B.教育部補助大專校院延攬國際頂尖人才執行績效報告

一、基本資料

學校名稱及聘任系所	國立清華大學統計學研究所		
玉山(青年)學者姓名	**蔡瑞胸**	**職稱**	**客座教授**
聘任方式	■**玉山學者** □專任教師 ■短期交流 □**玉山青年學者**		
經費執行期間	**107 年 12 月 1 日 至 108 年 11 月 30 日**		
聯絡人	**單位：統計學研究所** **職稱及姓名：銀慶剛所長** **聯絡電話：03-5715131＃42645** **傳　　真：03-5728318** **電子信箱：cking@stat.nthu.edu.tw**		

二、執行情形

(一)玉山(青年)學者工作項目及內容 (如教學工作或研究計畫等)

蔡瑞胸院士於清華大學客座期間與研究團隊針對三個資料科學相關主題，完成了四項研究工作。茲將計畫主題及研究成果之摘要，臚列如后。

研究計畫主題一：Model Selection for Unit-root Time Series with Many Predictors

參與人員：芝加哥大學 (蔡瑞胸、黃碩傑)、清華大學 (銀慶剛、黃學涵)

完成論文初稿，準備投稿中，摘要如下：

Abstract: In this work, we study the problem of model selection for non-stationary data when a large number of predictors are available. We propose an algorithm, FSR+HDIC+Trim, that combines a stepwise method known as Forward Stepwise Regression (FSR) and the high-dimensional information criterion (HDIC) introduced by Ing and Lai (2011). In particular, we establish a convergence rate of the empirical error for the FSR, and show the FSR+HDIC+Trim is model-selection consistent. Its performance is illustrated by simulation studies as well as an example application to U.S. macroeconomic data.

Keywords: Forward Stepwise Regression, Orthogonal Greedy Algorithm, LASSO, Adaptive LASSO, non-stationary time series, ARX model

研究計畫主題二：Spatial Temporal Modeling and Inference

參與人員：芝加哥大學 (蔡瑞胸)、清華大學 (徐南蓉、鄭又仁)、中研院 (黃信誠、黃世豪)

研究議題共二項，皆完成論文初稿，準備投稿中，摘要如下：

(1) Matrix-variate Autoregressive Spatial Temporal Models (N-J Hsu, H-C Huang, R.S. Tsay)

Abstract: Matrix-variate time series are common in economic, medical, environmental and atmospheric sciences, typically associated with large matrix dimensions nowadays. We introduce a structured autoregressive (AR) model to characterize the temporal dynamics in matrix variate time series, in particular the AR matrices are formulated as a bilinear form to reduce the model dimension. This bilinear parameter structure has the advantage of highlighting the dynamic interactions among column and row factors in the matrix, making the model highly explainable. For spatio-temporal matrix data, with spatial information implied from columns and rows, the AR coefficients

can be further constrained to ignore the dynamic effects beyond certain spatial neighborhoods. Furthermore, making use of spatial information again, we consider a non-stationary multi-resolution spatial covariance model for the innovation error matrices. The proposed spatio-temporal autoregressive model is very flexible of capturing heterogeneous spatial and temporal features, but still maintains a parsimonious parametrization. For inference, the maximum likelihood (ML) is considered and a fast algorithm is developed accordingly to implement the ML estimates for all the model parameters. A simulation study and an application to wind speeds in the western Pacific Ocean are provided to demonstrate merits of our proposed methodology.

Keywords: bilinear auto-regression, dimension reduction, matrix-variate time series, maximum likelihood, multi-resolution spline basis functions

(2) Testing Independence Between Two Spatial Random Fields (S-H Huang, H-C Huang, R.S. Tsay)

Abstract: In this article, we consider testing independence between two spatial random fields of dimension p and q with sample size n, where both p and q are allowed to be larger than n. We impose no spatial stationarity and no parametric structure for the two random fields. Our approach is based on canonical correlation analysis (CCA). But instead of applying CCA directly to the two random fields, which is not feasible for high-dimensional testing considered, we adopt a dimension-reduction approach using a special class of multiresolution spline basis functions. These functions are ordered in terms of their degrees of smoothness. By projecting the data to the function space spanned by a small to moderate number of leading basis functions, the spatial variation of the data can be effectively preserved. The first sample canonical correlation coefficient between the two random fields in the projected function space is used to construct the test

statistic, which has an asymptotic Tracy-Widom distribution under the null hypothesis. Our proposed method automatically detects the spatial structure between the two random fields. In addition, it is directly applicable to irregularly spaced data. We show that our test is consistent under mild conditions and provide two simulation experiments to demonstrate the effectiveness of the proposed method. Moreover, we apply our method to investigate whether the precipitation in continental east Africa is related to the sea surface temperature (SST) in the Indian Ocean, and whether the precipitation in western Australia is related to the SST in the north Atlantic Ocean.

Keywords: canonical correlation analysis, dimension reduction, high-dimensional test, irregularly spaced data, multiresolution spline basis functions, Tracy-Widom distribution

研究計畫主題三：Robust Estimation in High-Dimensional Vector Autoregressions

參與人員：芝加哥大學 (蔡瑞胸)、清華大學 (銀慶剛、邱海唐、黃碩傑)、中研院 (蔡恆修)

完成論文初稿，準備投稿中，摘要如下：

Robust Estimation in High-Dimensional Vector Autoregressions

Abstract: Vector autoregressive (VAR) models are commonly used to model time series data. However, financial time series data, e.g., stock returns and yields, are sometimes faced with The extreme values due to special events (such as the 2008 financial tsunami), and hence Prediction or parameter estimation based on least squares estimation is easily distorted by a few extreme values. Moreover, in the big data era, vector time series data with large dimension are collected in various scientific fields. When the dimension becomes large, the parameters in the VAR model can be much larger than

the number of observations. This leads us to consider high-dimensional VAR models with parameters estimated by some robust estimates instead of the least squares one. In this project, we aim at proposing some computationally efficient model selection and robust estimation methods in high-dimensional VAR models. We will establish the desired selection consistency and estimation efficiency results under some sparsity conditions. We will also illustrate the performance of the proposed algorithm through extensive simulation and several financial time series data.

Keywords: Vector autoregressive (VAR) models, High-dimensional time series, Model selection, Robust estimation

(二)玉山學者團隊合作情形 (請敘明團隊成員及合作方式)

1.團隊成員名單：(成員應包括校內副教授職級以下成員或博士後研究人員)

姓　名	單　位	職　稱
蔡瑞胸	芝加哥大學	講座教授
銀慶剛	國立清華大學	特聘教授
徐南蓉	國立清華大學	教　授
黃信誠	中央研究院	研究員
蔡恆修	中央研究院	研究員
鄭又仁	國立清華大學	副教授
黃世豪	國立中央大學	助理教授
黃學涵	國立清華大學	博士生
邱海唐	國立清華大學	博士後
黃碩傑	芝加哥大學	博士生

2.合作方式：

(1) 銀慶剛、黃學涵、黃碩傑參與研究計畫主題一：Model Selection for Unit-root Time Series with Many Predictors。

(2) 徐南蓉、黃信誠、鄭又仁、黃世豪參與研究計畫主題二：Spatial Temporal Modeling and Inference。

(3) 銀慶剛、蔡恆修、邱海唐與研究計畫主題三：Robust Estimation in High-Dimensional Vector Autoregressions。

(4)三個研究計畫的成員不僅專注於各自領域，且彼此間密切合作，透過互相觀摩刺激出更多創新的技術及觀念。

(三)績效說明 (請說明達到量化或質化之具體成果與績效、對學校發展之具體助益等)

1. 第一年完成 4 篇合作論文，並為第二及三年的玉山學者計畫，敲定多項合作議題。
2. 上述文章預期可發表在統計、計量經濟或機器學習的頂級期刊，增加本校在資料科學領域的國際能見度。
3. 指導數位本校統計所博士生，並將安排其中表現特優者至芝加哥大學訪問，為本校培育學術菁英。
4. 以清大玉山學者的身份，至國內多家大學擔任(國際)學術會議主講，除推動資料科學在台灣的發展外，亦提升本校在資料科學領域的聲望。
5. 安排多位國際知名統計學者至統計所訪問，協助本校統計所的年輕教師及研究生進一步與國際接軌。