 (Random Effect Estimation)	14
5	1982-2020 時間序列資料	18
5.1	一階差分資料	18
5.2	去時間趨勢估計 (De-time trend)	19
6	結論與比較	21

表格目錄

2.1	MLR on all variables (Hlavac 2022)	3
2.2	教育相關變數的聯合顯著性	4
2.3	以大專畢業生佔就業人口比例做為主要解釋變數的多變數回歸結果。	5
2.4	Heteroskedasticity Robust for college worker share	6
2.5	以各縣市 15 歲以上人口中受高等教育者比例作為主要解釋變數。	7
2.6	Heteroskedasticity Robust for higher edu share	8
3.1	IV first stage results.	10
3.2	IV 2SLS - 大專畢業生佔就業人口比例作為主要解釋變數。	11
3.3	IV 2SLS - 大專畢業生佔城市 15 歲以上人口比例作為主要解釋變數。	12
4.1	Panel data estimation - first differencing	13
4.2	Random effect estimation	15
4.3	表 4.2 的殘差估計。	16
4.4	Random effect estimation with IV - first stage.	16
4.5	Random effect estimation with IV - 2SLS.	17
5.1	時間序列資料估計結果。	20
5.2	時間序列資料的 serial correlation 檢驗。	21
5.3	De-time trend MLR	22
6.1	本研究中各項估計法的比較	23

圖片目錄

1	各縣市平均薪資對大專畢業生佔就業人口比例作圖。以 2020 年的資料繪成。	2
2	各縣市平均薪資對大專畢業生佔就業人口比例作圖。以 2018-2020 的資料繪成。	10
3	一階差分後的 panel data 資料	14

1 前言與研究目的

本文旨在探討我國高等教育是否具有外溢效應 (Spillover)，亦即政府在高等教育的補助是否能提升每個人的薪資，而非只有受高等教育者的薪資得到提升。從基本的供需模型中，可以得知針對外部效益的補助會最大化社會總剩餘，換句話說，高等教育的外部效益是政府補助在經濟學上的正當性來源。實務上，受到資料可及性的限制，我們希望量測高等教育對於就業者平均薪資的影響。在此我們做了一個武斷的假設：平均薪資的提升意味著所有就業者的薪資都得到提升，這意味著，若我們透過迴歸分析，發現高等教育的提升也會造成平均薪資的提升，那麼我們就能確認高等教育確實存在外部效益，並證實政府對高等教育的補助確實有基本的理論基礎。至於平均薪資的提升中，有多少部分是受高等教育者的薪資提升 (也就是高等教育的內部性)，又有多少是外溢效益的部分，需要就業市場中的個體資料，這是本研究現階段無法企及的。

在資料上，本研究的出發點是以全台灣 20 個縣市 (排除金門、馬祖) 的資料做橫斷面分析。我們收集到各縣市的就業者平均年薪資，並以此作為各個模型中主要的被解釋變數。接著，我們思考的是要以哪樣的資料作為衡量「高等教育」的變數，我們收集到各縣市就業人口中大學畢業者的比例、科學教育支出佔歲出比例以及 15 歲以上人口受大專以上教育之比例。

我們先簡單跑了一個各縣市就業人口中大學畢業者的比例與就業者平均年薪的單變數迴歸式，如圖1，可以發現它可能存在線性關係。但迴歸模型的實際衡量是另一回事，畢竟若要探討不同縣市中教育與薪資的關係，則勢必需要加入更多城市的特徵作為變數，以避免 OLS 估計的遺漏變數偏誤 (Omitted Variable Bias)，因此本文的首要任務便是找出顯著的控制變數，後續的篇幅則旨在使用工具變數、panel data 等和時間序列等各式迴歸模型與資料去印證高等教育與與薪資之間的關係。

1.1 資料來源與參考文獻

本研究中主要的被解釋變數，也就是各縣市的就業者平均薪資，其資料來自主計總處 (行政院主計總處 2022)，代表高等教育的三個解釋變數則來自行政院主計總處縣市重要統計指標查詢系統 (中華民國統計資訊網 2022a)，其餘的城市特徵資料也來自該系統。另有一些資料來自內政部戶政司人口統計資料 (內政部戶政司全球資訊網 2022)、中華民國統計資訊網的物價指數與家庭收支調查 (中華民國統計資訊網 2022b、中華民國統計資訊網 2022c)。

本文中多數的模型設定皆參考自 (Moretti 2004)。

2020 City Data

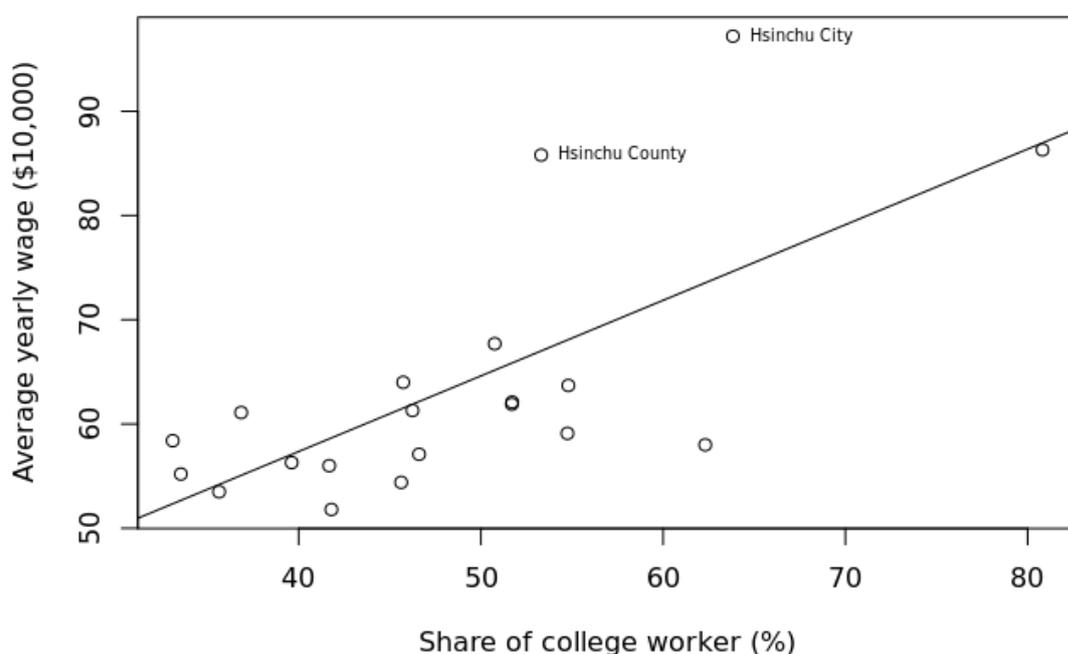


圖 1: 各縣市平均薪資對大專畢業生佔就業人口比例作圖。以 2020 年的資料繪成。

2 2020 年的多變數回歸模型

我們首先希望找到資料中具有足夠解釋力的變數作為模型中的控制變數，模型的設定如下：

$$wage2020 = \beta_0 + workforceCollege_2020 + \beta X + u \quad (2.1)$$

其中 $wage2020$ 即為被解釋變數，各縣市的就業者平均年薪資， $workforceCollege_2020$ 即為主要的解釋變數，就業人口中大專畢業生的比例， X 則包含其他各城市的特徵，包括：直轄市的虛擬變數 ($direct2020$)、受僱者佔勞動人口比例 ($hired2020$)、工業產值佔縣市生產總額比例 ($manufacture2020$)、服務業產值佔縣市生產總額比例 ($service2020$)、性別比 ($gender2020$)、科學教育支出佔歲出比例 ($eduexpense2020$)、15 歲以上人口受大專以上教育之比例 ($edulevel2020$)、結婚率 ($married2020$)、人均消費支出 ($expensepercapita2020$)、失業率 ($unemployment2020$)、直轄縣市的虛擬變數及大學畢業者佔就業人口比例的交互作用項 ($directedu2020$)。研究對象為台灣 20 個縣市（不納入金門、馬祖），共有 20 個樣本數。表 2.1 為我們多重線性迴歸的結果，我們將其中比較顯著的變數選出來，在後面的 2 個多重線性迴歸模型中當控制變數。

表 2.1: MLR on all variables (Hlavac 2022)

		<i>Dependent variable:</i>
		wage2020
workforceCollege_2020	-0.818 (1.814)	
direct	81.297** (23.442)	
hired2020	1.486** (0.608)	
manufacture2020	-2.045** (0.850)	
service2020	-1.049 (0.915)	
gender2020	2.499** (0.994)	
eduExpense2020	1.162* (0.572)	
eduLevel2020	1.866 (1.746)	
married2020	1.504 (1.141)	
expensePerCapita2020	0.003*** (0.001)	
unemployment2020	-31.468 (21.214)	
directEdu2020	-1.531*** (0.416)	
Constant	-268.635 (171.289)	
Observations	20	
R ²	0.959	
Adjusted R ²	0.888	
Residual Std. Error	4.055 (df = 7)	
F Statistic	13.604*** (df = 12; 7)	
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

在該模型中，我們發現代表高等教育的三個變數 *workforceCollege_2020*、*edulevel2020*、*eduexpense2020* 皆未能估計出顯著的係數，所以我們將此三個變數與被解釋變數，做多變數 OLS 估計，根據表2.2中的結果，這三個變數具有聯合顯著性。隨後，我們分別以這三個變數作為主要的解釋變數，與表2.1中較具解釋力的變數組成多變數迴歸模型。我們發現以「大專畢業生佔就業人口比例 (*workforceCollege_2020*)」與「各縣市 15 歲以上人口中受高等教育者比例 (*edulevel2020*)」作為主要的解釋變數，皆能得出顯著且正的 OLS 係數，且係數大小相當接近，其結果分別呈現於後面兩個小節中。

表 2.2: 教育相關變數的聯合顯著性

<i>Dependent variable:</i>	
wage2020	
<i>workforceCollege_2020</i>	0.946 (1.231)
<i>eduExpense2020</i>	0.650 (0.530)
<i>eduLevel2020</i>	-0.333 (1.259)
Constant	9.424 (18.926)
Observations	20
R ²	0.525
Adjusted R ²	0.436
Residual Std. Error	9.118 (df = 16)
F Statistic	5.889*** (df = 3; 16)
Note:	*p<0.1; **p<0.05; ***p<0.01

2.1 以大專畢業生佔就業人口比例做為主要解釋變數

在此，我們使用的模型如下：

$$\begin{aligned}
 \text{wage2020} = & \beta_0 + \beta_1 \text{workforceCollege_2020} + \delta_0 \text{direct} \\
 & + \beta_2 \text{wage2018} + \beta_3 \text{manufacture2020} + \beta_4 \text{hired2020} \quad (2.2)
 \end{aligned}$$

此處的高等教育變數,放的是各縣市大專畢業生佔就業人口比例 (*workforceCollege_2020*), 研究對象為台灣 20 個縣市(不納入金門及馬祖), 資料期間同樣為 2020 年, 共有 20 個樣本數, 其他控制變數包含了直轄市的虛擬變數 (*direct*)、工業佔 2020 年城市生產總額的比例 (*manufacture2020*)、受僱者佔整體勞動人口比例 (*hired2020*)。我們以落後兩期的被解釋變數 (*wage2018*) 作為代理變數, 來代表其他未加入的城市特徵。我們希望盡可能避免代理變數與主要解釋變數有內生性的問題, 因此選用落後兩年的被解釋變數。

根據模型估計結果發現, 我們主要的解釋變數大學生佔勞動人口比例, 確實是影響台灣各縣市平均薪資的決定因素。迴歸 F 值 =1175.673, p 值 <0.01, 達顯著水準, R² 為 0.998, 調整後 R 為 0.997, 此迴歸式有解釋能力。

表 2.3: 以大專畢業生佔就業人口比例做為主要解釋變數的多變數回歸結果。

Dependent variable:	
	wage2020
workforceCollege_2020	0.075** (0.029)
direct	-0.194 (0.413)
wage2018	1.017*** (0.021)
manufacture2020	0.025 (0.022)
hired2020	-0.032 (0.038)
Constant	-1.259 (2.005)
Observations	20
R ²	0.998
Adjusted R ²	0.997
Residual Std. Error	0.689 (df = 14)
F Statistic	1,175.873*** (df = 5; 14)
Note:	* p<0.1; ** p<0.05; *** p<0.01

2.1.1 變異數異質性

在使用橫斷面資料時，時常會遇到異質變異的問題，若使用具有異質變異性的模型加以估計，會導致錯誤的結果。故透過 Breusch-Pagan(BP) 統計值對此問題予以檢定。在本處算出模型 BP 值為 10.854, p 值為 0.05435，並不顯著，代表觀察值並不具異質變異之特性。我們仍然在表2.4中呈現 robust 的估計結果，並不影響主要解釋變數的顯著性。

表 2.4: Heteroskedasticity Robust for college worker share

<i>Dependent variable:</i>	
	wage2020
workforceCollege_2020	0.080*** (0.013)
direct	0.020 (0.240)
wage2018	0.997*** (0.011)
manufacture2020	0.009 (0.022)
hired2020	-0.034 (0.042)
Constant	0.290 (1.613)
Observations	20
R ²	0.998
Adjusted R ²	0.998
Residual Std. Error	0.474 (df = 14)

Note: *p<0.1; **p<0.05; ***p<0.01

2.2 以各縣市人口中受高等教育者比例做為主要解釋變數

第二個模型的高等教育變數，放的是各縣市 15 歲以上人口中受高等教育者比例 (*edulevel2020*)，模型的其餘設定都相同。根據模型估計結果發現，影響台灣各縣市

平均薪資的決定因素中, 顯著的因素有我們主要的解釋變數縣市接受高等教育者比例 (*eduLevel_2020*)。另外, 我們選用的代理變數 (落後 2 期的各縣市平均薪資 *wage2018*) 在兩個模型中也都有顯著的正效果。迴歸 F 值 =1220.078, p 值 <0.01, 達顯著水準, R² 為 0.998, 調整後 R 為 0.997, 此迴歸式有解釋能力。

表 2.5: 以各縣市 15 歲以上人口中受高等教育者比例作為主要解釋變數。

<i>Dependent variable:</i>	
	wage2020
direct	-0.322 (0.421)
manufacture2020	0.022 (0.020)
eduLevel2020	0.071** (0.026)
hired2020	-0.012 (0.033)
wage2018	1.015*** (0.021)
Constant	-1.967 (1.876)
Observations	20
R ²	0.998
Adjusted R ²	0.997
Residual Std. Error	0.677 (df = 14)
F Statistic	1,220.078*** (df = 5; 14)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

我們一樣透過 Breusch-Pagan(BP) 統計值檢定變異數異質性。在本模型 BP 值為 11.97, p 值為 0.0352, 結果顯著, 代表觀察值具異質變異之特性。不過從表 2.6 中可以看出, robust 的結果亦不會有太大差異。

表 2.6: Heteroskedasticity Robust for higher edu share

<i>Dependent variable:</i>	
	wage2020
direct	-0.107 (0.213)
manufacture2020	0.008 (0.021)
eduLevel2020	0.080*** (0.015)
hired2020	-0.017 (0.041)
wage2018	0.992*** (0.009)
Constant	-0.248 (1.673)
Observations	20
R ²	0.999
Adjusted R ²	0.998
Residual Std. Error	0.372 (df = 14)
<i>Note:</i>	* p<0.1; ** p<0.05; *** p<0.01

3 Instrumental Variable

在本節中，我們希望在前述的當期多變數回歸模型中再加入 IV，以試圖進一步減少估計的結果中可能有的遺漏變數偏誤。我們參考 Moretti 2004 中的做法，以每個城市的人口年齡結構作為 IV，在此先說明我們挑選這個變數作為 IV 的理由。

首先，我們知道一個好的 IV，應該要對被解釋變數沒有解釋力，但與模型中主要的解釋變數有相關性。更明確的說，我們選擇的 IV， z 應該符合以下條件：

$$\text{cov}(u, z) = 0 \quad (3.1)$$

我們認為一個城市的年齡結構不應該跟該地的薪資有明顯的相關性，因為年齡對薪資的影響，顯然會因不同的產業或其他因素而異。同時， z 還必須符合

$$\text{cov}(x, z) \neq 0 \quad (3.2)$$

由於台灣的人口中大學畢業生的比例越來越高，因此越年輕的人口族群中大學生的比例應該會越來越高，而年輕人較多的城市，大學畢業生的畢業就也應該較多。因此城市的年齡結構很可能跟我們的主要解釋變數，也就是大專畢業生佔就業人口的比例，有相關性存在。我們以 2010 年時每個城市中 15-24 歲人口的比例 (*workforceYoung_2010*)，作為模型中的 IV。選用落後十年的資料的原因是，我們希望盡可能提高讓式 3.1 成立的機率。一個城市該年度的薪資或其他模型中的特徵，要與十年前的年齡結構有相關性，應是不太可能。

表 3.1 中呈現出此 IV 的 First Stage Regression 結果。以我們挑選的 IV 對模型的主要解釋變數回歸，得到 -2.199 的 t-statistics，並無法滿足一般對於 IV 的要求 (t-statistics > 3.3)。因此可以看見，表 3.2、3.3 中的 2SLS 係數在統計上皆不顯著，不過值得提出的是，模型加上 IV 後，主要解釋變數的係數有明顯的提升 (0.076 versus 0.098)，或許還是可以對原本模型中的偏誤方向提供參考。

4 2018-2020 Panel Data

以上的當期多變數回歸分析，皆以台灣的 20 個縣市作為資料點，雖然模型中的主要解釋變數皆呈統計上顯著，我們還是希望能建立另一個資料數量更多的模型。因此我們將同樣的模型加入 2019 與 2018 年的資料，形成一個 panel data 模型。資料年份的限制主要來自於各縣市的平均薪資統計僅有 2018-2020 的資料。

由於我們總共只有三期的資料，因此我們將原本模型中的被解釋變數落後項移除，否則就必須放棄一期的資料。我們將圖 1 加上新增的兩期的資料點後，重新繪製於圖 2。

表 3.1: IV first stage results.

<i>Dependent variable:</i>	
workforceCollege_2020	
workforceYoung_2010	-5.845** (2.658)
Constant	91.140*** (19.525)
Observations	20
R ²	0.212
Adjusted R ²	0.168
Residual Std. Error	10.578 (df = 18)
F Statistic	4.835** (df = 1; 18)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

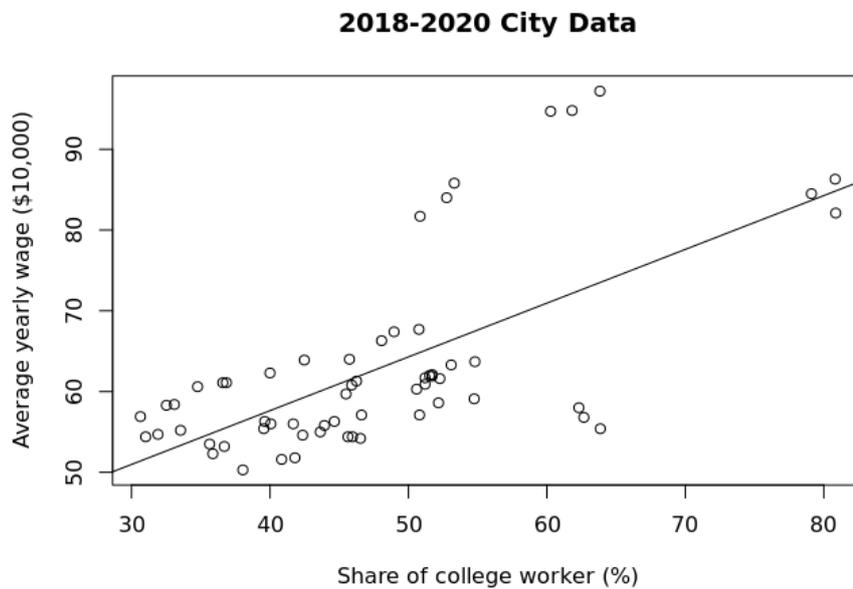


圖 2: 各縣市平均薪資對大專畢業生佔就業人口比例作圖。以 2018-2020 的資料繪成。

表 3.2: IV 2SLS - 大專畢業生佔就業人口比例作為主要解釋變數。

<i>Dependent variable:</i>	
wage2020	
workforceCollege_2020	0.098 (0.103)
direct	-0.308 (0.650)
wage2018	1.007*** (0.049)
manufacture2020	0.035 (0.048)
hired2020	-0.050 (0.087)
Constant	-0.643 (3.368)
Observations	20
R ²	0.998
Adjusted R ²	0.997
Residual Std. Error	0.704 (df = 14)
Wald test	1125 on 5 and 14 DF, p-value: < 2.2e-16

表 3.3: IV 2SLS - 大專畢業生佔城市 15 歲以上人口比例作為主要解釋變數。

<i>Dependent variable:</i>	
wage2020	
eduLevel2020	0.097 (0.101)
direct	-0.505 (0.820)
wage2018	1.002*** (0.054)
manufacture2020	0.033 (0.045)
hired2020	-0.026 (0.063)
Constant	-1.483 (2.668)
Observations	20
R ²	0.998
Adjusted R ²	0.997
Residual Std. Error	0.700 (df = 14)
Wald test	1138 on 5 and 14 DF, p-value: < 2.2e-16

4.1 一階差分法估計 (First Differencing Estimation)

我們首先以一階差分的方式消除 panel data 中所謂的固定效果 (Fixed Effect, or Unobserved Effect)，再進行一般的 OLS 估計。其結果如表4.1中所示，主要解釋變數-大專畢業生佔就業人口比例 (*workforceCollegeDiff*)-的係數並不顯著，該模型幾乎無法得到任何有用的估計結果。我們認為這是因為每個縣市的大專畢業生佔就業人口比例，在資料收集的這三年之間並沒有明顯的變化，在資料變異不足的情形下， SST_x 非常小時， β_x 的標準差就會很大¹，並導致不顯著的估計結果。我們將一階差分後的資料繪至於圖3中，可以看見原本圖2中的線性分布趨勢已完全消失。

表 4.1: Panal data estimation - first differencing

Dependent variable:	
	wageDiff
workforceCollegeDiff	-0.109 (0.086)
direct2	0.341 (0.283)
serviceDiff	0.358*** (0.103)
manufactDiff	-0.028 (0.133)
hiredDiff	-0.092 (0.120)
Constant	0.671*** (0.141)
Observations	40
R ²	0.279
Adjusted R ²	0.173
Residual Std. Error	0.601 (df = 34)

Note: *p<0.1; **p<0.05; ***p<0.01

¹ β_x 的標準差公式：
$$\frac{\hat{\sigma}}{\sqrt{SST_x(1-R^2)}}$$

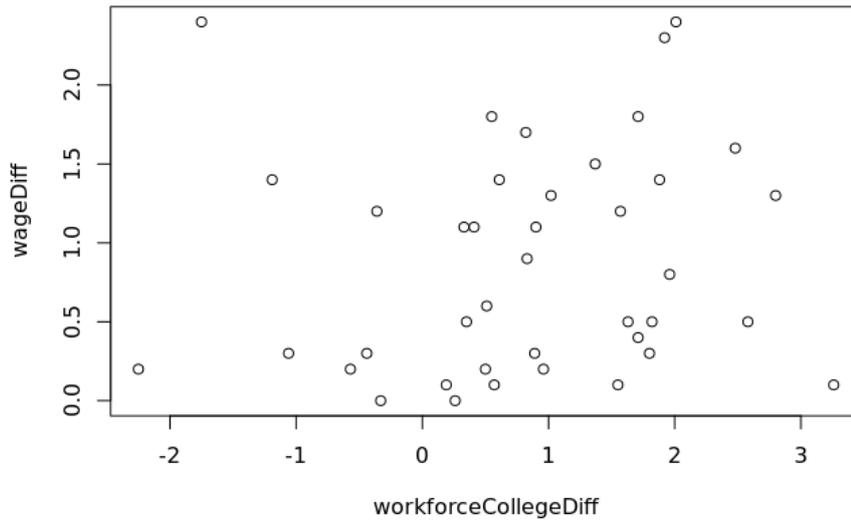


圖 3: 一階差分後的 panel data 資料

4.2 隨機效果法估計 (Random Effect Estimation)

在一階差分法無法得到有效的結果後，我們改以隨機效果法 (Random Effect) 估計。隨機效果法中，我們必須假設模型的固定效果與每一期的所有解釋變數皆為獨立，並以 quasi-demeaned data

$$y_{it} - \hat{\theta}\bar{y}, \quad x_{it} - \hat{\theta}\bar{x} \quad (4.1)$$

的方式取代原本的變數進行估計。當係數 $\hat{\theta} = 1$ 時，這個模型與固定效果法估計 (Fixed Effect Estimation) 相同。

我們以隨機效果法得到的估計結果，如表4.2所示，可以看見主要解釋變數的係數為統計上顯著的結果，且其數值大小與當期多變數回歸的模型 (表2.5) 相比，又提升了許多 (0.076 versus 0.403)。值得說明的是，我們在本模型中放入另一個變數: 每個縣市的年度人均支出 (*expensePerCapita2018*) 來增加控制變因。² 可以看見此變數的係數呈統計上顯著，但數值極小。

我們接著在這個 panel data 模型中也加入節3中的 IV，不過得出的結果與節3中並沒有不同，IV 本身對主要解釋變數的解釋力不夠顯著 (表4.4)，得出的 2SLS 係數雖然又近一步變高，但依然不顯著。(表4.5)

在以上的 panel data 模型中，我們成功地建立起一個資料點更多的模型，同樣印證了提升一地區的大專畢業生佔就業人口比例，可以提高該地的平均薪資，並且得出比當期多變數回歸模型更大的係數。除了上述的兩種估計方式外，我

²事實上，我們曾經在其他模型中也放入這個變數，但較果皆不佳，會導致我們的主要解釋變數失去顯著性。

表 4.2: Random effect estimation

<i>Dependent variable:</i>	
	wage2018
workforceCollege_2018	0.403*** (0.114)
manufacture2018	0.122 (0.127)
hired2018	-0.010 (0.169)
direct	-8.666 (10.479)
directEdu2018	0.098 (0.187)
expensePerCapita2018	0.001*** (0.0002)
Constant	24.754** (10.832)
Observations	60
R ²	0.539
Adjusted R ²	0.487
F Statistic	62.031***

Note: * p<0.1; ** p<0.05; *** p<0.01

	var	std.dev	share
idiosyncratic	0.5627	0.7501	0.015
individual	37.9300	6.1587	0.985

$$\hat{\theta} = 0.9299$$

表 4.3: 表4.2的殘差估計。

表 4.4: Random effect estimation with IV - first stage.

<i>Dependent variable:</i>	
workforceCollege_2018	
workforceYoung_2013	0.082 (0.584)
Constant	47.100*** (4.891)
Observations	60
R ²	0.0003
Adjusted R ²	-0.017
F Statistic	0.020

Note: *p<0.1; **p<0.05; ***p<0.01

們也以固定效果法進行了估計，發現與隨機效果的估計結果幾乎相同，這也可以從表4.3中接近 1 的 $\hat{\theta}$ 看出。我們也對該模型進行了獨立池化 (Independent Pooling) 的估計，但沒有得出合理的結果。³

雖然我們的隨機效果估計結果看似成功，但該模型有幾處弱點需要特別說明。首先，隨機效果法估計出的模型 R^2 ，只有 0.487，相比橫斷面多變數模型接近 1 的 R^2 有不少進步空間。其次，我們所使用的 panel data 估計方法其實都需要嚴格外生性 (Strict Exogeneity) 的假設成立，但我們無法確認或驗證該假設在我們的模型中是否成立。最後，隨機效果法要求模型中固定效果與每一期的所有解釋變數都獨立，這個假設我們也無法驗證。不過由於我們的模型在固定效果法下也得出類似的估計結果，而固定效果法不須這個假設，因此我們認為這點對估計結果的影響應該不大。

³僅得出工業佔生產總額有 0.4280 的顯著係數。

表 4.5: Random effect estimation with IV - 2SLS.

	<i>Dependent variable:</i>
	wage2018
workforceCollege_2018	1.776 (7.636)
manufacture2018	0.585 (2.097)
hired2018	-0.850 (4.705)
direct	35.758 (264.171)
directEdu2018	-0.853 (5.534)
expensePerCapita2018	0.0003 (0.003)
Constant	21.315 (22.243)
Observations	60
R ²	0.352
Adjusted R ²	0.279
F Statistic	17.148***

Note: *p<0.1; **p<0.05; ***p<0.01

5 1982-2020 時間序列資料

在經過以上的分析後，我們決定採取全台灣不分縣市的 1982-2020 年時間序列資料，對同樣的命題進行探討。基於資料收集上的限制，我們以台灣的家戶平均年所得 ($income_t$) 取代平均年薪資作為被解釋變數，而主要的解釋變數則改用政府的教育支出 ($edufund_t$)。由於我們發現家戶平均年所得與其落後一期的相關係數高達 0.9975，很有可能存在 highly persistent 問題，直接進行 OLS 估計會有 inconsistent 的風險。因此我們將所有資料皆進行一階差分

$$cincome_t = income_t - income_{t-1} \quad (5.1)$$

後才開始進行回歸估計。以下，所有開頭為 c 的變數名稱皆代表一階差分後的資料。

5.1 一階差分資料

我們以一階差分資料，建立了以下三個類似的模型：

$$cincome = cincome_lag + cedufund + cunem + cindpd + cavgGDP \quad (5.2)$$

$$cincome = cincome_lag + cedufund + cunem + cindpd \\ + cindpd_lag + cavgGDP \quad (5.3)$$

$$cincome = cincome_lag + cedufund + cunem + cservpd + cavgGDP \quad (5.4)$$

其中，

- $cincome_lag$ 是被解釋變數，家戶平均年所得的兩期落後項。由於我們擔心主要的解釋變數 $cedufund$ 與被解釋變數的落後項存在內生性，因此加入此變數。落後兩期是我們經過測試後最具解釋力的落後期數。
- $cedufund$ 是主要的解釋變數，政府教育支出。
- $cunem$ 是失業率。
- $cservpd$ 是服務業的生產總額，以新台幣百萬為單位。
- $cindpd$ 是工業的生產總額，以新台幣百萬為單位。

- $cindpd_lag$ 是 $cindpd$ 的一期落後項。
- $cavgGDP$ 是人均 GDP。

在以上三個模型中，主要解釋變數 $cedufund$ 的係數都呈顯著，且數值大小亦沒有明顯差異。模型5.2、5.3、5.4的結果分別呈現於表5.1的第一、二、三欄中。另外，這三個時間序列模型的 R^2 皆為 0.7 左右。為了確保該這些時間序列模型不會受到 serial correlation 的影響，我們將三個模型的估計殘差項 (residual, u_t) 分別對其殘差的落後一期項 (u_{t-1}) 進行 OLS 估計，發現期係數皆為不顯著 (表5.2)，因此在 contemporaneous exogeneity 的假設下，我們相信這些時間序列模型皆不受 serial correlation 影響。此外，這三個模型的 BP Test 結果分別為 p-value = 0.3591、p-value = 0.4967、p-value = 0.07205 因此在 95% 信心水準下皆無法拒絕變異數同質性的虛無假設。故我們的三個時間序列模型皆不需使用 HAC robust 估計。不過由於模型5.4的 BP Test 在 10% 信心水準下已呈顯著，我們仍然將其 HAC Robust 結果呈現於表5.1中第四欄。

5.2 去時間趨勢估計 (De-time trend)

除了前述的一階差分法估計外，我們也想確認時間序列模型是否受到時間趨勢 (time trend) 的影響。我們首先估計以下模型:

$$income = edufund + unem + indpd + avgGDP + t \quad (5.5)$$

並發現這個模型的解釋力很強，有極高的 R^2 。(表5.3)

接著，我們想要確認這個模型是否可能受到 highly persistent 的影響，因此，首先我們必須得到每個變數的去時間趨勢資料 (de-time trend data)。我們透過以下的迴歸

$$y_t = \alpha_0 + \alpha_1 t + e_t \quad (5.6)$$

得到變數 y_t 的去時間趨勢資料 $\ddot{y}_t = e_t$ 。

接著我們發現去時間趨勢的被解釋變數 \ddot{income}_t 與他的落後一期項 \ddot{income}_{t-1} 相關係數達 0.9570，很有可能存在 highly persistent 問題。因此我們再對所有已經去時間趨勢的資料進行一階差分

$$cincome_dt_t = \ddot{income}_t - \ddot{income}_{t-1} \quad (5.7)$$

並且在加入被解釋變數的兩期落後項後，重新一次5.5的迴歸，不過由於此時所有的變數都已經為去時間趨勢資料，因此模型中的變數 t 必須刪除。該迴歸估計的結果，與模型5.2的結果完全相同，僅在截距項有出入 (173 versus 206)。這是因為進行一階差分後， $t - (t - 1) = 1$ ，形同進行過了 de-time trend，因此除截距外其他係數的估計結果皆會相同。

表 5.1: 時間序列資料估計結果。

	<i>Dependent variable:</i>			
	cincome[3:38]			
		<i>OLS</i>		<i>MM-type linear</i>
	模型5.2	模型5.3	模型5.4	模型5.4HAC
cincome	0.555*** (0.121)	0.553*** (0.112)	0.729*** (0.126)	0.766*** (0.109)
cedufund_lag	0.0001** (0.00003)	0.0001*** (0.00003)	0.0001** (0.00003)	0.0001*** (0.00002)
cunem	-8,050.278*** (1,597.252)	-9,850.055*** (1,643.410)	-9,225.299*** (1,609.773)	-9,282.340*** (1,438.493)
cindpd	-0.001 (0.001)	-0.001* (0.001)		
cindpd_lag		-0.001** (0.001)		
cservpd			-0.009** (0.004)	-0.009 (0.006)
cavgGDP	0.149** (0.061)	0.077 (0.063)	0.201*** (0.063)	0.200** (0.095)
Constant	178.333 (1,437.168)	2,159.014 (1,548.312)	1,517.859 (1,488.989)	1,170.003 (1,755.452)
Observations	36	36	36	36
R ²	0.744	0.789	0.774	0.779
Adjusted R ²	0.702	0.745	0.737	0.742
Residual Std. Error	3,445.707 (df = 30)	3,182.269 (df = 29)	3,235.215 (df = 30)	3,290.781 (df = 30)
F Statistic	17.451*** (df = 5; 30)	18.079*** (df = 6; 29)	20.602*** (df = 5; 30)	

Note:

*p<0.1; **p<0.05; ***p<0.01

表 5.2: 時間序列資料的 serial correlation 檢驗。

	Dependent variable:		
	u_t		
	模型5.2	模型5.3	模型5.4
u_{t-1}	0.023 (0.194)	0.019 (0.198)	0.119 (0.181)
Constant	-52.015 (553.908)	-43.651 (503.820)	-19.639 (517.243)
Observations	35	35	35
R ²	0.0004	0.0003	0.013
Adjusted R ²	-0.030	-0.030	-0.017
Residual Std. Error (df = 33)	3,265.292	2,967.877	3,055.879
F Statistic (df = 1; 33)	0.014	0.009	0.435

Note:

*p<0.1; **p<0.05; ***p<0.01

6 結論與比較

在本論文中，我們用了數個回歸模型，希望印證高等教育的增加是否會對社會帶來外部效益。首先，在 2020 年的多變數回歸模型中，我們發現每增加就業人口中的大專生比例一個百分點，可以提高平均年薪資 0.075 萬元。接著，我們在同一樣的模型中再加入兩年的資料，並以 panel data 的方式進行估計。以隨機效果法估計的模型中，我們發現每增加就業人口中的大專生比例一個百分點，可以提高平均年薪資 0.403 萬元。(表6.1) 如果我們姑且假設平均薪資的提高，代表不只是受高等教育者的薪資提高，而是所有的就業者薪資都提高，那麼以上的結果即印證了提升高等教育確實有外部效益。不過平均薪資的提高，代表的是高等教育的內部性與外部性之總和，兩者的大小差異是我們在本研究中無法量測的。換句話說，在某些假設下，我們印證了高等教育存有外部效益，但該外部效益的大小則是我們無法量測的。

另外，我們也在整個台灣的時間序列資料中，發現政府的教育支出每提升一千元，可以提高平均家戶年所得 0.0001 萬元。這樣的分析雖然已經超過我們想針對高等教育的補助正當性做探討的範圍，但仍然某種程度上驗證了政府對教育的補助是符合最大化社會福利的。

不過本研究中所使用的數個迴歸模型，皆尚有進步空間。我們的 panel data

表 5.3: De-time trend MLR

<i>Dependent variable:</i>	
income	
edufund	0.0003*** (0.00002)
unem	-4,634.824*** (1,311.579)
indpd	-0.004*** (0.001)
avgGDP	0.275*** (0.066)
t	-2,095.528 (1,398.582)
Constant	32,611.320*** (5,618.118)
Observations	39
R ²	0.996
Adjusted R ²	0.995
Residual Std. Error	6,269.286 (df = 33)
F Statistic	1,575.655*** (df = 5; 33)
Note:	*p<0.1; **p<0.05; ***p<0.01

與時間序列模型在解釋力上都有明顯的不足，且每個模型中所使用的控制變因數量皆不多，再加上所使用的 IV 並不成功，因此對係數估計的偏誤並無法達到很好的控管。另外，panel data 模型中的嚴格外生性假設，也是我們無法驗證的。(表6.1)

最後，本文在最初即假設「平均薪資的提升代表所有人的薪資都提升。」實務上，這個假設很有可能不成立，平均薪資的提升甚至可能主要來受高等教育者的薪資大幅提升(內部效果)，而其他就業者的薪資下降。如何打破這個假設，以更具鑑別可能的資料確切量測出高等教育外溢效應的大小，才能真正對政府的「五年五百億」提供理論基礎。

估計方法	資料年分	主要解釋變數	被解釋變數		adjusted R^2
			平均薪資	平均家戶所得	
當期 MLR	2020	大專畢業生	0.075		0.997
		佔就業人口比例	(0.029)		
	2SLS		0.098		0.997
			(0.103)		
Panel Data - Random Effect	2018-2020	大專畢業生	0.403		0.487
		佔就業人口比例	(0.114)		
	2SLS		1.776		0.279
			(7.636)		
時間序列	1982-2021	政府教育支出		0.0001 (0.00003)	~0.7

表 6.1: 本研究中各項估計法的比較

參考資料

- [1] 行政院主計總處. 各縣市工業及服務業企業僱用員工薪資統計說明. Feb. 2022. URL: <https://www.dgbas.gov.tw/public/Attachment/22179491YWQ0W6Y9.pdf>. (accessed: 31.05.2021).
- [2] 中華民國統計資訊網. 縣市重要統計指標查詢系統. May 2022. URL: https://winsta.dgbas.gov.tw/DgbasWeb/ZWeb/StateFile_ZWeb.aspx. (accessed: 31.05.2021).
- [3] 內政部戶政司全球資訊網. 人口統計資料. May 2022. URL: <https://www.ris.gov.tw/app/portal/346>. (accessed: 31.05.2021).
- [4] 中華民國統計資訊網. 物價指數. May 2022. URL: <https://www.stat.gov.tw/ct.asp?xItem=35375%5C&CtNode=487%5C&mp=4>. (accessed: 31.05.2021).
- [5] 中華民國統計資訊網. 家庭收支調查. May 2022. URL: <https://www.stat.gov.tw/ct.asp?xItem=35375%5C&CtNode=487%5C&mp=4>. (accessed: 31.05.2021).
- [6] Enrico Moretti. “Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data”. In: *Journal of Econometrics* 121.1 (2004). Higher education (Annals issue), pp. 175–212. ISSN: 0304-4076. DOI: <https://doi.org/10.1016/j.jeconom.2003.10.015>. URL: <https://www.sciencedirect.com/science/article/pii/S0304407603002653>.
- [7] Marek Hlavac. *stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.3*. 2022. URL: <https://cran.r-project.org/web/packages/stargazer/index.html>.

r-project.org/web/packages/stargazer/index.html. (accessed: 31.05.2021).