



掃描資料在消費者物價指數上的應用

當今巨量資料應用蔚為顯學，運用零售業者於銷售時掃描商品條碼產生之巨量資料，可提升 CPI 統計確度並減少調查成本，惟相應挑戰也不小。本文介紹掃描資料特點、主要國家應用於 CPI 統計概況，以及我國推展情形。

李瑞彬、周治良（行政院主計總處綜合統計處科員、專員）

壹、前言

數位科技進步驅動巨量資料（Big Data）應用蓬勃發展，物價統計正是各國努力耕耘的領域之一，在資料蒐集與編製方法上不斷注入新思維。由於零售業者銷售商品時，透過掃描商品條碼（barcodes）可擷取交易資訊，包括交易日期、銷售商品之名稱、數量、單價及銷售收入等，這些交易活動所累積的巨量資料蘊含許多編算消費者物價指數（Consumer

Price Index，CPI）所需資訊，不少國家統計機構已投入研究，期能替代以往 CPI 採調查員實地查價方式，或彌補不足之處，俾有效提升調查樣本數及精確度，並降低調查成本。國際貨幣基金（IMF）與國際勞工組織（ILO）等共同編製的新（2020 年）版「消費者物價指數手冊」（Consumer Price Index Manual: Concepts and Methods，以下簡稱 2020 年 CPI 手冊），亦特闢專章介紹掃描資料（Scanner Data）的

應用與挑戰。

本文簡述掃描資料特點，綜整主要國家在 CPI 之應用現況，並說明我國目前推展方向與進度。

貳、掃描資料應用於物價統計之優點

2020 年 CPI 手冊指出，運用掃描資料於 CPI 統計具以下優點：

一、替代實地查價

調查員不再需要實地查

價，政府機關可將節省的調查資源重新配置。

二、提升查價商品籃的代表性

過去由於缺乏商品詳盡的營收與銷售量資料，調查員實地查價採立意抽樣（purposive sampling），詢問受訪者暢銷款或依貨架上商品陳列規模決定選入商品籃之產品；使用掃描資料可根據各商品銷售占比，選擇穩定銷售且較具代表性者納入商品籃，即可確保樣本的代表性。

三、更貼近消費者感受的均價

掃描資料包含一段期間內零售業者銷售各種商品的數量及營收，故可計算出同規格商品在該期間的銷售均價，包括折扣及折扣對商品銷售量的影響，故較原實地查價受限於調查資源僅能計入選定時點價格（point-in-time price）的方式，更能精準貼近消費者在該物價指數衡量期間內感受到的均價。

四、建構可按銷量加權的項目別物價指數

傳統實地查價因缺乏商品樣本花色的銷售量資訊，因此各國多採 CPI 手冊建議的未加權幾何平均法計算各基層項目指數，但透過掃描資料可即時取得樣本花色商品銷售資訊，各基層項目即可按銷量加權計算物價指數。

參、掃描資料應用於物價統計之挑戰

然而，各國政府統計機構要運用掃描資料於 CPI 編製作業時，須考量以下實務上的挑戰：

一、資料來源

政府統計機構可與零售業者協議取得，惟掃描資料具機敏性，政府須建立完善配套機制，確保資料之提供或運用安全無虞；另一來源係向第三方（如市場調查公司）購買，付費可降低資料取得難度，但會增加政府機關預算。

二、資訊系統

掃描資料屬巨量資料，故須投注相當人力及經費開發資訊系統，方能有效率地取得零售業者或第三方提供的資料，並將資料妥為儲存及處理等。

三、正確分類

首次取得掃描資料時，須花費大量資源，將各商品品項逐一對應至 CPI 各個項目，爾後雖可僅針對新增商品分類，惟仍需不斷檢視分類正確性，避免分類不當影響 CPI 統計結果。

四、資料檢核

掃描資料量龐大，須建立資料檢誤機制，並需與備用價格資料（如廠商傳單或是網路價格）比較，確保按掃描資料編製而成的物價指數合理。

五、品質調整

CPI 查價須固定品質規格，故傳統調查員實地查價時，須核對查價商品品質特徵，若

論述》統計·調查



遇商品特徵改變，則仔細記錄其變動明細，作為統計人員進行品質調整參考依據；惟掃描資料通常僅有商品名稱或型號，難有詳細商品特徵資料（如電腦的主機板、顯示卡、記憶體等規格），故較難觀察商品品質改變。

肆、各國掃描資料在 CPI 之應用

近年各國開始研究將掃描資料應用於 CPI，以下為主要國家發展近況：

一、英國國家統計局 (Office for National Statistics, ONS)

向主要商圈的零售業者索取銷售時點情報系統 (point of sale, POS) 資料，先研析適合納入 CPI 統計之掃描資料項目，並著手試編；規劃於 2022 年結合實地查價及掃描資料，按季發布試編 CPI 指數，2023 年第 1 季正式發布將掃描資料納入之 CPI 統計。

二、美國勞工統計局

(Bureau of Labor Statistics, BLS)

透過下述 2 種方式取得部分商品之掃描資料，從 2019 年起陸續納入 CPI 編算：

(一) 直接從零售業者取得銷售明細資料：2016 年 5 月起請某一零售業者提供每個銷售據點的商品（包括化妝品、珠寶、家庭用品、服飾類等）銷售價格及營收，已於 2019 年 3 月將 1,000 個商品納入 CPI 編算；2015 年 3 月起請另一零售業者每 2 個月提供 1 次處方藥的平均銷售價格，2021 年開始納入 CPI。

(二) 向第三方購買銷售數據：向市場調查公司 J.D. Power 購買新車銷售數據，2021 年開始納入 CPI。

三、日本統計局 (Statistics Bureau of Japan, SBJ)

2001 年 8 月起將 POS 資

料用於桌上型電腦及筆記型電腦的特徵迴歸模型，2006 年 8 月增加相機，2021 年 8 月再擴增電視等項目。

四、其他國家

除英國、美國與日本外，荷蘭、挪威、義大利、比利時及澳洲等國亦相繼發表相關研究，掃描資料應用於 CPI 項目如下頁表 1 所示。

伍、我國推展方向與進度

相較於其他國家在應用掃描資料時須向個別廠商索取 POS 資料庫或向資料公司洽購資料，我國因政府積極推動電子發票，且使用益趨普及，恰可作為掃描資料的來源；目前正積極探討應用電子發票資料補充 CPI 物價調查之可行性，利用便利商店、超市及量販賣場等業者開出之電子發票交易明細資料（受限於資料保護，內無業者名稱，僅包括交易日期、商品名稱、單價、數量及金額等資訊），逐步分析資料

特性，研究將其納入 CPI 編算作業模式，以提升 CPI 樣本代表性。

由於 CPI 係衡量家庭購買消費性商品及服務之價格變動情形，故查價商品之價格須為同月份各查價時點含折扣促銷之實際售價，而非原價；惟囿於部分業者電子發票交易明細係列示各販售商品名稱、單價（原價）、數量、總金額，個別商品折扣金額則以折扣代碼方式呈現，無法直接對應至應歸屬之商品，加以不同銷售管道的業者（如便利商店、超市、量販店）可能會販售相同商品，為編製符合 CPI 查價規範之商品價格，須進行資料處理，以計算出同月份某一業者銷售特定商品之折扣後實際售價，作為 CPI 查價商品之價格。

資料處理主要步驟係先將電子發票依內容適度歸類，以利判斷各張電子發票的折扣碼屬於何種商品，例如篩選出同日有同折扣碼的所有電子發票，分析共同購買的商品，藉此計算出折扣後之實際商品價格（表 2）。目前已藉由此方

表 1 各國掃描資料在 CPI 的應用

國家	CPI 已採用 / 預計採用品項
英國	預計 2023 年第 1 季起採用：燃料、家庭用品、個人照護產品、衣著、電子產品等。
美國	已採用：化妝品、珠寶、家庭用品、服飾類等。 預計 2021 年採用：處方藥、新車等。
日本	已採用：桌上型電腦、筆記型電腦、相機。 預計 2021 年 8 月採用：電視、電子辭典、攝影機及印表機等。
荷蘭	已採用：食品、個人照護產品。
挪威	已採用：食品、飲料、香菸、家庭用品、個人照護產品、汽油、衣著等。
義大利	已採用：食品（生鮮食品除外）。
比利時	已採用：食品、飲料、香菸、寵物用品、非耐久家庭用品、個人照護產品等。
澳洲	已採用：食品、飲料、香菸、寵物用品、清潔用品、非耐久家庭用品、個人照護產品等。

資料來源：作者參考美國、英國、日本相關文獻（參考文獻第 2、4 及 6 篇）自行整理。

表 2 電子發票交易明細－折扣碼處理釋例

發票	商品名稱	單價	數量	金額	實際單價
A	甲牌有機杏核仁粉	579	4	2,316	459(=579-120)
	乙牌抽取式衛生紙	309	1	309	309
	丙牌可頌 10 入	139	1	139	139
	COUPON #121941	-120	4	-480	
B	丁牌細關東麵	265	1	265	265
	甲牌有機杏核仁粉	579	1	579	459(=579-120)
	COUPON #121941	-120	1	-120	

資料來源：作者自行整理。

論述》統計・調查



式處理多數折扣促銷情形，惟仍有不同商品共用同一個折扣碼（如 3 號及 4 號電池）等特殊情況，尚須研究例外之處理方式，以得到實際的商品單價。

處理完折扣促銷、得到實際商品價格後，須正確地將商品逐一對應到 CPI 的 368 個項目。目前係以人工判斷方式進行首度歸類，並同步建立關鍵字分類機制，期能應用至爾後擴增之商品，俾商品歸類作業自動化。

陸、結論

CPI 是一項總體經濟指標，除為政府制定各項財經決策的重要參據外，也是民間薪資調整、契約價金連動的參考指標，受各界關注程度不斷升高；加上市售商品推陳出新加速、銷售管道多元，對 CPI 編製工作的挑戰日增。然隨巨量資料應用蓬勃發展，也為物價統計帶來新契機，各國政府統計機構均戮力研究零售業者的掃描資料，以作為傳統實地調查的替代或補充資料來源之一。我國由於有電子發票制度

恰可作為掃描資料的來源，以電子發票資料補充既有查價之作業已有雛型；惟電子發票資料量龐大，從資料整理、折扣碼的處理、商品與 CPI 項目對應，及至最後如何反映在 CPI 上，皆須自行撰寫高運算效能程式，以協助處理大量資料。刻正努力集思務實解決方法與標準化作業流程，期我國 CPI 能在網路爬梳（web scraping）後，續隨國際腳步以掃描資料提升物價統計之廣度與品質。

參考文獻

1. International Monetary Fund (2020), Consumer Price Index Manual: Concepts and Methods, 2020.
2. Konny et al. (2019), Big Data in the U.S. Consumer Price Index: Experiences & Plans.
3. Office for National Statistics (2020), Using alternative data sources in consumer price indices: May 2020.
4. Office for National Statistics (2021), Research into the use of scanner data for constructing UK

consumer price statistics, Apr. 2012.

5. 日本總務省統計局（2011），2010 年 CPI 消費者物價指數解說，<https://www.stat.go.jp/data/cpi/2010/kaisetsu/index.html>。
6. 日本總務省統計局（2019），第 14-15 次物價指數研究會，<https://www.stat.go.jp/info/kenkyu/cpi/index.html>。❖