行政院國家科學委員會專題研究計畫 成果報告

使用標的對應訊息的新經濟預測方法
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 99-2410-H-004-058-
執 行 期 間 ：99 年 08 月 01 日至 100 年 07 月 31 日
執 行 單 位 ：國立政治大學經濟學系

計 畫 主 持 人：徐士勛

計畫參與人員：碩士班研究生-兼任助理人員：張慈恬
博士班研究生-兼任助理人員：曾憲政
博士班研究生-兼任助理人員：陳玟吟
博士班研究生-兼任助理人員：林于智
博士班研究生-兼任助理人員：徐兆璿

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 100 年 10 月 24 日

1 Introduction

Because of the recent advances in information technology, literally thousands of time series data are available and this progress motivates people to take advantage of the richer bulk of information in forecasting. In the literature, one of the most popular approaches recently is the factor approach framework of Stock and Watson (1998, 2002). In this framework, these unknown factors are estimated by using the principal component analysis (PCA). then the linear model for the target variable is estimated by the ordinary least squares (OLS) when these estimated factors are plugged in. Since the number of factors is unknown in practice, the information criteria is suggested to introduce in either estimation stage or forecasting stage, when these "natural ranked" principal components are considered in sequence, see e.g., Stock and Watson (1998, 2002) and Bai and Ng (2002).

The factor approach is easy to implement and better performance of this approach relative to the conventional AR model is also investigated in the literature. However, since the prediction accuracy measured by the mean-squared error (MSE) for the target variable is our major concern, there are some remarks in either stages in this framework. First, no matter what the target variable is, the factor estimates and their "natural rankings" are determined by PCA in the estimation stage only. Some empirical studies showed that however, these resulting dominant principal components may not have the forecasting power for some specific target variable. The "natural ranking" property of principal components should be carefully re-investigated in the forecasting stage and the prediction accuracy may be improved if the "targeted ranking" is provided instead. On the other hand, the prediction accuracy may also be improved if the construction of factors is connected to the target variable. The hard thresholding and soft thresholding methods of Bai and Ng (2008a) and combining forecast and then principal component (CF-PC) method of Huang and Lee (2009) are thus proposed to supervise or to weight variables based on the predictive powers to the target variable before implementing PCA in the estimation stage. Second, either in the original factor models or the above supervised methods, the linear model in forecasting stage is estimated by using OLS. It is well known that the OLS estimates often have low bias but large variance. The prediction accuracy can sometimes be improved by shrinking or setting to zero some coefficients. There have been lots of subset selections and shrinkage methods proposed in the literature, such as forward- and backward-stepwise selection, forward-stagewise regression, ridge regression, the least absolute shrinkage and selection operator (LASSO), least angle regression (LARS), and so on, see e.g., Tibshirani (1996), Efron et al. (2004), De Mol et al. (2008) and Hastie et al. (2009). The prediction accuracy could be further improved if these shrinkage and subset selections methods are introduced in the forecasting stage in place of the OLS.

In what follows, we present the forecasting model in which we are interested, the proposed two-step approach including PCA and LASSO, the properties of the proposed, simulation results and an empirical study. Then we conclude this project briefly.

2 The forecasting models

Let y_t denote the target variable of interest and $\boldsymbol{X}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ consists of N observed variables (predictors)at time t, for $t = 1, \ldots, T$. In this project, because our focus is on the selection or aggregation of $\boldsymbol{X}_t$ in the forecasting model, we simply consider the following model of y_{t+h} without other variables involved:

$$y_{t+h} = \boldsymbol{\beta}'\boldsymbol{X}_t + e_{t+h}, \quad t = 1, 2, \ldots, T-h. \tag{1}$$

If $N \ll T$, then the consistent estimator of $\boldsymbol{\beta}$, denoted $\hat{\boldsymbol{\beta}}$, respectively, are obtained by the OLS estimation method. When N is quite large, it is well known that the estimates may not so much reliable, and consistently estimating the model is even not possible when N is larger than the sample size T. Therefore, we may try to aggregate or to extract information of $\boldsymbol{X}_t$ by some transformations. Let $\mathcal{T}$ be a transformation of $\boldsymbol{X}_t$ such that $\mathcal{T}(\boldsymbol{X}_t)$ is a $r \times 1$ vector of functions of $\boldsymbol{X}_t$ for some r. If $r \ll N$, then we may also consider a smaller but effective regression model based on these new r regressors:

$$y_{t+h} = \boldsymbol{\beta}'\mathcal{T}(\boldsymbol{X}_t) + e_{t+h}, \quad t = 1, 2, \ldots, T-h. \tag{2}$$

In the literature, there are many methods for constructing $\mathcal{T}(\boldsymbol{X}_t)$ in the forecasting model (2). In particular, factor approach of Stock and Watson (1998, 2002) is the popular one recently. In this approach, it relies on an additional assumption that $\boldsymbol{X}_t$ is driven by only a few underlying unknown factors, $\mathcal{T}(\boldsymbol{X}_t)$ can be formed as the estimates of these factors, and then the estimates of $\boldsymbol{\beta}$ is obtained by OLS.

3 The proposed approach

In this project, I propose a promising approach to forecast y_{t+h} based on extracting useful information from large dimensional $\boldsymbol{X}_t$. In the first step, I search for orthogonally linear combinations of $\boldsymbol{X}_t$ by principal component analysis. In the second step, I determine the $\mathcal{T}(\boldsymbol{X}_t)$ from these N linear combinations of $\boldsymbol{X}_t$ and the estimates of $\boldsymbol{\beta}$ simultaneously by implementing the LASSO.

3.1 Principal component analysis (PCA)

Assume $N \times 1$ vector $\boldsymbol{X}_t$ has the covariance matrix Σ_x, then the PCA searches for the N mutually orthogonally normalized linear combinations of $\boldsymbol{X}_t$, say $h_1'\boldsymbol{X}_t, h_2'\boldsymbol{X}_t, \ldots, h_N'\boldsymbol{X}_t$, such that for $m = 1, \ldots, N$,

$$h_m = \arg\max_h h'\Sigma_x h, \quad s.t. \;\; h'h = 1, \quad h'\Sigma_z h_i = 0, i = 1, 2, \ldots, m-1.$$

As a consequence, $h_1'\boldsymbol{X}_t$, known as the first principal component of $\boldsymbol{X}_t$, is the linear combination with largest variance; $h_2'\boldsymbol{X}_t$, the second principal component of $\boldsymbol{X}_t$, is with second largest variance, and so on. In practice, we often rescale $\boldsymbol{X}_t$ by standardizing it to have mean zero and unit variance. The sample version of PCA is then carried out by replacing Σ_x with the sample covariance matrix $\widehat{\Sigma}_x = T^{-1}\sum_{t=1}^{T}\boldsymbol{X}_t\boldsymbol{X}_t'$. Denote $\hat{h}_1, \ldots, \hat{h}_N$ as the sample version of $h_1, \ldots, h_N$, then we further define the mapping

$$\mathrm{PC}(\boldsymbol{X}_t; \kappa) = [\hat{h}_1, \ldots, \hat{h}_\kappa]'\boldsymbol{X}_t, \ \kappa = 1, \ldots, N. \tag{3}$$

3.2 LASSO

Consider a general linear model $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_o + u_t$, $t = 1, \ldots, T$, where y is the regressand, $\boldsymbol{x}_t$ is the $k \times 1$ vector of standardized regressors and $\boldsymbol{\beta}_o$ is the corresponding unknown coefficients. Denote $\mathrm{RSS}(y, \boldsymbol{x}; \boldsymbol{\beta})$ the residual sum of squares from this linear model given some $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_k\}'$, then we represent the LASSO estimates as

$$\widehat{\boldsymbol{\beta}}^{\text{lasso}}(y, \boldsymbol{x}) = \arg\min_{\boldsymbol{\beta}} \ \mathrm{RSS}(y, \boldsymbol{x}; \boldsymbol{\beta}) \quad \text{subject to} \sum_{i=1}^{k} |\beta_i| \le c, \tag{4}$$

where $c \ge 0$ controls the amount of shrinkage applied to the estimates, and it shrinks some coefficients and sets others to zero. Besides, the LASSO also provides a ranking of $\boldsymbol{x}$ according to the absolute value of $\widehat{\boldsymbol{\beta}}^{\text{lasso}}(y, \boldsymbol{x})$ since $\boldsymbol{x}$ is standardized. This ranking is targeted because it reveals the predictive power of each elements of $\boldsymbol{x}$ in the model. When the regressors $\boldsymbol{x}$ are orthogonal, the LASSO estimate gives an analytic form:

$$\widehat{\beta}_i^{\text{lasso}}(y, \boldsymbol{x}) = \mathrm{sign}\left(\widehat{\beta}_i^{\text{ols}}(y, \boldsymbol{x})\right)\left(\left|\widehat{\beta}_i^{\text{ols}}(y, \boldsymbol{x})\right| - \mu/2\right)_+, \tag{5}$$

where $\mathrm{sign}(\cdot)$ denotes the sign of its argument (± 1), and $z_+ = z$ if $z > 0$ and 0 otherwise. It indicates that under the LASSO, only the variables with absolute value of OLS estimates larger than $\mu/2$ will be retained in the model, and the LASSO estimate of these retained variables are the corresponding OLS estimates shrunk by the amount $\mu/2$.[1] In what follows, we sometimes call $\mu/2$ as the "threshold" of the LASSO estimates.

3.3 The proposed

In this project, I take both the advantages of the PCA and the LASSO by a two-step approach. In the first step, I re-organize the information containing in the original regressors $\boldsymbol{X}_t$ by using PCA. K_{NT}

[1] As noted by Tibsirani (1996), the parameter μ in the LASSO could be determined by three methods, they are the cross-validation, generalized cross-validation and an analytical unbiased estimate of risk.

orthogonally linear combinations of $\boldsymbol{X}_t$ constructed by PC($\boldsymbol{X}_t$; K_{NT}) will be introduced in forecasting y_{t+h}. Note that K_{NT} could be a (increasing) large number depending on N and T. In the second step, the $\mathcal{T}(\boldsymbol{X}_t)$ and the estimates of $\boldsymbol{\beta}$ in forecasting model (2) will be simultaneously determined by implementing the LASSO, that is, the estimate of $\boldsymbol{\beta}$ would be $\widehat{\boldsymbol{\beta}}^{\text{lasso}}(y_{t+h}, \text{PC}(\boldsymbol{X}_t;\, K_{NT}))$, and non-zero elements of $\widehat{\boldsymbol{\beta}}^{\text{lasso}}(y_{t+h}, \text{PC}(\boldsymbol{X}_t;\, K_{NT}))$ indicates which principal components are helpful to predict y_{t+h}.

There are some remarks. In the first step of the proposed, as what emphasized above, I consider all principal components of $\boldsymbol{X}_t$ instead of only some of them in practice. It differs greatly from the factor approach of Stock and Watson(1998, 2002) and other related. In the second step, I implement the LASSO to decide which principal components of $\boldsymbol{X}_t$ would helpful to predict y_{t+h}, the helpless principal components would receive zero LASSO coefficients. Unlike the "natural ranking" of PC($\boldsymbol{X}_t$; N), it gives the "targeted ranking" instead. Moreover, because the principal components are orthogonal, the LASSO gives the analytic form of $\widehat{\boldsymbol{\beta}}^{\text{lasso}}(y_{t+h}, \text{PC}(\boldsymbol{X}_t;\, N))$, it is more easily to implement than the estimate $\widehat{\boldsymbol{\beta}}^{\text{lasso}}(y_{t+h}, \boldsymbol{X}_t)$ proposed by De Mol et al. (2008), and good properties of the LASSO, comparing to other shrinkage methods, are kept especially when $\boldsymbol{X}_t$ are highly correlated.

4 The properties of the proposed approach

In the proposed two-step approach, the forecast performance in the second step heavily depends on the principal components constructed in the first step. Given the vector form representation of $\boldsymbol{X}_t$:

$$X_t = \Lambda F_t + u_t,$$

where $F_t = (f_{1t}, f_{2t}, \ldots, f_{Kt})'$ is a $K \times 1$ vector of unobserved common factors at time t, Λ is the $N \times 1$ corresponding matrix of factor loadings, and u_t is $N \times 1$ idiosyncratic errors. We consider the assumptions of $\boldsymbol{X}_t$ as follows:

Assumption 4.1

(a) $\boldsymbol{X}_t$ *is a* $N \times 1$ *vector of covariance-stationary processes with mean zero.*

(b) $\mathbb{E}[F_t] = \mathbb{E}[u_t] = \mathbb{E}[F_t u_t] = 0$, $\mathbb{E}[F_t F_t'] = I_K$, *the covariance structure of* $\boldsymbol{X}_t$ *is given by* $\Sigma_X = \Lambda\Lambda' + \Omega$, *where* Σ_X *and* Ω *are the* $N \times N$ *population covariance matrix of* $\boldsymbol{X}_t$ *and* u_t, *respectively.*

(c) Denote $\lambda_{\min}(A)$ *and* $\lambda_{\max}(A)$ *the smallest and the greatest eigenvalues of a matrix* A, *and* $\|A\| = (\lambda_{\max}(A'A))^{1/2}$,

$$0 < \liminf_{n\to\infty} \frac{1}{N}\lambda_{\min}(\Lambda'\Lambda) \leq \limsup_{n\to\infty} \frac{1}{N}\lambda_{\max}(\Lambda'\Lambda) < \infty$$

These assumptions on $\boldsymbol{X}_t$ are quite standard in the literature, and these assumptions ensure that the PCA can provide an consistent estimators of F_t, provided the N and T go to infinity; see Bai and Ng (2008) and De Mol et al. (2008) for example. Let $\hat{F}_t$ and $\hat{\lambda}$ be the estimators of F_t and λ by PCA, then the consistency and asymptotic normality of these estimators are directly follows the Result A in Bai and Ng (2008b):

Theorem 4.2 (Consistency of factor estimators)
Given Assumption 4.1,

(a) If $N = o(T^2)$ as $N, T \to \infty$, then for each t, $\sqrt{N}(\hat{F}_t - F_t)$ is asymptotic normal distributed with zero mean.

(b) If $T = o(N^2)$ as $N, T \to \infty$, then for each i, $\sqrt{T}(\hat{\lambda}_i - \lambda_i)$ is asymptotic normal distributed with zero mean.

Assumption 4.3

(a) $y_{t+h} = \boldsymbol{\beta}'\boldsymbol{F}_t + e_{t+h}$, where $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients and some coefficients are zeros, and e_{t+h} is orthogonal to $\boldsymbol{X}_t$.

(b) Assume that all elements of $\boldsymbol{\beta}$ are shrunk to zero and the prior distributions of which are i.i.d double-exponential, that is,

$$f(\beta_k^o) = \frac{1}{2\sigma_{\beta^o}^2} \exp\left(\frac{-|\beta_k|}{\sigma_{\beta^o}^2}\right)$$

Assumption 4.3(a) is typically considered in the factor model literature, except that we explicitly assume that not all of factors in Assumption 4.1 are effective for forecasting y. Assumption 4.3(b) states the prior distribution of $\boldsymbol{\beta}$, which links the Lasso estimators of $\boldsymbol{\beta}$ to the posterior mean of $\boldsymbol{\beta}$.

Theorem 4.4 (The property of Lasso estimators)
Given Assumptions 4.1 and 4.3, for $k = 1, \ldots, K$, we have the posterior mean of β_k as

$$\beta_k^p = sign\left(\widehat{\beta}_k^{ols}(y, \hat{F}_t)\right)\left(\left|\widehat{\beta}_k^{ols}(y, \hat{F}_t)\right| - \frac{1}{2\sigma_{\beta^o}^2}\right)_+,$$

where $\widehat{\beta}_k^{ols}(y, \hat{F}_t)$ is the kth element of OLS estimates when regressing y on $\hat{F}_t$, and $\hat{F}_t$ is the estimator obtained form PCA of $\boldsymbol{X}_t$. Moreover, β_k^p shares the same form as the LASSO estimator for β_k.

This Theorem implies that some β_ks will exactly be set to zero if their magnitudes are less than the "threshold" $(2\sigma_{\beta^o}^2)^{-1}$. It means that the less important principal components induced from $\boldsymbol{X}_t$ for forecasting y would be discarded.

5 Simulation

In this section, we report one of the simulation results we did. This simulation called "oversampling" is introduced in Boivin and Ng (2006). The oversampling means that the data are more informative on only some most dominant factors. The experiment is designed as follows. First,two serially correlated factors drive the data:

$$F_{kt} = 0.5F_{kt-1} + u_{kt},\ u_{kt} \sim N(0, 1),\ k = 1, 2$$

Second, there are two target series:

$$y^A_{t+1} = \beta^A F_{1t} + \varepsilon^A_{t+1},\ \varepsilon^A \sim N(0, \sigma^A),\quad y^B_{t+1} = \beta^B F_{2t} + \varepsilon^B_{t+1},\ \varepsilon^B \sim N(0, \sigma^B),$$

where $\beta^A = 1 = \beta^B$, $\sigma^A = 1 = \sigma^B$. Third, five types of data with sample size N_s, $s = 1, 2, 3, 4, 5$, $N_1 : X_{it} = 0.8F_{1t} + e_{it},\ e_{it} \sim N(0, 1 - 0.8^2)$; $N_2 : X_{it} = 0.6F_{2t} + e_{it},\ e_{it} \sim N(0, 1 - 0.6^2)$; $N_3 : X_{it} = 0.4F_{1t}+0.1F_{2t}+e_{it},\ e_{it} \sim N(0, 1-0.4^2-0.1^2)$; $N_4 : X_{it} = 0.1F_{1t}+0.4F_{2t}+e_{it},\ e_{it} \sim N(0, 1 - 0.1^2 - 0.4^2)$; $N_5 : X_{it} = e_{it},\ e_{it} \sim N(0, 1)$. Fourth, 14 cases are considered by various combinations of N_i. They are divided into three groups. In the first group, factor F_1 dominates F_2, the cases are Case 1($N_1 = 20$), Case 5 ($N_1 = 20$, $N_3 = 20$) and Case 11 ($N_1 = 20$, $N_3 = 20$, $N_5 = 40$). The second group considers the cases that F_2 dominates F_1, they are Case 2 ($N_2 = 20$), Case 4 ($N_2 = 20$, $N_3 = 20$) and Case 6 ($N_2 = 20$, $N_4 = 40$). The others are Case 3 ($N_1 = 20$, $N_2 = 20$), Case 10 ($N_1 = 20$, $N_2 = 20$, $N_5 = 40$), Case 7 ($N_1 = 20$, $N_2 = 20$, $N_3 = 20$), Case 12 ($N_1 = 20$, $N_2 = 20$, $N_3 = 20$, $N_5 = 40$), Case 8 ($N_1 = 20$, $N_2 = 20$, $N_4 = 40$), Case 13 ($N_1 = 20$, $N_2 = 20$, $N_4 = 40$, $N_5 = 40$), Case 9 ($N_1 = 20$, $N_2 = 20$, $N_3 = 20$, $N_4 = 40$), and Case 14 ($N_1 = 20$, $N_2 = 20$, $N_3 = 20$, $N_4 = 40$, $N_5 = 40$).

The number of replications of each case is 200, and we consider two measures of estimates in each case, they are Average of in-sample MSEs (based on 200 in-sample periods) and out-of-sample MSEs(based on 200 out-of-sample periods). Different prior variance of $\boldsymbol{\beta}$ are chosen such that the thresholds of the proposed Lasso method equal 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3. We then compare the results of the proposed with the original forecasting model proposed by Stock and Watson (2002)(denoted SW), and the hard thresholding model with threshold for t-ratio equals 1.28 (denoted HT1) and 1.65(denoted HT2). Besides, because only two factors drive the data, we consider $q = 1$ and 2, the number of factors, in SW, HT1 and HT2 methods. The simulation results, the ratio of the MSE for a given method to the MSE of SW model, are reported in Table 1 for y_A.[2]

When $q = 1$ for variable y_A, the proposed Lasso method performs best in the cases that F_2 dominates F_1, for example case 2, 4 and 6, no matter what $1/2\sigma^2_{\beta^o}$ is chosen. Similarly, when $q = 1$ for variable y_B, we get the similar results when in the case 3 and 5. However, when $q = 2$, the first

[2] For saving the space of this report, we do not report the results for y_B.

two estimated factors capture all important information about F_1 and F_2 as well as y_A and y_B, so the advantages of Lasso method gradually vanish, but Lasso estimates with some prior variances are also comparable with HT1 and HT2.

6 Empirical Study

In the empirical study, we consider the data set which has been analyzed in Stock and Watson (2005), Bai and Ng(2008a) and Lee and Tu (2009). We briefly summarize the features of this data set and this study as follows. First, monthly series available from 1960m1 to 2003m12 for a total of $T = 528$ observations, and $N = 132$ variables in this data set. Second, the target variable y is the logarithm of PUNEW (CPI all items, the 115th variable in the data set.), as what in Bai and Ng(2008a), Lee and Tu(2009). Third, We consider a h step-ahead forecasting Model of y for $h = 1$, 3, 6, 12, 18, 24, 30 and 36. Fourth, nine values of threshold from 0.3 to 0.9 as well as the one decided by BIC are considered in the proposed Lasso estimation method. The benchmark model is AR(4) model and all results the ratio of the MSE for a given method to that AR(4)model. Three out-of-sample periods (1971m1 to 2003m12, 1971m1 to 1979m12, 1980m1 to 1898m12)are considered, and we summarize the results in Table 2.

There are some remarks in Table 2. First, except the cases with thresholds smaller than 0.2 and $h = 1$, the proposed method beats AR(4) model is most cases. Second, as h increases, the advantages of the proposed is more significant. Third, the performance of the proposed with the BIC-chosen threshold is good in the third sample period(1980m1 to 1989 m12) but is not in the other two sample periods. We also observed the similar phenomenon in the unreported results when other data-driven methods such as AIC and cross-validation are considered. It suggests that it would be interesting to find a better data-driven threshold method when computing Lasso.

7 Concluding Remarks

In this project, I proposed a promising approach to forecast target variable based on extracting useful information from large dimensional variables. I take both the advantages of the PCA and the LASSO by a two-step approach. In the first step, I re-organize the information containing in these many predictors by using PCA. All these orthogonally principal components are then introduced in forecasting target variable. It differs greatly from the factor approach of Stock and Watson(1998, 2002) and other related, when only the dominant principal components (according to the“natural ranking”) would be considered in these factor approaches. In the second step, the targeted principal components and the estimates of the linear forecasting model will be simultaneously determined by implementing the LASSO. The helpless principal components would receive zero LASSO coefficients. Unlike the "natural ranking" of PCA, it gives the "targeted ranking" instead. Moreover, because the principal components are orthogonal, good properties of the LASSO, comparing to other shrinkage methods, are kept, and it also gives the analytic form of the estimates. In contrast with the other forecasting approaches introduced in the literature, the proposed approach in this project is easily to implement. The properties of the proposed estimates are also given when we equip $\boldsymbol{X}$ with the factor structure as typically introduced in the literature but with many factors, and link the LASSO estimates to the posterior mean of $\boldsymbol{\beta}$ in Bayesian analysis. The simulation results show that the proposed performs good in some cases where the "natural ranking" of PCA is helpless for forecasting the target variable. However, the simulation results also suggest us that there does not exist a best method/approach to forecasting, it heavily depends on the relationships among the data, the target variable and the underlying factors. More effort to clearly investigate the properties of the existing methods/approaches is needed to make in the future. On the other hand, the results of the empirical study also indicates that a good data-driven method for threshold in LASSO is worth studying in the future.

References

Bai, J. and S. Ng (2002), Determine the number of factors in approximate factor models, *Econometrica*, **71**, 135-171.

Bai, J. and S. Ng (2008a), Forecasting economic time series using targeted predictors. *Journal of Econometrics*, **146**, 304-317.

Bai, J. and S. Ng (2008b), Large dimensional factor analysis. *Foundations and Trends ® in Econometrics*, **3**, 89-163.

Boivin, J. and S. Ng (2006). Are More Data Always Better for Factor Analysis. *Journal of Econometrics*, **132**, 169-194.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian regression a valid alternative to principal components. *Journal of Econometrics*, **146**, 318-328.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, (2004), Least angle regression. *Annals of Statistics*, **32**, 407-499.

Hastie, T., R. Tibshirani, and J. Friedman (2009), The Elements of Statistical Learning. Springer-Verlag.

Huang, H. and T.H., Lee (2009), To combine forecasts or to combine information? *Econometric Reviews*, forecasting.

Hyvarinen, A., J. Karhunen, and E. Oja (2001), Independent Component Analysis. John Wiley & Sons, Inc., New York.

Hyvarinen, A. and E. Oja (1997), A fast fixed-point algorithm for independent component analysis. *Neural Computation*, **9**, 1483-1492.

Lee, T.H. and Y.D. Tu (2009), Forecasting using supervised factor models. Working paper.

Stock, J. H. and M. W. Watson (1998), Diffusion Indexes, NBER Working Paper 6702.

Stock, J. H. and M. W. Watson (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics*, **20**, 147-162.

Tibshirani, R., (1996), Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267-288.

Table 1: Oversampling simulations for y_A.

YA		SW	HT1	HT2	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO	SW	HT1	HT2	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO
	q,u/2	1	1	1	0.05	0.1	0.15	0.2	0.25	0.3	2	2	2	0.05	0.1	0.15	0.2	0.25	0.3
Ave. In-Sample MSE	Case 1	1.035504	1	1	0.962111	0.997686	1.019758	1.038629	1.060699	1.087419	1.028884	1	1	0.966643	1.001892	1.023451	1.042383	1.064638	1.0916
Out-of-Sample MSE	Case 1	1.043134	1	1	1.019968	1.013766	1.024475	1.042299	1.065141	1.092896	1.053917	1	1	1.014134	1.009241	1.019103	1.035733	1.057367	1.083859
Ave. In-Sample MSE	Case 5	1.041354	0.999969	0.999954	0.915337	0.983566	1.017066	1.038081	1.060308	1.086918	1.029105	0.999998	0.999996	0.918876	0.987281	1.021502	1.042991	1.065495	1.092419
Out-of-Sample MSE	Case 5	1.049858	0.99999	0.99998	1.038344	1.018082	1.026803	1.044755	1.067973	1.096092	1.046233	0.999992	1.000004	1.034605	1.013602	1.02159	1.038347	1.06011	1.086843
Ave. In-Sample MSE	Case 11	1.055978	0.995898	0.996819	0.835217	0.960353	1.013138	1.037227	1.059424	1.085647	1.046481	0.99256	0.99318	0.839398	0.965121	1.017327	1.0411	1.063465	1.089956
Out-of-Sample MSE	Case 11	1.04399	0.99988	0.999854	1.090569	1.025212	1.025292	1.041394	1.063674	1.090761	1.060932	1.002773	1.00289	1.083201	1.02086	1.022512	1.038997	1.061015	1.087766
Ave. In-Sample MSE	Case 2	2.325695			0.951273	0.976298	0.993365	1.00106	1.003531	1.004154	2.311983			0.952888	0.978342	0.996118	1.004435	1.00738	1.008194
Out-of-Sample MSE	Case 2	2.324229			1.02561	1.00594	0.999264	0.996926	0.996381	0.996276	2.332114			1.024825	1.004493	0.997791	0.996036	0.99565	0.995498
Ave. In-Sample MSE	Case 4	2.266269	0.642712	0.601877	0.502081	0.539336	0.562196	0.57824	0.59398	0.611037	1.216955	1.007206	1.017857	0.919738	0.98751	1.027702	1.055711	1.083526	1.114028
Out-of-Sample MSE	Case 4	2.289398	0.649963	0.609224	0.575274	0.563673	0.569638	0.581344	0.595594	0.611486	1.246868	1.017205	1.028538	1.042325	1.024849	1.03738	1.059771	1.086406	1.11598
Ave. In-Sample MSE	Case 6	2.321974	0.974092	0.962513	0.724916	0.816683	0.897211	0.950771	0.981527	0.997788	2.261416	0.981525		0.740593	0.834649	0.917362	0.972212	1.003386	1.019915
Out-of-Sample MSE	Case 6	2.320652	0.994379	0.995319	0.964516	0.945088	0.959554	0.977966	0.990917	0.998713	2.328189	1.006542		0.97504	0.956505	0.971783	0.990237	1.003213	1.010698
Ave. In-Sample MSE	Case 3	1.05899	0.982904	0.982836	0.909126	0.975214	1.010185	1.033559	1.05745	1.084774	1.046889	0.996332	0.995852	0.92519	0.992959	1.028834	1.05346	1.078177	1.105942
Out-of-Sample MSE	Case 3	1.060646	0.987329	0.986412	1.025378	1.005969	1.01729	1.036906	1.060944	1.08871	1.052888	1.003659	1.003822	1.038925	1.018124	1.028516	1.04778	1.071244	1.098535
Ave. In-Sample MSE	Case 10	1.052647	0.977402	0.978263	0.82675	0.951609	1.006065	1.033468	1.058197	1.085674	1.047848	0.989819	0.989112	0.846029	0.971199	1.025511	1.051982	1.076174	1.103716
Out-of-Sample MSE	Case 10	1.067328	0.983931	0.982206	1.06641	1.012641	1.021389	1.042641	1.067712	1.096299	1.060878	1.002977	1.004536	1.088354	1.028573	1.035046	1.054624	1.078853	1.107036
Ave. In-Sample MSE	Case 7	1.062233	0.986168	0.9858	0.863983	0.961078	1.007543	1.03368	1.057785	1.084994	1.039819	0.996423	0.997024	0.88186	0.980737	1.028614	1.056578	1.082513	1.110597
Out-of-Sample MSE	Case 7	1.065685	0.989185	0.988491	1.044645	1.009055	1.019219	1.038668	1.061972	1.089308	1.047863	1.003079	1.003513	1.0543	1.020675	1.029978	1.050007	1.073804	1.10103
Ave. In-Sample MSE	Case 12	1.072619	0.978589	0.978989	0.776662	0.932269	0.999358	1.029956	1.055537	1.083357	1.033713	0.99188	0.992592	0.798704	0.957959	1.024757	1.055082	1.081022	1.109602
Out-of-Sample MSE	Case 12	1.068699	0.985533	0.984347	1.106124	1.016701	1.016807	1.035968	1.060199	1.088093	1.059701	1.003524	1.004684	1.111762	1.029002	1.031655	1.050173	1.073738	1.101335
Ave. In-Sample MSE	Case 8	1.989246	0.56293	0.542031	0.443201	0.512775	0.546214	0.567361	0.589834	0.616437	1.035638	0.996454	0.995401	0.841633	0.972222	1.035067	1.074792	1.117113	1.167123
Out-of-Sample MSE	Case 8	1.990182	0.575688	0.553427	0.578961	0.548499	0.55479	0.571312	0.593035	0.619213	1.05088	1.001687	1.003327	1.094668	1.03645	1.049069	1.08219	1.124669	1.175393
Ave. In-Sample MSE	Case 13	1.980986	0.556909	0.53878	0.401443	0.501402	0.544836	0.567477	0.589997	0.616513	1.03469	0.991138	0.990992	0.760518	0.950089	1.031642	1.074547	1.117449	1.168015
Out-of-Sample MSE	Case 13	2.024501	0.572741	0.553349	0.624566	0.56019	0.56215	0.579005	0.601374	0.627995	1.031015	1.001752	1.003438	1.173897	1.046216	1.051072	1.083729	1.12665	1.177874
Ave. In-Sample MSE	Case 9	1.509624	0.727487	0.711251	0.551279	0.662803	0.714008	0.743447	0.774062	0.81061	1.031607	1.000142	1.000201	0.799562	0.960943	1.034369	1.076779	1.121177	1.174347
Out-of-Sample MSE	Case 9	1.534479	0.736108	0.719121	0.777647	0.720305	0.728074	0.75155	0.782357	0.819667	1.047845	1.001994	1.003141	1.121689	1.038165	1.045161	1.076893	1.119977	1.172568
Ave. In-Sample MSE	Case 14	1.515154	0.731451	0.71505	0.500982	0.653525	0.717755	0.749379	0.780151	0.816614	1.038282	0.996434	0.997658	0.716326	0.936839	1.030137	1.075953	1.120427	1.173211
Out-of-Sample MSE	Case 14	1.525263	0.743799	0.726612	0.86475	0.742609	0.738105	0.760246	0.791039	0.828381	1.058136	0.999768	1.000831	1.230755	1.05741	1.051544	1.081986	1.125038	1.177416

Table 2: Empirical Study for CPI growth rate.

h-step-ahead	MSE(AR4)	Threshold 0.1	Threshold 0.2	Threshold 0.3	Threshold 0.4	Threshold 0.5	Threshold 0.6	Threshold 0.7	Threshold 0.8	Threshold 0.9	BIC
1	7.44933793	3.214958	2.210637	1.458799	1.114014	1.056621	1.007355	1.001233	1	1	1.164701
3	4.77968046	2.594722	1.583455	0.967536	0.942586	0.944539	0.968834	0.973249	1.00612	1.003375	1.742428
6	3.55952325	1.553412	0.981737	0.864835	0.831469	0.873659	0.873068	0.872723	0.922865	0.990213	1.071023
12	3.85573211	1.232488	0.908485	0.742233	0.737798	0.764266	0.768346	0.844843	0.844815	0.844395	0.867362
18	4.68277377	1.288417	0.814997	0.697931	0.707077	0.694798	0.719145	0.740757	0.756269	0.834292	0.771476
24	5.65276712	1.372287	0.914035	0.659706	0.67841	0.695208	0.693266	0.713335	0.727315	0.80949	0.825372
30	6.27362068	1.405901	0.878487	0.677889	0.687993	0.702269	0.698987	0.754162	0.773561	0.780782	0.725426
36	6.70593233	1.302993	0.804089	0.669728	0.675233	0.723606	0.750969	0.785298	0.774824	0.797068	0.800743

(a)Sample Period: 1971m1-2003m12

h-step-ahead	MSE(AR4)	Threshold 0.1	Threshold 0.2	Threshold 0.3	Threshold 0.4	Threshold 0.5	Threshold 0.6	Threshold 0.7	Threshold 0.8	Threshold 0.9	BIC
1	8.63021973	6.524515	3.832826	1.995896	1.244299	1.133836	1.009041	1	1	1	1.501026
3	4.9947297	5.151742	2.587515	0.953879	0.998404	1.048725	1.057679	1.037843	1	1	3.572843
6	4.12840449	1.903407	0.779259	0.83787	0.865485	1.030239	1.028507	1.027494	1.021918	1	1.501665
12	5.37354376	1.052408	0.838365	0.690966	0.766233	0.793093	0.790746	0.972632	0.972564	0.971538	1.009397
18	7.03283582	0.978252	0.732098	0.653813	0.728387	0.729377	0.763228	0.765244	0.772605	0.955498	0.877289
24	8.0300859	1.128017	0.86551	0.658841	0.726202	0.75616	0.763285	0.806948	0.841166	0.910754	1.057574
30	7.78617708	1.154358	0.828098	0.742234	0.819216	0.837153	0.846924	0.959724	0.992254	0.99506	0.892022
36	7.87843836	1.053357	0.783491	0.802046	0.895199	0.997057	1.015269	0.999771	1.003947	0.999332	1.152659

(b)Sample Period: 1971m1-1979m12

h-step-ahead	MSE(AR4)	Threshold 0.1	Threshold 0.2	Threshold 0.3	Threshold 0.4	Threshold 0.5	Threshold 0.6	Threshold 0.7	Threshold 0.8	Threshold 0.9	BIC
1	9.53430908	1.54827	1.504408	1.267371	1.077957	1.035635	1.011354	1.003277	1	1	0.990803
3	7.88725786	1.352604	1.094414	0.940778	0.898488	0.874369	0.918756	0.919338	1.00233	1.006953	0.909773
6	5.74212726	1.254716	0.97937	0.829878	0.778447	0.748915	0.748915	0.748915	0.852603	0.959102	0.809958
12	5.72576391	1.115446	0.721649	0.66673	0.644208	0.68791	0.697773	0.70222	0.70222	0.70222	0.711529
18	6.96269477	1.271482	0.654978	0.59268	0.591288	0.586351	0.599037	0.660268	0.688305	0.681983	0.623104
24	9.18811977	1.308011	0.733926	0.553825	0.568212	0.589932	0.587527	0.594559	0.589406	0.700479	0.61031
30	11.6025394	1.213681	0.732582	0.557883	0.574639	0.590995	0.579743	0.607123	0.628192	0.639583	0.605708
36	13.120488	1.173306	0.689655	0.527198	0.518669	0.543315	0.572515	0.634667	0.61816	0.669337	0.595905

(c)Sample Period: 1980m1-1989m12

國科會補助計畫衍生研發成果推廣資料表

日期:2011/10/17

<table>
<tr><td rowspan="3">國科會補助計畫</td><td colspan="2">計畫名稱: 使用標的對應訊息的新經濟預測方法</td></tr>
<tr><td colspan="2">計畫主持人: 徐士勛</td></tr>
<tr><td>計畫編號: 99-2410-H-004-058-</td><td>學門領域: 數理與數量方法</td></tr>
<tr><td colspan="3">無研發成果推廣資料</td></tr>
</table>

99 年度專題研究計畫研究成果彙整表

計畫主持人：徐士勛	計畫編號：99-2410-H-004-058-
計畫名稱：使用標的對應訊息的新經濟預測方法	

成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	1	1	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	1	1	100%	人次	
		博士生	4	3	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估 ■達成目標 □未達成目標（請說明，以 100 字為限） □實驗失敗 □因故實驗中斷 □其他原因 說明：
2. 研究成果在學術期刊發表或申請專利等情形： 論文：□已發表 □未發表之文稿 ■撰寫中 □無 專利：□已獲得 □申請中 ■無 技轉：□已技轉 □洽談中 ■無 其他：（以 100 字為限）
3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限） 近來由於大量時間序列資料的建立，文獻上已有許多嘗試從其中萃取有用的訊息並應用於經濟預測上的研究，這個計畫的研究範疇也聚焦於此。相較於目前文獻上的方法或模型，我認為此計畫的兩階段預測方法有下列優點。1.容易操作：第一階段中使用的主成分分析法幾乎是所有套裝軟體能計算的統計方法，而第二階段中的 LASSO 運算法也因為第一階段主成分都具有正交的特性而具有明確的形式； 2. 性質清楚：主成分分析法和因子模型的關係與 LASOO 運算法在正交化變數下的結果都相當清楚，同時可以根據不同預測標的變數而得到對應之訊息(主成分)排序。模擬和實證研究結果也顯示這樣的設計雖然不是最佳的模型，但是大部份可以和因子相關模型，如 Stock and Watson~(2002, Journal of Business and Economic Statistics)，Bai and Ng(2008, Journal of Econometrics) 等相抗衡。若能進一步的藉由不同的模擬設計更清楚凸顯該方法與其他方法的各自適用時機，相信能讓研究者對這些方法有更深瞭解，並且研究結果也能發表在較好的國際期刊中。這樣的努力我現在仍再進行中。