校務研究: 教育巨量資料的應用
Prepared for 靜宜大學 財務與計算數學系

林世昌

國立清華大學 學習評鑑中心主任/經濟學系教授

September 10, 2015
演講大綱

- 經濟研究
- 計量經濟以及經濟實證研究
- 教育研究 – 使用巨量資料
- 校務研究
- 校務研究實例
- 結語
Cores in Economics

- Microeconomics
- Macroeconomics
- Econometrics
Fields in Economics

- **Microeconomics**
  - Game theory, math economics, market (mechanism) design, industrial organization,...

- **Macroeconomics**
  - International finance, international trade, growth,...

- **Econometrics**
  - Time series, microeconometrics, semi(non)-parametrics,...

- **Applied econometrics**
  - Empirical IO, labor economics, education economics, defense economics, sports economics, economics of law,...

- To form the efficient GMM, the optimal weight matrix is:

\[
\hat{W}_n = \left( \frac{Z' \hat{\Omega} Z}{n} \right)^{-1}
\]

- Traditional way: \( \hat{\Omega}_{HC0} = \text{diag}(e_i^2) \) or \( \hat{\Omega}_{HC1} = \frac{n}{n-k} \text{diag}(e_i^2) \)

- With estimated quasi-hat matrix \( \hat{H}_i \), we have
  - \( \hat{\Omega}_{HC2} = \text{diag}\left( \frac{e_i^2}{1-\hat{H}_i} \right) \): Horn, Horn, and Duncan (1975)
  - \( \hat{\Omega}_{HC3} = \text{diag}\left( \frac{e_i^2}{(1-\hat{H}_i)^2} \right) \): MacKinnon and White (1985)
  - \( \hat{\Omega}_{HC4} = \text{diag}\left( \frac{e_i^2}{(1-\hat{H}_i)^{\delta_i}} \right) \): Cribari-Neto (2003)
  - \( \hat{\Omega}_{HC5} = \text{diag}\left( \frac{e_i^2}{(1-\hat{H}_i)^{\alpha_i}} \right) \): Cribari-Neto et al. (2007)
Bayesian Interpretation: Lin and Chou (2012)

Use the quasi-hat matrix derived in Lin and Chou (2013) and set the prior parameter \( v_i \) by \( v_{HCS,i} = (1 - \hat{H}_i)^{-\eta_{s,i}} - 1 \) such that the corresponding posterior parameter \( \tilde{v}_{HCS,i} = (1 - \hat{H}_i)^{-\eta_{s,i}} \), it is easy to show that the approximation of \( \mathbb{E}[\beta(\theta)|D_n] \) is:

\[
\beta^*(\tilde{\theta}_{HCS}) = \beta^* \left( \frac{H_s}{\sum_{i=1}^{n}(1 - \hat{H}_i)^{-\eta_{s,i}}} \right),
\]

where \( H_s = ((1 - \hat{H}_1)^{-\eta_{s,1}}, ..., (1 - \hat{H}_n)^{-\eta_{s,n}})' \), and the approximation of \( \text{var}[\beta(\theta)|D_n] \) is:

\[
V_{HCS} = J_s C (H_s)' Z' \text{diag} \left( \frac{u_1^*(\tilde{\theta}_{HCS})^2}{(1 - \hat{H}_1)^{\eta_{s,1}}}, ..., \frac{u_n^*(\tilde{\theta}_{HCS})^2}{(1 - \hat{H}_n)^{\eta_{s,n}}} \right) Z C (H_s),
\]

where \( J_s = (1 + \sum_{i=1}^{n}(1 - \hat{H}_i)^{-\eta_{s,i}})^{-1} \sum_{i=1}^{n}(1 - \hat{H}_i)^{-\eta_{s,i}} \). In addition, \( \eta_{2,i} = 1, \eta_{3,i} = 2, \eta_{4,i} = \delta_i \), and \( \eta_{5,i} = \alpha_i \).
Forecasting: Lin, Chou and Chou (2010)

- **Ratio-based test:**

\[
RB_T = \sqrt{T} \left( \frac{\bar{a}_T}{b_T} - c_0 \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \frac{a_t - c_0 \cdot b_t}{b_T} \right) \xrightarrow{D} \mathcal{N}(0, V_o)
\]

- **Difference-based test:**

\[
DB_T = \frac{\sqrt{T}d_T}{\sqrt{f}} \xrightarrow{D} \mathcal{N}(0, 1),
\]

- **Bayesian approach:** The likelihood function \( f(Y^T|\theta) \) is given by:

\[
(2\pi)^{-T}(\sigma_a^2\sigma_b^2)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \left( \frac{\sum_{t=1}^{T} (A(L)(e_t^2 - (x_t - \mu)^2 - D))^2}{\sigma_a^2} + \frac{\sum_{t=1}^{T} (B(L)(x_t - \mu))^2}{\sigma_b^2} \right) \right\}.
\]

The joint posterior density can be expressed as follows:

\[
p(\theta|Y^T) \propto p(\theta) \cdot f(Y^T|\theta) \propto \mathcal{K}(\theta).
\]
Decomposition by Sub-groups: Lin and Yun (2015)

**Standard OB decomposition**

\[
\hat{y}_j^f - \hat{y}_j^m = \sum_{k=2}^{K} \hat{\pi}_{jk}^m (\bar{q}_{jk}^f - \bar{q}_{jk}^m) + (\bar{x}_j^f - \bar{x}_j^m)\hat{\theta}^f \\
+ (\hat{\alpha}_f^f - \hat{\alpha}_j^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + \bar{x}_j^f (\hat{\theta}_j^f - \hat{\theta}_j^m) + \sum_{k=2}^{K} (\hat{\pi}_{jk}^f - \hat{\pi}_{jk}^m)\bar{q}_{jk}^f
\]

- **Explained**
- **Unexplained**

**Normalized OB decomposition**

\[
\hat{y}_j^f - \hat{y}_j^m = (\hat{\beta}_j^f - \hat{\beta}_j^m) - (\hat{\beta}_j^m - \hat{\beta}_j^m) \\
+ (\hat{\alpha}_f^f - \hat{\alpha}_j^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + (\hat{\pi}_j^f - \hat{\pi}_j^m) + \sum_{k=2}^{K} (\hat{\pi}_{jk}^f - \hat{\pi}_{jk}^m + \hat{\pi}_{jk}^m)\bar{q}_{jk}^f + \bar{x}_j^f (\hat{\theta}_j^f - \hat{\theta}_j^m) + \text{Explained}
\]

- **New measure**
- **Unexplained**
labor supply, wages, employment

- Review the development of estimating female labor supply in Taiwan (MUS)
- Does EITC induce more labor supply? (SFIE, MUS)
- Gender wage gap by college major (MUS)
- Gender wage gap by industry (MUS)
- Employment status and choice of employment sector of foreign spouse (Survey of Foreign and Mainland Spouses’ Life Status)
Method: Incidental Censoring & Multinomial Logit

 Heckman (1979): self-selection model

- consisting of an outcome equation and a selection equation

  **outcome equation:** \( y_1 = x'\beta + \sigma_1 u_1 \)

  **selection equation:** \( y_2 = z'\gamma + u_2 \),

where \( x \) and \( z \) may overlap.

- Multinomial logit specifies:

\[
P_{mi} = \frac{\exp [X_i'\beta_m]}{\sum_{l=1}^{M} \exp [X_i'\beta_l]}
\]

- In-determinancy in choice probability \( P_{mi} \)? Let \( \beta_m^* = \beta_m + \delta \).

- A solution is to set \( \beta_1 = 0 \). Therefore,

\[
P_{1i} = \frac{1}{\sum_{l=1}^{M} \exp [X_i'\beta_l]}
\]
Empirical Industrial Organization

- Firm performance vs. FDI, cluster, RD, innovation activities
  - IFDI, OFDI, Import, Export vs. product innovation/effective patents (TTIS-I, SOFTM)
  - FDI Strategies (expansive and defensive) vs. productivity and innovation capability growth (ICSC, SOFTM)
  - Complementarities of R&D Strategies (internal, external, cooperative) on innovation performance (new product sales, patents, RoR on RD) (TTIS-I, ICSC)
  - Local industrial agglomeration and vertical FDI to China vs. productivity (TSCMO and TEJ)
  - Choices between standard (V1, H1) and non-standard FDI (V2, V3, H2) strategies (SOFTM)
  - Innovation strategies (product, process, and both) vs. firm performance (TTIS II, ICSC)
Method: Patent Counts Modelling

- The two-parameter Negative Binomial model is as follows:

\[
Pr[W_i = k | X_i] = \frac{\Gamma(k + \theta)}{\Gamma(\theta)\Gamma(k + 1)} \left(\frac{\theta}{\exp(X_i'\beta) + \theta}\right)^\theta \left(\frac{\exp(X_i'\beta)}{\exp(X_i'\beta) + \theta}\right)^k,
\]

where \( W_i \) denotes the number of valid patents for firm \( i \) and \( \Gamma(\cdot) \) is the Gamma function such that \( \Gamma(s) = \int_0^\infty z^{s-1}e^{-z}dz \) for \( s > 0 \).

- Nonlinear GMM panel count model can be specified as:

\[
P_{it} = \exp(x_{it}'\beta + \eta_i) + u_{it},
\]

where \( P_{it} \) is a nonnegative integer count variable with observed frequencies, \( i = 1, \cdots, N, t = 1, \cdots, T_i \) and \( T_i \) permits unbalanced panels, \( x_{it} \) denotes a vector of firm \( i \)'s determinants of patenting at time \( t \), \( \eta_i \) represents the unobserved heterogeneity specific to firm \( i \), and \( u_{it} \) is an error term.
Military expenditures (Milex) related issues

- Milex vs. inequality with global panel data
- Milex vs. unemployment with global panel data
- Milex vs. inflation rates by different stages of defense policy in Taiwan
- Milex vs. welfare expenditures – “guns-butter” argument
- Impact of war types (international or civil) on food price with global panel data
Method: Dealing with Panel Regression

▶ The original model in level form:

\[ \text{Health}_{it} = \alpha_0 + \alpha_1 \text{Edu}_{it} + \alpha_2 \text{Milex}_{it} + \mathcal{W}_{it}'\eta + c_{1i} + \nu_{it} \]

▶ weakly exogeneity assumption implies:

\[ \text{E}[\text{Milex}_{is}\nu_{it}] = 0 \text{ for } s \leq t - 1 \]

▶ The differenced panel model removes fixed effect \(c_{1i}:\)

\[ \Delta\text{Health}_{it} = \alpha_1 \Delta\text{Edu}_{it} + \alpha_2 \Delta\text{Milex}_{it} + \Delta\mathcal{W}_{it}'\eta + \Delta\nu_{it} \]

▶ the moment condition becomes:

\[ \text{E}[\text{Milex}_{is}\Delta\nu_{it}] = 0 \text{ for } s \leq t - 2 \]

▶ Milex lagged two or more can serve as an instrument in the FD model
Sports Economics

- Focusing on major league baseball (MLB)
- Using fruitful panel data to do the following:
  - Effect of players’ wage dispersion on team performance
  - Postseason participation and payroll disparity
  - Factors that drive the team performance at different stages
  - Impact of policy change (wild card system) on attendance
  - Effect of contract length on batters’ performance
Method: Regressions at Different Stage Outcomes

- # winning games: RE/ FE panel or probably panel IV/GMM
  - attendance: endogenous

- winning percentage: panel proportional data method
  - `glm meals yrrnd parented api99, link(logit)`
    - `family(binomial)` robust

- whether enter the postseason at year \( t \): panel probit/logit or dynamic panel probit
  - `xtprobit (RE); xtlogit (FE, RE)`

- whether win the division (AL or NL) championship at year \( t \): panel probit/logit or dynamic panel probit
  - `xtprobit (RE); xtlogit (FE, RE)`

- whether win the world series at year \( t \): panel probit/logit or dynamic panel probit
  - `xtprobit (RE); xtlogit (FE, RE)`
研究主體: 國小、國中、高中、大學、研究所、博士、博士後
政策意涵: 個人、家庭、學校、政府
資料形態: 校外、清大校內資料庫

1. 台灣學生學習成就評量資料庫 (Taiwan Assessment of Student Achievement, TASA)
2. 台灣教育長期追蹤資料庫 (Taiwan Education Panel Survey, TEPS)
3. 台灣高等教育資料庫 (Taiwan Higher Education Database, THED)
4. 國家科技人力資料庫 (National Profiles of Human Resources in Science and Technology, NPHRST)
5. 清大教育整合資料計畫 (by CLASS)
「臺灣學生學習成就評量資料庫」(Taiwan Assessment of Student Achievement, TASA) 是由國家教育研究院策劃督導，所進行之長期性研究計畫，該研究所蒐集之資料涵蓋國小、國中、高中 (高職) 此三個教育階段，用以蒐集學生國語、英語、數學、社會、自然此五科之學習成就表現。

資料蒐集期間 2005 至 2011 年：
① 2005 年：國小六年級之國語、英語、數學
② 2006–2007 年：國小四、六年級、國中二年級、高中及高職二年級之國語、英語、數學、自然、社會 (社會國小四年級無資料提供)
③ 2009 年：國小四、六年級之國語、英語、數學、自然、社會 (英語、社會國小四年級無資料提供)
④ 2010 年：國中二年級之國語、英語、數學、自然、社會
⑤ 2011 年：高中及高職二年級之國語、英語、數學、自然、社會
Example

We utilize TASA to empirically evaluate whether the test score differentials between pupils with immigrant and native mothers are substantial across subjects, grades and years.

1. There exist test score differentials between the two groups after controlling for the students’ individual characteristics and family background – particularly Chinese, Math and English.

2. The academic gaps between native students and pupils with mothers from Southeast Asian countries tend to widen, while the students’ performance is about the same as that for native students if their mothers are from mainland China.


2001年開始, 對當年為國中一年級以及高中、高職和五專二年級之學生、學生家長、老師、和學校, 進行二至四次的資料收集。目前資料庫已經完成調查與資料釋出, 並委託中央研究院人文社會科學研究中心「調查研究專題中心」管理資料釋出事宜。

後續調查 (TEPS-B): 教育和勞力市場的連結, 委由國立政治大學團隊 (計畫網址: http://tepsb.nccu.edu.tw) 持續進行。
Example (Cram Schooling Timing Decision)

Lin and Lu (2010) take advantage of Taiwan Education Panel Survey data to evaluate the timing of the coaching effect on mathematics learning for junior high-school students. Our main finding suggests that the best strategy for the budget-limited parents is to send their children to cram schools for intensive learning in the 8th rather than earlier in the 7th grade, in order to enhance the scores for the math tests being held in the 9th grade.

Example (BMI and academic performance)

Lu, Chou and Lin (2014) take advantage of the Taiwan Education Panel Survey (TEPS) data set to empirically evaluate whether a student’s academic performance, measured by his or her test score, is affected by the body weight status of this student. In particular, we simultaneously consider the impact of being ”overweight” and ”underweight” on students’ academic performance. The differentiated effect of weight status by gender is also studied in depth.

政策的施行往往需要詳細的資料加以評估, 教育議題也不例外, 不論是先進國家 (如美國), 或開發中的國家 (如埃及), 都有教育資料庫的建置, 提供高品質的教育資料供決策參考。

鑑於我國縱貫的教育追蹤資料較少, 自 2003 年彭森明教授在國立清華大學開始執行「台灣高等教育資料系統之建置與運用」, 之後由國立師範大學承接該調查並維護「台灣高等教育資料庫」(Taiwan Higher Education Database, THED)。THED 是當前我國最重要的高等教育資料庫之一。

調查對象為包含大學生, 碩士生和博士生。其中除了橫向的調查外, 也有追蹤調查, 包括大學一年級, 三年級, 大四, 及畢業後一年的狀況調查。其中也含有各技職院校的學生, 並在幾個年度進行大專院校教師 (2004; 2009) 的調查。
THED – Policy Implication

Example

We extensively explore the impact of socio-economic status (SES) on a variety of students’ outcome in different stages (i.e., junior high, high school, college/university)

1. whether attend a public/private high school; high school grade/ranking
2. competence test scores; college entrance methods
3. academic performance in college; book coupon award
4. job search time; monthly salary one year after graduation from college

國家科技人力資料庫

有鑑於科技人力資源關乎國家競爭優勢的建立與維持，先進國家與著名國際組織早已著手進行科技人力的觀測與探討，國家實驗研究院科技政策研究與資訊中心於 2000 年底進行「國家科技人力資源庫」(National Profiles of Human Resources in Science and Technology, NPHRST)之規劃與建置。

NPHRST 以居住在台、澎、金、馬三個月以上且獲有博士學位者為對象。目的在蒐集我國博士人力的教育及就業等資訊，提供博士人力配置之資訊，作爲政府施政及學、研、產各界人士之參考。

NPHRST 自 2002 年開始運作迄今已逾十年，所收錄之人才資料截至 2013 年 10 月 1 日爲止，共計 74,301 筆個人資料。
We investigate whether doctorates can be completed faster overseas than domestically. By taking advantage of a large data set compiled by the NPHRST in Taiwan, we find that

1. Taiwanese students studying in foreign schools (and who will come back to Taiwan) on average take 9.61 months less than those earning PhD degrees domestically.

2. 9.61 months reflect approximately a 60,000 USD opportunity cost.

3. Other controls, such as gender, age, fields of study, prior degrees, cohorts, school/major matching scenarios and tiers are net out.

Definition (Yale University)

The Office of Institutional Research (OIR) provides information to support university decision-making through a variety of analytic activities, data-gathering tasks, and research projects. OIR serves as a clearinghouse for most statistical information about the University. The office reports to and places first priority on work for the Provost and for the Officers of the University, while also working regularly with many other offices and individuals both inside and outside of Yale.
Definition (Univ. of Texas at Austin)

The Office of Institutional Accreditation and Effectiveness (IAE) establishes vision, leadership, and processes to strengthen the university’s institutional effectiveness and accreditation efforts. We value strategies that advance educational, programmatic, and operational excellence in a way that is evidence based, locally defined, and efficiency driven. Across campus, we promote practices that result in:

1. higher student academic achievement
2. an enhanced student experience
3. aligned and transparent decisions
4. readily available information for improvement, accountability, and accreditation
Definition (Association for Institutional Research)

The field of institutional research (IR) is over 50 years old and is embedded in nearly every college and university in the United States and many others around the world. Often working behind-the-scenes, IR professionals support campus leaders and policy makers in wise planning, programming, and fiscal decisions covering a broad range of institutional responsibilities. These areas can include research support to senior academic leaders, admissions, financial aid, curriculum, enrollment management, staffing, student life, finance, facilities, athletics, alumni relations and many others. In addition to providing the data-informed foundation for good decision making, institutional researchers use the data they collect for governmental reporting and to benchmark their results against similar institutions.
什麼是校務研究?

Definition (彭森明講座教授)
針對學校內部之行政運作、管理、環境設備、教學措施、學生學習歷程與表現、學術研究，以及師生背景等過去與現在的資料，進行彙整、分析與闡述，使之轉換成為有用的資訊，供校務規畫、政策制定，以及方案設計、執行與評鑑等決策論證之用，所演化成的一種特殊研究領域。
什麼是校務研究？

- 由上述定義可得知「校務研究」的研究範疇相當廣泛
- 『校務研究辦公室』扮演的角色
  - IR vs. 大專院校
  - 國發會 vs. 行政院—Dr. Kuan’s view
- 我國的大專院校即使沒有明文設立『校務研究辦公室』，但平時已經都在從事「校務研究」的相關議題
  - Retention rate, transfer rate, graduation rate, enrollment rate, etc
  - SAT scores vs. required courses learning
  - 中國醫藥大學、台北醫學大學、逢甲、弘光、玄奘、中山大學
- 相對於西方常設的『校務研究辦公室』，我國目前的「校務研究」的現況
  - 不同議題由校內不同的處室負責
  - 資料也通常散在個別負責的辦公室 (如註冊組、招生組、課務組、畢業輔導室、就業輔導室、計算機中心等等)
- 研究方法以敘述統計 (bar, pie charts/group means/cross tables) 爲主
Escalating demands on limited public funds have made governments increasingly concerned with greater effectiveness and efficiency in the allocation and use of resources, and with maximising the achievement of desired outcomes. This has engendered an increasing emphasis on quality management as a vehicle for implementing public accountability. In the tertiary education sector, the achievement of quality within funding constraints is now an imperative for all institutions.
校務研究的決策方式

「校務研究」決策方式的特點 (有別於我國當前「廣義的」校務研究):

1. 跳脫「有限理性」(Bounded rationality) 的侷限
2. 統整數據並萃取有價值的訊息給決策者
3. 訊息的提供以嚴謹的統計方法為之
4. 資料與訊息的提供須有效率的進行
5. 以 IR 蒐集的資訊作爲決策重要依據，亦即
   - 以資料爲基礎的決策 (Data-driven decision making)
   - 以證據爲基礎的決策 (Evidence-based decision making)
Evidence-based Decision in Daily Life

- A recent work from a team at the University of California - San Diego and Harvard University


- Over time, **too little deep sleep** may also take a toll on your heart by contributing to **high blood pressure**
Evidence-based Decision in Daily Life

(a) My BP Trend

(b) My BP Daily Record
Evidence-based Decision in Daily Life

(c) Deep Sleep Record

(d) Heart Rate Record
校務研究的決策方式

► 「校務研究」決策程序等同於從事實證研究 (empirical study)
► 基本的程序如下：

1. 確定議題和資料需求
2. 開始資料蒐集
3. 進行資料分析
4. 分析報告與交流
5. 形成最終決策
校務研究的簡史 – 美國

早期的校務研究活動
- Cowley (1966): 1701 年 Yale University 組織管理研究
- Stephens (1966): 1847 年 University of Missouri 學生問卷調查
- Stephens (1966): 1856 年 University of Missouri 董事會的要求

50、60 年代開始成長
- 戰後嬰兒潮、高教擴張
- 教師、學生、教學資料需求孔急
- 1960 年, 美國「校務研究協會」(Association for Institution Research, AIR) 成立, 爲校務研究提供資訊與經驗交流的平台

70、80 年代電腦使用
- 調查問卷、電腦、統計方法使用
- 校務研究機構紛紛成立

90 年代至今
- 先進資料庫、網路
- 研究內容系統化、研究方法多元化
- 決策透明、跨校比較
校務研究的簡史 – 日本

- 日本於 2004 年以「認證評鑑」與「國立大學法人評鑑」為基礎，展開首次的大學評鑑制度。
- 2005 – 2011 年為「第一期認證評鑑」實施階段，校務研究的推動與「大學資料庫」(University Portrait) 的建構，為 IR 的前導。
- 2010年「大學評價學位授予機構」成立「IR 研究會」，用以宣導 IR 的概念。
校務研究的簡史 – 日本

► 2009年同志社大學、北海道大學、大阪府立大學與甲南大學等4所大學共同成立「4大學 IR 策略聯盟」，結合校務研究與教學來提高教學品質
► 2012年擴大至 8 所大學成立「教學評價體制 (IR network) 大學課程教育品質保證」，採相同問卷，以「學生調查、畢業生調查、英語調查」為主題
In additional to AIR in the US, there are...

1. **EAIR** (1979; Amsterdam, the Netherlands; Tertiary Education and Management)
2. **AAIR** (1988; Australia; JIR)
3. **SEAAIR** (2010; IMU, Malaysia; JIRSEA)
4. **SAAIR** (1994; South Africa)
5. **DAIR** (1997; Dutch)
6. Other affiliated organizations (known as AOs)

**TAIR** (2015; forthcoming)
校務研究的範疇

定期報告

年報: annual report – Yale “Factsheet”, UCI “College portrait”
共同報告: Common Data Set (CDS) – College Board, Peterson’s guide, US News & World Report

主題式分析

1. 畢業率分析
2. 學生學習投入
3. 學生學習成效
4. 學生滿意度
5. 畢業生追蹤調查
6. 老師教學成效
7. 校園氣氛、種族關係
8. 評估管理部門效率
9. 財務規劃 etc...
校務研究的運作條件

► 理念得到校方支持
  ▶ Vice President, Assistant to the President, Provost, even President/Chancellor

► 成立校務研究專責辦公室
  ▶ Institutional Research, Research and Planning, Institutional Planning, Academia Planning

► 校務研究辦公室人力配置
  ▶ UC Berkeley (10), UC Irvine (7), Bates College (2)

► 校務研究辦公室人力資源
  ▶ Computer Science, (Higher) Education, Statistics
  ▶ IR Director is key to successful IR!
RESPONSIBILITIES:

- Provides oversight and supervision for the institutional research and assessment, and contributes to institutional planning ensuring that meaningful, appropriate/accurate data and supporting documentation are made available to meet the decision-making needs of the campus;
- Works with others to coordinate and improve systems for assessing effectiveness of campus programs and services, including; works with IT on building institutional data warehouse.

PREFERRED SKILLS & KNOWLEDGE:

- Doctorate in a related field preferred, experience in quantitative and qualitative research, statistical analysis, and reporting required.
- Excellent technology skills, particularly with data presentation and analysis software such as SAS, SPSS, Excel, Access, R, etc.

SALARY:

- The anticipated hiring salary range is $85,000 to $112,000.
校務研究的資料整合議題

- 如前面概念篇所述，校務研究包山包海，平時大家也都在進行
- 先前也強調 (相對嚴謹的) 校務研究具以下特質:
  - 資料整合程度
  - 研究方法嚴謹度 (to be discussed later)
- 資料是否整合完善攸關於校務研究的效率性和完整性
  - 四處要資料、串資料
  - 不僅檢視少數變項之間的關連性
校務研究的資料整合議題

- 資料庫整合並不容易
  - 並非所有資料均進到資料倉儲 (Personal computer)
  - 並非所有資料有同樣的編碼 (Big 5 or unicode)
  - 資料庫管理系統亦可能不同 (MS SQL or Sybase)
  - 諮商中心資料

- 校務資料整合流程：入學前、在學端以及畢業後

  原始資料
  教務處
  學務處
  總務處
  研發處
  其他部門

  \[ \Rightarrow \]

  資料倉儲
  data warehouse

  \[ \Rightarrow \] 使用者界面
  \[ \Rightarrow \] 資料輸出&應用

- 資料授權使用
以評估多元入學成效為例

- 檢視推薦甄試 (個人申請、學校推薦、繁星計畫) 以及指定考試等入學管道之學業 (或非學業) 表現
  - admission methods vs. academic performance (avg. final grade)
  - summary stat approach
  - group mean comparison

- 但是, 和 (相對嚴謹的) 校務研究有何差異？
  - 資料來源
    1. need more data – family background, student engagement (time allocation), association (part-time job)
    2. administrative-level + survey data

- 研究方法
  1. regression approach by controlling confounding factors
  2. propensity score matching method (treatment vs. control)
  3. non-linear models may be required

- 可獲致較為可信且精確的評估
國立清華大學教育資料整合計畫

- 2008年成立教學發展中心 (Center for Teaching and Learning Development, CTLD):
  - 教師發展與助教培訓組
  - 學習促進與研究規劃組: 2010 teaching evaluation survey
- 2011年成立學習評鑑中心 (Center for Learning Assessment Service, CLASS)
- 增進 (評估) 教師教學、學生學習成效
- 行政業務 + 研究導向
- A postdoc allocation, an assistant and me (3)
- 整合校內教育資料庫進行細緻、嚴謹的政策評估、建議
- semi-IR – 「準校務研究機構」 – (校務評鑑與研究中心)
- administrative-level (教務、學務) vs. survey data (學評中心)
清大學習評鑑中心主持之調查研究

- **2010–2014 畢業生學習回顧調查**
  - 校、系滿意度、教育目標、核心能力、未來發展...

- **2012–2014 大二學生學習經驗調查**
  - 學習適應、時間分配、上課方式、教育目標...
  - can be compiled to be a panel data set

- **2013–2014 研究生學習回顧調查**
  - 和畢業生問卷同但多加研究生議題、指導教授...

- **2011 (2015??) 教師對「教學意見評量調查」之調查**
  - 攸關教學成效甚至教師薪資...

- **2013–2014 新生報到系統填答高中以及家庭背景資料**
  - 盡量蒐集以避免重複詢問
Example

Examine the outcomes of the Multi-Channel College Entrance System by comparing four different types of students’ academic performance

- (1) School Recommendation; (2) Individual Application; (3) Multi-star Project, and (4) Joint College Entrance Examination

- Outcome variables include: overall GPAs, semester GPA, # of receiving the University Presidential Award, and whether the students are experienced a half of course credits failed (aka 2-1)

Example

Examine the learning outcomes of the Multi-Channel College Entrance System for economics majors on the required and selective courses

- (1) School Recommendation; (2) Individual Application; (3) Multi-star Project, and (4) Joint College Entrance Examination

- Outcome variables include: GPA for each required course, GPA for each selective course

Example

Examine the learning outcomes of the Multi-Channel College Entrance System on academic and non-academic performance

- (1) Application-based method, and (2) Exam-based method

- Outcome variables include: class ranking for each semester; acquired ability on social science and science (two factors); degrees of satisfaction on courses, teaching, jobs; engagement in club association/ part-time jobs; participation in international exchange program/ international volunteer program

The overall cumulative GPAs/semester GPAs are continuous outcome variables. The *multiple regression model* can be estimated:

\[ Y_i = \alpha + D_i^{AM} \beta + X_i' \gamma + \varepsilon_i, \quad i = 1, \ldots, n, \]

Due to the non-negative integer of the “extremely good” outcome variable, the *negative binomial regression model* is formulated below:

\[ E[Y_i|D_i^{AM}, X_i] = \exp[\alpha + D_i^{AM} \beta + X_i' \gamma], \quad i = 1, \ldots, n. \]

For this binary “extremely bad” outcome variable, a typical *logistic regression* can be estimated as follows:

\[ \Pr[Y_i = 1|D_i^{AM}, X_i] = \frac{1}{1 + \exp[-(\alpha + D_i^{AM} \beta + X_i' \gamma)]}, \quad i = 1, \ldots, n, \]

where \( Y_i = 1 \) if a student has experienced a half of his/her course credits failed, and \( Y_i = 0 \) otherwise.
The impact of the college entrance routes on the academic performance in a type \( j \) status can be expressed as follows:

\[
E[\gamma^j_i - \gamma^o_i | \omega_j \neq o = 1] = E[\gamma^j_i | \omega_j \neq o = 1] - E[\gamma^o_i | \omega_j \neq o = 1], \quad j = 1, 2, 3.
\]

Nearest neighbor matching:

\[
\tau_{NR}^{\text{NR}} = \frac{1}{N_T} \sum_{i \in T} \left\{ Y^T_i - \sum_{j \in C(i)} w_{ij} Y^C_j \right\}
\]

Kernel matching estimator is given by:

\[
\tau^K = \frac{1}{N_T} \sum_{i \in T} \left\{ Y^T_i - \frac{\sum_{j \in C} Y^C_j K(\frac{p_j - p_i}{h_n})}{\sum_{k \in C} K(\frac{p_k - p_i}{h_n})} \right\},
\]

where \( K(\cdot) \) is kernel function and \( h_n \) is a bandwidth parameter.

Eric S. Lin (CLASS & Economics, NTHU)  IR and Education Big Data  September 10, 2015  54 / 60
進行中或未來的研究

Example

- Teaching and research – substitute or complementary?
- PhD (local vs. foreign students) funding allocation
- Impact of OCW usage on campus/off-campus learning
- Admission methods vs. performance on required/selected courses
- Teaching evaluation vs. grading policy
- Evaluation of NTHU Residential College
- Peer effects investigation in NTHU dormitories
- Learning effect of English teaching
- Comparison of graduate attributes between Taiwan and Australia
“Education is the most powerful weapon which you can use to change the world” by Nelson Rolihlahla Mandela (1918–2013)

教育/校務研究提供 empirical evidence 給政府/校方的決策者
  ▶ 協助大學 (各級學校) 追求卓越, 確保行政、教學、研究、服務的有效發展與品質保證

政府/校方應該大力支援教育/校務研究的發展
  ▶ 即使沒有正式的校務研究機構
  ▶ 鼓勵老師利用學校資料進行有意義的教育研究

我國校務研究的推廣有助於提升校務研究的品質
  ▶ 校務研究在台灣成為一門專業
  ▶ Graduate Certificate in Institutional Research – Penn State
  ▶ Master of Arts in Higher Education IR – Michigan Univ.
  ▶ Graduate Certificate in Institutional Research & Policy Analysis – ASU
Thanks for your attention!

Comments are welcome!


